Carnegie Mellon Symposia
on Cognition

David Klahr, Series Editor

Anderson: Cognitive Skills and Their Acquisition

Carroll/Payne: Cognition and Social Behavior

Clark/Fiske: Affect and Cognition

Cohen/Schooler: Scientific Approaches to Consciousness

Cole: Perception and Production of Fluent Speech

Farah/Ratcliff: The Neuropsychology of High-Level Vision: Collected Tutorial Essays

Granrud: Visual Perception and Cognition in Infancy

Gregg: Knowledge and Cognition

Just/Carpenter: Cognitive Processes in Comprehension

Klahr: Cognition and Instruction

Lau/Sears: Political Cognition

MacWhinney: The Emergence of Language

MacWhinney: Mechanisms of Language Acquisition

Reder: Implicit Memory and Metacognition

Siegler: Children's Thinking: What Develops?

Sophian: Origins of Cognitive Skills

Steier/Mitchell: Mind Matters: A Tribute to Allen Newell

VanLehn: Architectures for Intelligence

THE EMERGENCE OF LANGUAGE

Edited by

Brian MacWhinney
Carnegie Mellon University

LEA
1999
LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Mahwah, New Jersey London

Schlesinger, I. M. (1977). *Production and comprehension of utterances.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Slobin, D. I. (1997). The origins of grammaticizable notions: Beyond the individual mind. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 5: Expanding the contexts.* Mahwah, NJ: Lawrence Erlbaum Associates.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development.* Cambridge, England: Cambridge University Press.

Tomasello, M., & Brooks, P. J. (1998). Early syntactic development: A construction grammar approach. In M. Barrett (Ed.), *The development of language.* London: UCL Press.

Zipf, G. K. (1935). *The psycho-biology of language.* Boston: Houghton Mifflin.

Chapter 8

The Emergence of Language From Embodiment

Brian MacWhinney
Carnegie Mellon University

"Man is the measure of all things."

–Protagoras

The basic function of language is communication. When the listener succeeds in decoding the message intended by the speaker, the communication has been a success. But exactly how does the speaker package information to make sure that the listener will succeed? What does the listener have to do to build up a mental representation that echoes the original representation in the speaker's mind?

The traditional approach to this problem is one that has focused on the construction of propositional representations (Clark & Clark, 1977; Kintsch, 1974; Levelt, 1989; Schank & Abelson, 1977; Sowa, 1984). In this standard model, a message is represented by a directed graph in which words are joined together by labeled arcs. Although these graphs allow for multiple attachments to a single node, they otherwise resemble the phrase structure tree used in linguistics. This standard, graph-based approach provides a good way of depicting patterns of connectedness between words, but it fails in terms of providing a deeper account of meaning. There is a big gap between the schematic representation provided in a propositional graph and our actual understanding of the activity underlying a sentence. When we look at a picture of a boy letting a frog out of a glass jar, we can form a dynamic representation of the boy turning the lid of the glass jar and the frog hopping out of the jar. Although we could notate structural

aspects of this action sequence through a propositional graph, we cannot use this graph to capture the actual flow of action or the unitized nature of the whole scene. The graph tells us little about the ways in which we move our elbow and wrist to unscrew the lid or the ways in which the frog leaps and jumps.

To deepen the linkage of propositional representations to cognition, cognitive linguists have often sought to unpack lexical forms into component propositional structures. For example, generative semantics provided an account of the meaning of a verb like *kill* that unpacked its lexical form into the predicate structure (cause(become(not(alive)))). Pursuing this form of decompositional semantic analysis, Miller and Johnson-Laird (1976) showed how the meaning of a simple noun like *table* can be unpacked into a series of propositions that explain how the top of the table fits onto its legs, how we place things on tables, and how we work at tables. Langacker (1989) and Talmy (1988) showed how we can enrich propositional representations by linking them to pictorial representations based on the theory of space grammar. These decompositional approaches succeed at enriching propositional representations by unpacking the meaning components of individual lexical items. However, they still leave us with a large gap between the notational system being used and our actual understandings of the meanings of sentences.

By treating communication as the construction of links between abstract symbols, or even pictorial configurations, the standard approach has implicitly accepted a fully Platonic characterization of sentence meaning. In this Platonic view, the meaning of a sentence has its truest existence outside of the minds of individual speakers and listeners. The "true" meaning of an utterance is not dependent on the cognitions of individual speakers or listeners; rather, it is a general cultural possession, or perhaps even a reflection of abstract semantic theory. Although this view makes good sense to those who wish to construct a logical theory of linguistic meaning, it cannot provide a psychologically satisfying account of the processing of meaning. The core problem with the propositional account is that it encourages us to think of representations in terms of disembodied graphs and schematic diagrams.

There is an interesting alternative to the standard propositional account. This alternative account treats language processing as a process of "perspective taking." According to this view, language comprehension and production are embodied processes whose goal is the creation and extraction of embodied meanings. Speakers and listeners use language as a way of working through various perspectives and shifts in perspective grounded on the objects and actions described by language. We can refer to these processes of active embodiment as the *perspective-taking system*. In order to understand sentences, we must become actively involved with a starting

point or initial perspective. We use this perspective as the foundation for building an embodied understanding of the sentence. For example, when we listen to a sentence such as *The skateboarder vaulted over the railing*, 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 understand this utterance. If we know only a little about skateboarding, the perspective we assume will be monochromatic and superficial. We will simply imagine the skateboard and rider somehow jumping 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 that we have available, the more deeply we can understand the sentence.

FOUR PERSPECTIVAL SYSTEMS

The human mind constructs perspectives on four levels: (1) affordances, (2) spatio-temporal reference frames, (3) causal action chains, and (4) social roles. The hypothesis being developed here is that these four perspectival systems are grounded on specific brain structures that have evolved to solve major adaptive challenges. Each of the four perspectival systems relies on cognitive simulation—also known as *representation* (Karmiloff-Smith, 1982) and *redintegration* (Horowitz & Prytulak, 1969)—to construct meaning by reenacting the sensory and motoric experiences. Let us take a brief glance at how perspective taking works in each of these four perspectival systems.

In the first perspectival system, language and cognition relate to individual objects and actions through affordances. When we think of an object like a banana, we think of it in terms of its colors, textures, and odors. All of these properties provide us with intimate affordances regarding this object. The perspective we assume when we evaluate the affordances provided by the banana is invariably the perspective of our own body as it acts on the banana.

The second perspectival system is the set of competing spatio-temporal reference frames. The three alternative spatial frames are an object-centered frame, a speaker-centered frame, and an environment-centered frame. Temporal relations are also perceived through three analogous frames. Shifting between these frames involves competition and cooperation between perspectives.

The third perspectival system is the one that is most centrally involved in the emergence of grammar. This is the system of causal action frames that allows us to understand the action of a verb from the perspective of the subject in nominative-accusative language or the focus in ergative-absolutive languages.

The fourth perspectival system is the one that allows us to adopt the social and cognitive perspectives of other human beings. In terms of its linguistic reflexes, this system supports the use of discourse devices such as anaphora, deixis, aspect, conjunction, and backgrounding. Perspective switching on this highest level places heavy demands on working memory, planning, strategy selection, and social referencing.

Three Examples

To get a sense of the ways in which perspectives can be instantiated and modified in discourse, consider Sentence (1):

1. As far as the eye could see, stalks of corn were bending in waves under the battering force of a surging curtain of rain.

If we listen to this sentence in a fairly passive way, we may extract a vague picture of strong rain coming down on a large field of corn. However, if we take a more embodied stance, it allows us to "get into" the meaning of the sentence. To do this, we first assume the perspective of an *eye* that scans the full distance from the foreground out to the end of vision at the horizon. This scanning sets up a spatial frame for grounding *as far as the eye can see.* We then shift perspective to the *stalks of corn.* After constructing *stalks* as an initial perspective, we hear the word *corn.* At this point, we see *stalks* as a distributive figure located against the general ground of *corn.* From this distributive perspective, we trace the bending across repeated *waves* or rows of corn. As we begin to do this, we use the word *under* to spawn a secondary causal perspective for the *battering force* that is seen to emerge from a *surging curtain of rain.* All of these shifts in perspective are driven by specific linguistic devices such as *under, of,* and *as far as.* An embodied understanding of this sentence involves a movement across at least four perspectives: eyes, stalks of corn, battering force, and curtain of rain. Because we used the *corn* as our starting point for the main clause, we end up with an understanding that emphasizes the corn as the actor responding to external pressures, rather than an understanding that starts with the rain as the first mover. In this sense, the exact syntactic form we select when speaking constrains the dynamics of the listener's understanding of the flow of perspective in a sentence.

Let us look at a second example of a slightly different type. Sentence (2) allows us to examine ways in which two alternative perspectives are yoked together in terms of a reciprocal interaction.

2. The harder you try to clamp the pipe, the more the water spurts out into the room.

To understand this sentence fully, we imagine clamping a pipe by using some tool that we either squeeze or turn. Because the tool is not specified, the exact shape of the action cannot be precisely embodied, but we feel ourselves exerting some type of pressure against the pipe. At the same time, we assume the secondary perspective of a stream of water that shoots out into the room. In order to understand the yoked relation between these two perspectives, we must notch up our hold on the pipe by degrees and imagine corresponding increases in the stream of water. In fact, one such imagining of a notching up of pressure on the pipe and subsequent increase in the water spurt is enough to give us the embodied sense of an ongoing linkage between the two yoked processes. We then simply assume that repeated increases of pressure on the pipe will lead to repeated increases of spurting by the water.

As a third example of the process of embodied representation, let us take a look at how we understand Sentence (3) with a more metaphoric content.

3. Casting furtive glances at the seamstress, he wormed his way into her heart.

To understand this sentence, we begin by embodying its literal meaning. We first take the viewpoint of the implied subject and imagine casting glances at a seamstress. Next we assume the guise of a worm and imagine trying to enter into the seamstress. Of course, we would avoid interpreting this too graphically. The repulsiveness of assuming the guise of a worm triggers selection of a more metaphorical interpretation for *worming his way into her heart.* In this metaphorical interpretation, the suitor merely acts like a guileful person who is trying to get emotionally closer all the time and the seamstress accepts these advances, allowing the suitor to enter into her affections. However, the juxtaposition of the figurative and literal interpretation gives this metaphor its unique flavor.

The shape of human language is strongly influenced by the way in which perspective promotes the extraction of embodied meanings. To put this more succinctly, we can say that language emerges from embodiment. The grammatical systems that mark functions such as tense, transitivity, deixis, aspect, and agency have as their sole purpose the elaboration of embodied

representations through perspective taking. Although languages vary widely in the ways they mark these basic functions, the need to mark these relations is universal. When children learn language, they use perspectival relations (MacWhinney & Bates, 1989) as a key to acquiring lexical and grammatical forms. They do this by focusing on activities that correspond to their own embodied perspectives (Huttenlocher, Smiley, & Charney, 1983).

Having sketched out the basics of the impact of perspective taking on the emergence of language, we next turn to a more detailed look at the four levels of perspective taking: affordances, spatial reference systems, causal action, and social referencing. We begin with an examination of the role of perspective taking in organizing affordances.

AFFORDANCES

Affordances (Gibson, 1966) are sensations that we experience when we interact with individual objects. When we grab a banana, our hands experience the texture of the banana 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 affordances are coordinated with visual affordances 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 affordances involving the action of peeling, as well as the peel itself. With the peel removed, we can access new affordances from the meat of the banana. When we eat a banana, our whole body becomes involved in chewing, swallowing, and digestion. All of these affordances in vision, smell, taste, touch, skeletal postures, haptic actions, and even locomotion are provided by a single object that we categorize as a "banana." It is this rich and diverse set of affordances 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. Our first line understanding of the word *banana* is grounded not on these declarative facts, but on affordances. When we hear the word *banana*, each of these affordances becomes potentially activated. The visual affordances or images may be the quickest to receive activation. If the sentence requires nothing more, this may be all that we experience. However, just activating the raw visual image is enough to enable embodied processing of the word *banana*.

Affordances are thoroughly grounded in both the motor and the sensory systems. The perspective that we adopt to understand these individual words

is one that reactivates our normal, personal encounters with these objects. These encounters involve both motoric actions and sensory perceptions. When we hear the word *banana*, we activate neural pathways that are involved in our nonfictive interactions with real bananas. In this sense, understanding of the meaning of an object involves running a "cognitive simulation" of our interactions with that object in terms of its most salient affordances.

Sometimes languages reflect affordances directly in their names for things. For example, in Navajo, a chair is *bikáá'dah'asdáhí* or "on-it-one-sits." In this example, the object is being characterized primarily in terms of the actions it affords. Or to take a more familiar example, many languages refer to a corkscrew as a "cork puller." Here, again, the object is being characterized in terms of the action that it affords. 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 effective as definitions in terms of 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 on 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."

Verbs and adjectives also provide affordances. 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, pressure on the knees and other joints, and the sense of our weight coming down on the earth. Although all of these affordances are eventually available, only the tip of this iceberg of is activated unless the sentence calls for the activation of the full set. Adjectives and adverbs also activate affordances, but only in consort with accompanying nouns and verbs. Consider the combination of the adjective *red* with different nouns. A red tomato is redder than is red lettuce. Squeezing a towel affords different sensations from squeezing a sugar cube. Pointing a football toward something involves different orientations and affordances from pointing a tennis ball toward something. The various affordances provided by these words interact through a system of competitive interactions and polysemic processes, as discussed in MacWhinney (1989).

Children tend to learn new words by matching up their concepts to the affordances provided by these words. Lise Menn (personal communication, 1997) observed her son looking at a bird and declaring "have no hands." It appears that the child was relating to the bird by assuming its perspective and this led immediately to the observation that the bird had no hands. Similarly, Marilyn Shatz (personal communication, 1997) 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 new animate objects by mapping them onto their own human perspective. This mode of apperception can also extend to nonanimate agents, as we take an embodied approach to understanding the shapes, postures, and positions of trees, cars, house, and even waterfalls (Werner & Kaplan, 1963).

Languages often directly reflect the embodied nature of object name affordances. 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.

Psychologists have often noted that the compatibility between ideas tends to facilitate recall and recognition. Sometimes this compatibility also provides evidence for the construction of embodied representations. For example, Klatzky, Pellegrino, McCloskey, and Doherty (1989) asked participants 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 names that matched the action, such as *pinch* for *aim a dart*, these judgments were facilitated. It appears that generating the appropriate hand shape facilitated construction of the representation. This occurs because the affordances provided by a word like *dart* match up well with the hand-shape cues given by Klatzky et al. Affordances can also involve emotional and affectual attitudes. Ertel and Bloemer (1975) asked participants to verify sentences that sometimes contained negative elements. They found that judgments for negative sentences were facilitated when participants were separating blocks, rather than putting them together. Similar findings for a relation between physical states such as smiling or frowning and affective judgments about people mentioned in sentences were reported by Berkowitz and Trocolli (1990).

This view of perception as requiring active motoric involvement is supported by a wide variety of neurological and physiological findings. Psychophysiologists have often noted a general match between emotions, thoughts, and efferent responses (Cuthbert, Vrana, & Bradley, 1991). For example, when we imagine performing bicep curls, there are discharges to the biceps. Similarly, when we imagine eating, there is an increase in salivation. However, the precise match between such efferent discharge and particular linguistic structures has not been delineated using these techniques. For example, studies of galvanic skin response (GSR) and salivation rate have not been able to tell us how people understand words like *stab* or *lettuce*.

SPACE, TIME, AND MOTION

The second major perspective-taking system is the one that deals with position in and through 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 are (1) the ego-centered frame, (2) the object-centered frame, and (3) the environment-centered frame.

The most basic of these three frames is the ego-centered frame, because it encodes the perspective of the speaker. The spatial position of the speaker is given by the deictic term *here* and locations away from the speaker are given by the deictic term *there*. The function of deixis is absolutely fundamental to all perspective taking in space and time, because everything must eventually be referenced to the here and now. The speaker may 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.*

Within the object-centered frame, locations are defined in terms of their position relative to an external object. 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 object-centered 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 object-centered frame. In both ego-centered and object-centered frames, positions are understood relative to a figural perspective that has an orientational field like that of the upright human body.

The use of the ego-centered frame as the basis for the object-centered frame leads to a variety of consequences for both sentence processing and memory. In their classic sentence-picture verification study, Chase and Clark (1972) examined reaction times to sentences like *the star is above the line* or *the star is under the line.* They found that participants were quicker to verify the sentence when it contained an unmarked preposition like *above* instead of a marked preposition like *under.* Clark (1973) noted that unmarked prepositions like *above* or *in front of* tend to reflect orientation to the favored human perspective, whereas marked prepositions like *under* or *behind* deviate from this preferred human perspective. Additional support for the notion of a basic human spatial perspective comes from a

study by Bryant, Tversky, and Franklin (1992) in which participants were asked to memorize spatial layouts from particular perspectives. For example, in the hotel scene, they were told "To your left, you see a shimmering indoor fountain." Each object was located at some point in reference to the observer in the imagined scene. After the scene was stored in memory, the time taken to retrieve a particular object was measured. Participants responded most quickly to objects located on the head–feet axis, followed by the front–back axis, followed by the left–right axis. This ordering of perspective reflects the fact that the head–feet axis is the most fundamental human dimension, followed by front–back and then left–right.

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 object-centered field by our own ego perspective. For example, when having a party in the back 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 ego position and perspective as basic and locating the external object within that ego-centered 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, object-centered reference is best viewed as an extension of ego-centered reference grounded on a shift of perspective from ego to an external object. Both ego-centered and object-centered perspectives are governed by the basic human perspective. All that is involved in moving between ego-centered and object-centered reference frames is explicit tracking of perspective shift in a way that allows the two frames to be active in parallel.

The third spatial reference system, the environment-centered 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, because they are taken as the basis for a full-blown Cartesian coordinate system. The Guugu Yimithirr language in northeast Queensland (Haviland, 1993) 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." Of course, we can use this type of geocentric or environment-centered 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 specific local landmarks in English. For example, we can describe a position as being *50 yards behind the school.* In this case, we are adopting an initial perspective that is determined either by our own ego-centered location (e.g., facing the school) or by the object-centered 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 *50 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 *50 yards* to locate a given object.

In all three reference systems, there is a perspective (ego, object, or reference landmark), a reference object (the school), and a location being specified (the position 50 yards away from the school). As long as we are working in the real world, the shifting of perspective within and between these three frames is not that difficult. However, there is always a certain preference for the ego-centered and object-centered frames over the more difficult environment-centered frame. Children who are learning languages, like Tzotzil, that make extensive use of all three frames tend to begin with ego-centered and object-centered frames and only later acquire environment-centered frames (de Leon, 1994). It makes sense that children learn to make spatial reference by first starting out from their own point of view (Piaget, 1952).

The major tasks involved in spatial processing are choice of a reference frame, assignment of position within a frame, shifting between reference frames, and managing competing reference frames. Language use places particularly high demands on frame shifting and integration. Consider a sentence like *I found a ring over there under the bench east of the swimming pool.* The phrase *over there* invokes an ego-centered reference frame that points the listener's attention to a position distant from either the speaker or the listener. The phrase *under the bench* invokes an object-centered reference frame that locates the position under a bench. However, the position of the bench in the overall field is not yet well determined. The phrase *east of the swimming pool* completes the identification of the location by invoking the environment-centered frame of compass positions. As long as we are clear about the locations of the perspectives involved, these shifts between perspectives are easy to manage.

Asking participants to construct coherent maps of new spatial arrays from sketchy verbal descriptions is a more difficult task. Studies have shown that participants shift between multiple competing frames in accord with task demands (Carlson-Radvansky & Logan, 1997; De Vega, 1994; Franklin, Tversky, & Coon, 1992; Klatzky, Loomis, Beall, Chance, & Golledge, 1998; Maki & Marek, 1997), and that learning to manipulate these competing frames is a skill that develops gradually through the school-age years (Rieser, Garing, & Young, 1994).

Mental Models and Spatial Perspectives

The effects of perspective are not confined to linguistic expression; rather, they also extend to the mental models that we extract from linguistic descriptions. The impact of perspective on mental models can be seen both in the process of constructing internalized models and in the use of these models, once they are constructed. Studies of comprehension often emphasize the online incremental nature of the comprehension process (Hess, Foss, & Carroll, 1995; Marslen-Wilson, 1975; Marslen-Wilson & Tyler, 1980; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). It also appears that incrementality is facilitated when interpretations can be organized about the perspective of the main character or protagonist. The following example passages adapted from Sanford and Moxey (1995) illustrate these effects:

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

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

A propositional analysis does not reveal anything odd about (5), but an embodied perspectival analysis reveals that, when Fred comes to put his mug onto the wallpaper, it is 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 (1995) used passages of this type to argue that textual coherence depends on the construction of full representations of passages.

If we look at other studies in the discourse comprehension literature, we find further evidence that readers use perspective to construct mental models. For example, Murray, Klin, and Meyers (1993) and Keefe and McDaniel (1993) provide evidence that readers tend to follow along closely with the perspective of the protagonist in the passage. In their experiments, subjects may read a sentence like "After standing through the three-hour debate, the tired speaker walked over to his chair." Alternatively, they may read the sentence "The tired speaker moved the chair that was in his way and walked to the podium to continue his three-hour debate." After reading one of these two sentences, subjects were asked to pronounce the visually presented probe word "sat." They were faster at pronouncing the probe when it followed the first sentence, and slower when it followed the second. In the first sentence, the speaker is about to sit down and it makes a lot of sense to pronounce the probe word "sat." In the second sentence, it makes no sense to suddenly have the speaker sitting down when he is actually ready to continue his debate. Glenburg, Meyer, and Linden (1987) report similar results using a probe recognition latency measure.

Morrow, Bower, and Greenspan (1989) asked subjects to read passages describing buildings and the rooms and objects in those buildings. They then read a passage that told about how a protagonist moved through the building. Their task was to decide if particular objects were in particular rooms. It turns out that they were quicker to make this judgment when the rooms and the objects were on the path that the protagonist had taken. The fact that they were quicker in assessing the position of objects directly along the path is consistent with the idea that the path is being encoded from the viewpoint of the imagined protagonist who is touring the imagined building.

O'Brien and Albrecht (1992) gave subjects sentences to read, such as "As Kim stood outside the health club, she felt a little sluggish." Having read this, subjects would then be given the sentence, "She decided to go outside. . . ." Because the continuation is not congruent with the previous spatial arrangement, subjects had trouble reading these continuation sentences.

These experiments in text comprehension have shown that mental models are constructed from the perspective of the protagonist. Relations between objects that lie outside of the main path followed by the protagonist are not as fully encoded as relations between on-path objects and the movements of the protagonist. Embodied representations include not only spatial relations, but also information about the body position of the protagonist (Keefe & McDaniel, 1993), orientations of objects (Sanford & Moxey, 1995), shapes of objects (Klatzky et al., 1989), and other affordances. In addition to information about spatial perspectives and affordances, embodied representations also include information organized around causal action perspectives and social reference perspectives, as we discuss in the next two sections.

Tense, Aspect, and Modality

The processing of temporal relations through adverbials, tense-aspect marking, and modality parallels the processing of spatial relations. Processing of the verb features of tense, aspect, and modality involves movement through the worlds of time, fictive action, and social obligation. Our movements through these worlds are all conducted from a specified perspective that matches up with the aspect or tense given in the sentence. As in the case of spatial processing, the initial basis for temporal deixis is the time of the speech act in which the ego is present. Often, we need to deal with shifts and splits in perspective across time. Vendler (1957) showed how tense can encode three different temporal perspectives: speaking time, action time, and reference time, which are parallel to ego-centered, object-centered, and environment-centered spatial frames, respectively.

Several recent experimental studies have shown that listeners use tense-aspect markings to add richness to the mental models they construct. Earlier, we saw how listeners use spatial relations to construct mental models that embody alternative perspectives. This same principle also extends to the processing of temporal relations through aspects and tense. For example, Carrieras, Carriedo, Alonso, and Fernández (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 reinstantiate that secondary perspective.

CAUSAL ACTION CHAINS

The two levels of perspectival organization we have discussed so far provide grounding for many of the basic units of language. Affordances ground individual open-class words such as *banana*, *warm*, and *run*. Spatial relations ground closed-class words such as *now*, *behind*, and *until*. Spatial relations also 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 rich enough relational system to understand the ways in which objects act on other objects. In particular, a major problem facing both language and cognition is the task of understanding who has acted on what in a causal action chain. Language provides a separate series of devices for solving this problem.

Intransitive Verbs

Predications using intransitive verbs constitute the lowest level building blocks from which causal action chains are constructed. Intransitives 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. In this way, we can treat intransitive verbs as a rich set of embodied affordances. Unlike the affordances provided by nouns, we are not evaluating our actions on an external object. Instead, with 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. When verbs like *twist* or *spread* are used transitively, the basic embodied perspective is still that of the object that is twisting or spreading.

Transitivity Systems

Causal action chains arise from the linking of intransitive activities first into transitive descriptions that are then chained together to form longer narratives. Consider a simple transitive sentence like *The farmer grew the corn*. In this relation, the basic intransitive process of growing is evaluated from the perspective of the corn. However, in terms of causation, it is the farmer who acts on the corn and makes it grow. The farmer plays the role of the agent and the corn plays the role of the patient. When describing relations of this type, languages have to decide whether to focus on the external causal actor or the more directly embodied patient. Nominative-accusative languages, like English, place focus on the actor by treating it as the perspective for the clause. In these languages, the grammatical role of subject is tightly linked to the function of perspective taking. Even in a passive sentence, like *The corn was grown by the farmer*, the subject still marks the farmer as the initial causor. In this case, however, the sentence

is understood from the perspective of the patient (the corn), rather than the agent (the farmer).

In ergative-absolutive languages, like Basque or Djirbal, the primary focus is on the participant undergoing 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. Ergative-absolutive languages place a narrow, close focus on process and leave causation as a secondary fact, which is then assigned to the ergative case role. Du Bois (1987) noted that this tendency may be supported by the fact that the actors in transitive sentences are often omitted or pronominalized in causal discourse. 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 a nominative-accusative language like English can illustrate occasional ergativity effects, languages like Hindi or Samoan can illustrate incomplete or "split" ergativity (Delancey, 1981; Silverstein, 1976). 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 tense, 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 tense. 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 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 causor 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 Australian languages and many languages in 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. Other factors that can lead to splits in ergative marking include inferential markers and certain discourse structures.

Ergative marking can also be used to mark intentionality. Delancey (1981) described this for the Caucasian language Batsbi, which uses ergative case for the subject of a sentence like (6) when the falling is intentional and absolutive marking of the subject when the falling is unintentional.

6. Bill fell to the ground.

This use of alternative constructions to mark intentionality closely resembles a similar use of the passive in English.

Variations in transitivity can also be used as a way of shifting between various causal perspectives in discourse. In an example like (7), perspective is initially assigned to the first noun (*car*) as starting point.

7. 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. Studies by Sachs (1967) and Lombardi and Potter (1992) 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 its full form (Keenan, MacWhinney, & Mayhew, 1977). 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).

Packaging and Conflation

Individual clauses are the basic links in causal action 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:

8. The beam fell.
9. The beam fell when the crane operator released a lever.
10. The crane dropped the beam.

11. The crane operator released the beam.
12. The crane operator pulled a lever to release the beam.
13. 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 (8), we restrict our perspective entirely to *the beam*. If we select (9), we take *beam* as our first perspective, but then add *the crane operator* as a secondary perspective. In (10) through (13), 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:

14. The lion pushed the giraffe, and the giraffe bumped into the table.
15. The lion hit the giraffe and it bumped into the table.
16. The lion bumped the giraffe into the cube.
17. The lion bumped the giraffe, sending it flying toward the table.

In (14), 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 that bumps into the table. Sentence (15) has the same structure, but the pronoun *it* refers ambiguously to either the initial perspective of *lion* or the new perspective of *giraffe*. In (16) and (17), 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 (18):

18. 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 (16), 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:

19. The light bulb flashed.
20. The light bulb flashed until morning.

In Example (19), the default reading is that the light bulb flashed no more than once or twice. However, in (20) 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.

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 nonanimate, it still carries the full force of a causal perspective. Sentence (21) illustrates this effect:

21. The library contains three major collections.

Here, the library is viewed as an agent that actively holds collections of books. As Talmy (1988) noted, this is only a fictive agency, because 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 Sentences (22) and (23).

22. She walked down through the cornfields, out to the river.
23. The path winds down through the cornfields, out to the river.

In (22), the human agent moves over a real path. In (23), on the other hand, the path moves fictively over the same path. Similarly, in (25) 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.

24. The carpenter holds the four legs firmly against the center pedestal.
25. 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 (26) the activities of the initial perspective (Tim) trigger a series of activities in the secondary perspective (Mary).

26. 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 nonactions. Although this causation is on the social level, we apply a basic physical causal model to even these social effects.

C-Command and Starting Points

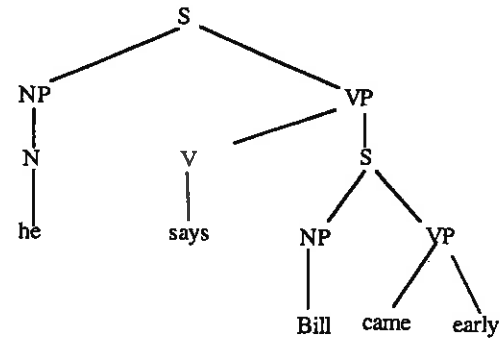
Perspective taking in causal action chains impacts certain key aspects of the grammar of pronominal coreference. This effect results from a basic fact about language use, which is that starting points must be fully referential (MacWhinney, 1977). Gernsbacher (1990) discussed this requirement in terms of her theory of "structure building," which holds that the incremental nature of sentence comprehension requires the starting point to be fully identified, because it is the basic building block on which the rest of the interpretation will be constructed. In dozens of psycholinguistic investigations, Gernsbacher has shown that the initial nominal phrase has the "advantage of first mention." This advantage makes it more memorable and more accessible for further meaningful processing. However, when the first noun is not sufficiently referential, the foundation is unclear and the process of comprehension through structure building is thwarted. If the starting point is a nominal, referentiality is not at issue. However, if the starting point is a pronoun, then there must be a procedure for making it referential. One way of doing this is to link the pronoun up to an entity mentioned in the previous discourse. In a sequence like (27), it is easy to link up *he* in the second sentence with *John* from the first sentence, because John has already been established as an available discourse referent. However, in (28) there is nothing to link *he* to and the second sentence seems awkward without this previous link.

27. John was listing the guests at last week's party. He says Bill came, along with Mary and Tom.
28. 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.

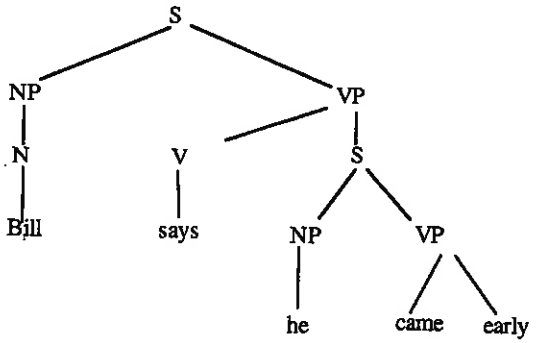
The generative theory of Government and Binding (Chomsky, 1982) treats this same phenomenon in terms of structural relations in a phrase-marker tree. According to this theory, the problem in the second sentence in (28) is that Bill does not "c-command" the pronoun *he* and cannot be coreferential. The phrase marker involved is given in (29):

29.



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* in this tree, we get (30):

30.



Here, *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 (31).

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

This effect is not a simple matter of linear order, because coreference between a pronoun and a following noun is perfectly good when the pronoun is in an initial subordinate clause. Consider the contrast between (32) and (33), where the asterisk on (33) indicates that *he* cannot be coreferential with *Lester*.

32. After he drank the vodka, Lester started to feel dizzy.

33. *He started to feel dizzy, after Lester drank the vodka.

Contrasts of this type create problems for the simplest versions of the formalist approach, because they involve identical structural relations. However, they follow immediately from the theory of perspective taking, because a preposed subordinate clause is not a main clause and the process of structure building only requires referentiality for the subject of the main clause. In effect, subordinate conjunctions like *after* in (32) work as cues to place the following material on hold for structure building, until the main clause is encountered. Some additional examples of these effects are given in (34) through (39).

34. *She found a snake near Sue.
35. *Near Sue, she found a snake.
36. *She denied that Martha was a robber.
37. *She liked some of Mary's dates.
38. *He was adored by the students who studied with John.
39. *I think she found a snake near Sue.

In all these examples, clause-initial pronouns that are not protected by placement into an adverbial phrase must be referential so they can serve as the bases for structure building.

The same principle that requires that subjects be referential also applies in a somewhat weakened form to the direct and indirect objects of verbs, as illustrated in (40) through (42).

40. *John told him that Bill was crazy.
41. *I'm willing to give him fifty dollars for Ben's bike.
42. *Him, John's mother likes.

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

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

By the time we reach elements that are no longer in the main clause, as in (45), coreference back to the main clause is not blocked, because elements in a subordinate clause are not crucial perspectives for the structure-building process.

45. 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 nonreferential perspectives comes from sentences with special aspect markings, as in (46) and (47).

46. She had just gotten back from vacation, when Mary saw the stack of unopened mail piled up at her front door.
47. *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 (46), 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. Just as relevance markers can increase the openness of a main clause pronoun to coreference, so the openness of a subordinate clause noun for coreference can be decreased by indefinite marking, as in (49).

48. While Ruth argued with the man, he cooked dinner.
49. ?While Ruth argued with a man, he cooked dinner.
50. While Ruth was arguing with a man, he was cooking dinner.

The addition of an aspectual marker of current relevance in (50) overcomes the effect of indefiniteness in (49), 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 pattern 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 (51), the initial *wh*-word *who* indicates the presence of information that needs to be identified.

51. *Who does he hate most?

In this case, it is the pronoun *he*, rather than the initial word *who*, that serves as the starting point for structure building. Because 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 (52) through (56).

52. Who is hated by his brother most?
53. Who thought that Mary loved him?
54. Who hates his mother most?
55. Who said Mary kissed him?
56. Who hates himself most?

In these examples, the *wh*-word can be coreferent with noncentral components, such as objects and elements from embedded clauses. Only coreference with subjects, as in (51), 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) presented 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.

Relative Clauses, Conjoined Clauses, and Possessives

Perspective taking also has an important effect on the grammar and processing of various forms of syntactic embedding and conjunction. Let us first look at the impact of perspective on relative clause interpretation. The account 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 th perspective of the relative clause. In the SO relative clause type, there i a double perspective shift. Perspective begins with the main clause subjec (*dog*). When the next noun (*cat*) is encountered, perspective shifts once However, at the second verb (*kicked*), perspective has to shift back to th initial perspective (*dog*) to complete the construction of the interpretation Sentences that have further embeddings have even more switches. Fo example, Sentence (61) has six perspective switches.

61. The dog the cat the boy liked chased snarled. 6 switches

Sentences that have as much perspective switching as (61) without addi tional lexical or pragmatic support are basically incomprehensible, at leas at first hearing.

Studies of the acquisition of relative clauses by children largely suppor the order of difficulty predicted by the perspective-taking account (Mac Whinney, 1982). This predicted order is: SS > OO = OS > SO. This orde appears to predict results across a wide variety of experimental paradigm including imitation, comprehension, and sentence memory. In addition a study of online sentence processing effects in Hungarian (MacWhinne & Pléh, 1988) with adult participants further supported a role for the perspective account. However, there was also evidence in Hungarian fo the importance of additional parallel structure effects. In Hungarian, al six orders of the subject, object, and verb are grammatical. In three o these orders (SOV, SVO, and VSO), the subject is the topic; in three othe 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 (62), 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.

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 showed that

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 (52) through (56).

52. Who is hated by his brother most?
53. Who thought that Mary loved him?
54. Who hates his mother most?
55. Who said Mary kissed him?
56. Who hates himself most?

In these examples, the *wh*-word can be coreferent with noncentral components, such as objects and elements from embedded clauses. Only coreference with subjects, as in (51), 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) presented 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.

Relative Clauses, Conjoined Clauses, and Possessives

Perspective taking also has an important effect on the grammar and processing of various forms of syntactic embedding and conjunction. Let us first look at the impact of perspective on relative clause interpretation. The account 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. Sentences that have further embeddings have even more switches. For example, Sentence (61) has six perspective switches.

61. The dog the cat the boy liked chased snarled. 6 switches

Sentences that have as much perspective switching as (61) without additional lexical or pragmatic support are basically incomprehensible, at least 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 participants 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 (62), 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.

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 showed that

young children find it easier to imitate a sentence like (64), as opposed to one like (65).

64. Mary cooked the meal and ate the bread.
65. Mary cooked and John ate the bread.

In (64), there is no perspective shift, because the perspective of Mary is maintained throughout. In (65), on the other hand, perspective shifts from Mary to John. Moreover, in order to find out what Mary is cooking, we have to maintain both the perspective of Mary and John until the end of the sentence.

We can distinguish structures that require the maintenance of multiple perspectives from those that simply require repeated perspective shifting. Sentence (66) illustrates how the possessive construction can require repeated perspective shifting.

66. 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.

Retracing Perspectives

The perspective-taking process also influences ways in which sentences are retraced or reformulated 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 (67) and another showed (68).

67. A cat gives flowers to a bunny.
68. A cat gives flowers to a boy.

MacWhinney (1977) found that, for pictures like these, participants sometimes produced retraces like (69) or (70), but never produced retraces like (71) or (72).

69. A bu # a kitty's giving a flower to a bunny.
70. A boy # the cat's giving a boy a flower.
71. A ca # a bunny gets flowers from a cat.
72. 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, 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.

In this section, we examined various ways in which perspective taking affects grammar and sentence processing on the level of the clause. This discussion passed over many other areas of clausal grammar where perspective taking has a similarly important impact. These additional areas include quantifier scope, word order, and attachment. A fuller account would describe how grammar emerges from perspective in each of these domains. For now, however, we move on to an examination of the fourth and highest level of perspectival organization.

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, propositional 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.

It seems unimaginable that the complexity of human society could ever emerge without support from linguistic expression, and herein lies a so-

lution to understanding the great mystery of language evolution. 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 (1960, 1975) and others, Vygotsky (1962) 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 on 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.

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.

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 (73) illustrates the use of the experiencer-stimulus verb *admire,* whereas Sentence (74) illustrates the use of a stimulus-experiencer verb like *apologize.*

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

McDonald and MacWhinney (1995) asked participants to listen to sentences like (73) and (74) 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 (74) 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 (73) 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 online is good evidence in support of the perspective-taking account of sentence processing.

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 of presuppositions and perspective. Consider the conjunctions *but* and *although* in sentences like (75) and (76).

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

To understand (75), we have to figure out why Mary's winning of the race would weaken her allegiance to Helen. To understand (76), 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 (77) illustrates this.

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

To extract the meaning of (77), 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.

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 (78), 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.

78. 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, as these form the backbone of a great deal of oral communication.

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, 1994). 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 (79) illustrates the type of embedded representations computed through a theory of mind.

79. 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.

Contrasting Perspectives

The various social conventions and forms we 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 (80) from Fauconnier and Turner (1996). In Example (80), a contemporary philosopher is imagining a dialog with Kant.

80. 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 *Critique of Pure Reason*, that only innate ideas have power. But I say to that, what about neuronal group selection? And he gives no answer.

Fauconnier and Turner noted 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 (80) 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 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.

Multifocal Chains

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 traveling 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 or-

ganization 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 by 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 on them.

MERGING THE FOUR LEVELS

This chapter developed a view of language–thought relations that emphasizes the construction of a human perspective across four major cognitive systems: affordances, spatio-temporal frames, causal action chains, and social frames. Each of these systems establishes a partial cognitive reflection of the entire human being. The affordance system internalizes and adapts to the ways in which humans act on the world using sensation and action. Spatio-temporal frames internalize our mental models of positions, moments, and movements in the world. Causal action chains allow us to encode the activities of the world in terms of our own causative perspective. Social frames allow us to view actions in terms of their personal consequences and implications.

Operating by themselves, these four systems would not give rise to the unitary experience of human consciousness. Without language, our minds would remain prisoners of a certain internal modularity (Fodor, 1983). It is language that provides the real-time, dynamic, symbolic links that merge these four separate perspectives into the integrated human perspective we call consciousness. Language, both in its social form and in the guise of

inner speech, links these four separate frames into a functional neural circuit that embodies a complete mental homunculus. It is this complete perspective-taking system that we use to solve scientific problems, form narrations, and develop social relations.

In Examples (1) through (3), at the beginning of this chapter, we examined ways in which language could express interacting and switching perspectives. Let us now consider a more extended example of how language works to blend together information from these four separate systems. This example, given in (83), comes from an Associated Press release of May 20, 1997.

83. A cyclone hammered the Bangladesh coast Monday with the force of "hundreds of demons" leveling entire villages of mud and thatch huts, flooding crops, and killing at least six people.

    Three men and two children were crushed under collapsed buildings or hit by flying pieces of tin roofs in the southern port of Chittagong. One man died in Teknaf, about 110 miles down the coast, when he was blown off his roof, while trying to secure it.

    The storm roared in from the Bay of Bengal with wind gusts of 125 mph, forcing a half-million people to flee their huts and huddle in concrete shelters. Many power and telephone lines were down, so a full account of casualties and damage was not available.

To comprehend this passage fully, we first assume the perspective of the cyclone hammering the coast. We know that storms do not use literal hammers to beat down on the land, but we sense the driving nature of the cyclone pounding the coast. The affordances of hammers, pounding, and rain are fairly clear. We then convert our image of the cyclone to that of hundreds of demons who are now pounding the coast in the guise of a storm. Further concretizing our vision, we now see this demon storm leveling and killing. The perspective now shifts to the people who are being killed. Here the article uses the split perspective of the passive to focus the people against the background of the cyclone and flying roofs. We learn about a man who died in Teknaf. At first his perspective is a passive one (blown off his roof), but then we see him play a more active role (trying to secure it). We also begin to see a shift between spatial perspectives with movements from Chittagong and Teknaf back to the Bay of Bengal. Finally, we shift from the fact that the storm has downed the power lines back to the overall perspective of the writer of the press release who is explaining to us that conditions on the ground made it difficult to write a complete account of this event. Together, these various shifts of perspective give this short press release a rich, dynamic quality that allows us to partially understand key aspects of the catastrophe. This use of dy-

namic perspective taking and perspective switching substantially enriches our ability to form rich interpretations that support the acquisition of this new information.

NEUROPHYSIOLOGICAL IMPLICATIONS

The claim being made is that the human brain has evolved in a way that allows it to run a high-level simulation of the human body and its positioning in the spatial, social, and causal world. It is this continually running cognitive simulation that constitutes human consciousness. Moreover, language, particularly in the form of inner speech (Sokolov, 1972; Vygotsky, 1962), plays a pivotal role in supporting this continually running perspectival simulation. Language does this by coordinating sensory and motoric systems in posterior brain areas with attentional and planning systems in frontal areas. By continually accessing and refreshing posterior areas, frontal areas allow us to interpret and anticipate experiences in terms of basic affordances. In addition, the frontal cortex is specifically adapted to support dynamic integration of the four perspectival levels we discussed. In particular, frontal cortex has separate mechanisms for refreshing and integrating affordances, spatial referencing, causal action chains, and social frames. In this section, we examine evidence from cognitive neuropsychology that supports this view of cortical processing.

Control of Affordances

A variety of evidence indicates that frontal cortex works together with posterior cortex in a perception–action chain (Neisser, 1976) that allows us to process affordances. 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 matches up with others that suggest that, even on the levels of affordances and spatial frame processing, perspective switching is controlled by frontal structures that associate perception with action.

In the last few years, imaging studies have provided additional evidence regarding the control of affordance processing. Studies by Parsons et al. (1995), Martin, Wiggs, Ungerleider, and Haxby (1996), and Cohen et al. (1996) showed that when participants are asked to engage in mental imagery, they use modality-specific cortical systems. There is growing evi-

dence for 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). Studies using functional magnetic imaging resonance (fMRI) technology have shown that left inferior prefrontal cortex (LIPC) is activated for initial presentations of words, but not for repeated presentations (Demb et al., 1995; Gabrieli et al., 1996). These findings call into question the traditional view of language processing, which locates semantic processing exclusively in posterior errors such as the inferior parietal or the superior temporal gyrus. In addition, lesion studies (Gainotti, Silveri, Daniele, & Giustolisi, 1995), positron emission tomography (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 the meanings that are eventually expressed by words in general and verbs in particular. Within the current framework, this process is best understood as involving the active generation of a motoric perspective that is compatible with the affordances of particular objects. For example, in the verb generation paradigm, the participant sees a picture of a chair. The task is to activate a verb that is appropriate for this object. To do this, the participant must utilize the affordances of the chair to activate a motoric perspective that is then used to activate a verbal label like *sit*. This interaction between affordances and actions is controlled by areas of the frontal lobes that generate action plans and perspectives.

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). Goodale noted that patients with lesions to the dorsal "where" 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 on. 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.

Control of Spatial Perspective

Studies of neurological patients suffering from visual neglect provide striking support for the role of perspective in both affordance and spatial level processing. Behrmann and Tipper (1998) looked at patients with right parietal lesions who showed neglect in the left visual field. When asked to

copy a picture with three flowers in flower pots, these participants would typically draw the right flower pot accurately and have problems with the left flower pot. However, when they were then asked to copy a picture of a single flower in a flower pot, they were able to accurately copy one side of the flower but would have trouble with copying the side of the flower on their neglected side. Studies of this type have shown that neglect is based not on absolute properties of the visual field, but on the representation of objects cognitively. This level of field independence is apparently based on primarily parietal mechanisms, or at least connections between parietal and frontal cortex, because the lesions involved are in the parietal lobe. However, there is also evidence for an interaction between frontal and parietal areas in the control of spatial perspective. 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 lend further weight to the idea that the prefrontal visual area works together with parietal areas to facilitate object-centered processing of affordances and spatial representations.

Control of Action Chains

It has long been suspected that parts of inferior frontal cortex around Broca's area play an important role in controlling action sequences (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 perspective theory, it would be easy to suspect that these frontal areas are particularly involved in the construction and support of causal action chains. In fact, recent fMRI work has linked the processing of syntactically complex sentences to Broca's area (Geschwind, 1965; Just, Carpenter, Keller, Eddy, & Thulborn, 1996).

Because language functions simultaneously on so many levels, it is not surprising to find that several frontal areas must all work in concert during language processing. The rich pattern of interconnectivity between frontal areas and from frontal areas to posterior, thalamic, and cingulate areas (Fuster, 1989; Kolb & Whishaw, 1995) underscores the extent to which the frontal system works to integrate a variety of mental facilities, including attention (Cohen & Bookheimer, 1994), memory (Shimamura, Janowsky, & Squire, 1990), inhibition, motor planning, and goal formation, all in the service of perspective taking. Mesulam (1990, p. 610) asked "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 perspectival account suggests that the answer to this question lies in the fact that the frontal cortex is not only attempting to integrate perspective across four levels of cognition but also to support the emergent frontal homunculus through access to memory, attention, and systems for inhibiting prepotent responses.

The level of frontal integration we described so far is available in our primate cousins, albeit in a somewhat less elaborated fashion (Wilkins & Wakefield, 1995). What is unique to man is the linkage of this rich frontal system to language. Both 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 and the linguistic forms it controls provide an additional bridge between the control architecture of frontal cortex and the affordance and spatial processing of posterior cortex. More important, language links us to our social world and allows us to share in the perspectives of others. These processes of verbal sharing encourage the formation of inner speech that serves to progressively knit together the four levels of perspective taking.

HOW CAN THIS ACCOUNT BE ELABORATED?

This account is just a hypothesis. Although there are hundreds of pieces of linguistic, cognitive, and neurological evidence pointing toward the importance of embodiment and perspective as the central organizing principles of the mind, the current version of the perspective hypothesis needs to be elaborated in much greater detail before it can really be tested. But exactly how should this general claim be cashed out in terms of a specific mechanistic model? Ideally, it would be nice to have a full, simulated model of the human brain. However, evolution has devoted several hundred million years to crafting the basic neural structure and at least another 4 million in building up the specific human adaptations that support language. It may take us a while to catch up with all this handiwork, using our best digital computers. However, we can already begin to see how the notion of embodiment has begun to illuminate work in Artificial Intelligence (Brooks, 1991; Feldman et al., 1996; Harnad, 1990, 1995; Regier, 1996). While this simulation work is progressing, there are several areas in which the empirical claims of the perspective hypothesis can be further elaborated.

From the viewpoint of psycholinguistics, the perspective hypothesis generates many important predictions. Whenever a structure shifts to a marked perspective or forces integration of competing perspectives, it should be difficult to produce, comprehend, imitate, and recall. In fact, this type of prediction has already received extensive support, but it needs to be tested out now against the full range of linguistic structures and psycholinguistic

tasks. Not all effects in psycholinguistics can be attributed to perspective, but its influence is pervasive enough to require a complete reexamination of the experimental literature in this new light. Currently, there is no reasonably complete model of sentence processing that properly incorporates perspective-taking effects. However, models developed by Gernsbacher (1990), MacDonald, Pearlmutter, and Seidenberg (1994), and MacWhinney and Bates (1989) could all be adapted in ways that would foreground the impact of perspective on sentence processing.

From the viewpoint of linguistics, the perspective hypothesis unifies a great deal of thinking about functional pressures on language form by Fauconnier, Givon, Kuno, Langacker, Talmy, and others. A theory grounded on perspective and embodiment can provide new motivation for typological theories of language universals as they affect lexical, grammatical, and discourse structure. The individual theories of aspect, pronominal coreference, reflexivization, transitivity, conjunction, and relativization with which linguists are currently working could all be restructured to deal specifically with perspective as a force unifying language and cognition. In this way, linguistics would be drawn more deeply into association with the whole of cognitive science.

From the viewpoint of developmental psychology, the role of perspective in organizing learning reawakens the attention to issues originally debated by Piaget and Vygotsky. Both Piaget (1959) and Vygotsky (1962) recognized the importance of the child's perspective for grounding cognitive development. Piaget viewed this in terms of an egocentric perspective, whereas Vygotsky emphasized social influences on inner speech. The perspective hypothesis views these two approaches as fully compatible. In fact, our understanding of cognitive development would be incomplete without emphasizing both views. Other researchers such as Dewey (1933), Huttenlocher and Presson (1973), Case (1997), and Montessori (1913) emphasized the importance of "learning by doing." These linkages between perception and action are closely compatible with the perspective hypothesis.

From the viewpoint of cognitive psychology, the perspective account matches up well with a number of current research trends. Several years ago, Kolers and Roediger (1984) argued that memories are shaped by the action of encoding. Since then, researchers have increasingly emphasized the impact of embodiment on both memory and categorization. Glenberg's (1997) approach is particularly close to the one developed here. One of Glenberg's important contributions is the notion of "mesh," which emphasizes the extent to which default properties stored in memory interact with new embodied perceptions to give rise to subjective experience and new learning. Glenberg's ideas about mesh match up well with models of schema application in the connectionist framework (McClelland, St. John, & Taraban, 1989; Rumelhart, Smolensky, McClelland, & Hinton, 1986).

By modeling interactions of memory with current embodied perception, cognitive psychologists can make good use of the concept of perspective to generate new experiments and more detailed hypotheses. Within the general area of cognitive psychology, there are further applications of perspective theory to both spatial representation theory and problem-solving theory. In spatial representation theory, there is already a strong tendency to view cognition in terms of conflicting and converging spatial perspectives. The current approach simply rearticulates ideas already common in that field. However, in the field of problem solving, the role of perspective remains largely unexplored. For studies of man–machine interaction, problem representation, and search strategies, the perspective account makes many interesting and testable predictions.

From the viewpoint of social psychology, perspective theory provides an interesting way of rethinking social cognition. Perspective directly expresses basic social concepts such as theory of mind (Flavell & Miller, 1997), social referencing, symbolic interaction (Blumer, 1969), and perspective taking. Rather than viewing social processes in terms of "cold" as opposed to "hot" cognition, perspective theory emphasizes the way in which embodiment unifies hot emotional and instinctual affordances with colder discourse-based structures for social relations. The linkage of social structures to linguistic processes offers a variety of additional ways of conceptualizing the learning and application of roles, rules, and relationship.

Finally, from the viewpoint of cognitive neuropsychology, the implications of the four-level perspectival account of frontal functioning are new and largely unexplored. Through the combined application of imaging techniques such as event-related potentials (ERP) and functional magnetic resonance imaging (fMRI) with data from clinical populations, we can begin the slow process of testing out and elaborating the perspective hypothesis in ways that can possibly reshape our thinking about the human mind. There are also clear implications of this hypothesis for our thinking about the ways in which the human brain developed from the primate brain (Wilkins & Wakefield, 1995). By relating changes in the architecture and connectivity of frontal cortex with posterior and limbic areas to changes across primate species, we can gain a still richer view of the emergence of language from embodied cognition.

REFERENCES

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

Bartsch, K., & Wellman, H. (1994). *Children talk about the mind.* New York: Oxford University Press.

Behrmann, M., & Tipper, S. P. (1998). Attention accesses multiple reference frames: Evidence from visual neglect. *Journal of Experimental Psychology: Human Perception and Performance.*

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

Blumer, H. (1969). *Symbolic interactionism: Perspective and method.* Englewood Cliffs, NJ: Prentice-Hall.

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

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 Associates.

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.

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.* New York: Wiley.

Chase, W., & Clark, H. (1972). Mental operations in the comparison of sentences and pictures. In L. W. Gregg (Ed.), *Cognition in learning and memory* (pp. 205–232). New York: Wiley.

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

Clark, H., & Clark, E. (1977). *Psychology and language: An introduction to psycholinguistics.* New York: Harcourt, Brace, & Jovanovich.

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, M. S., & Bookheimer, S. Y. (1994). Localization of brain function using magnetic resonance imaging. *Trends in Neurosciences, 17*(7), 268–277.

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.

Cuthbert, B. N., Vrana, S. R., & Bradley, M. M. (1991). Imagery: Function and physiology. *Advances in Psychophysiology* (Vol. 4).

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

De Vega, M. (1994). Characters and their perspectives in narratives describing spatial environments. *Psychological Research, 56,* 116–126.

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

Demb, J. B., Desmond, J. E., Wagner, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *Journal of Neuroscience, 15*(9), 5870–5878.

Dewey, J. (1933). *How we think.* Boston: Heath.

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., & Turner, M. (1996). Blending as a central process of grammar. In A. Goldberg (Ed.), *Conceptual structure, discourse, and language* (pp. 113–130). Stanford, CA: Center for the Study of Language and Information.

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. Flores 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. M. Coltheart (Ed.), *Attention and performance VII: The psychology of reading* (Vol. 12, pp. 619–654). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fodor, J. (1983). *The modularity of mind: As essay on faculty psychology.* Cambridge, MA: MIT 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.

Gabrieli, J. D. E., Desmond, J. E., Demb, J. B., Wagner, A. D., Stone, M. V., Vaidya, C. J., & Glover, G. H. (1996). Functional magnetic resonance imaging of semantic memory processes in the frontal lobes. *Psychological Science, 7,* 278–283.

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

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

Geschwind, N. (1965). Disconnection syndromes in animals and men. *Brain, 88,* 585–644.

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

Glenberg, A. (1997). What memory is for. *Behavioral and Brain Sciences, 220,* 1–55.

Glenberg, A. M., Meyer, M., & Linden, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26,* 69–83.

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

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 hierarchically organized sequential behavior. *Behavioral and Brain Sciences, 14,* 531–595.

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

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 Associates.

Haviland, J. B. (1993). Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology, 3,* 3–45.

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,* 27–60.

Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General, 124,* 62–82.

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

Huttenlocher, J., & Presson, C. (1973). Mental rotation and the perspective problem. *Cognitive Psychology, 4,* 277–299.

Huttenlocher, J., Smiley, P., & Charney, R. (1983). Emergence of action categories in the child: Evidence from verb meanings. *Psychological Review, 90,* 72–93.

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.

Karmiloff-Smith, A. (1982). Language as a formal problem space. In W. Deutsch (Ed.), *Child language: Beyond description.* New York: Springer.

Keefe, D., & McDaniel, M. (1993). The time course and durability of predictive inferences. *Journal of Memory and Language, 32,* 446–463.

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.

Kintsch, W. (1974). *The representation of meaning in memory.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Klatzky, R. L., Loomis, J. M., Beall, A. C., Chance, S. S., & Golledge, R. G. (1998). Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychological Science, 9,* 293–298.

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* (4th ed.). New York: Freeman.

Kolers, P. A., & Roediger, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior, 23,* 425–449.

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

Langacker, R. (1989). *Foundations of cognitive grammar. Vol. 2: Applications.* Stanford, CA: Stanford University Press.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

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. 311–354). New York: Academic Press.

Lombardi, L., & Potter, M. (1997). The regeneration of syntax in short term memory. *Journal of Memory and Language, 31,* 713–733.

Luria, A. (1960). Verbal regulation of behavior. In M. A. B. Brazier (Ed.), *The central nervous system and behavior.* New York: Macy Foundation.

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.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676–703.

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 Associates.

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). New York: 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., & Bates, E. (Eds.). (1989). *The crosslinguistic study of sentence processing.* New York: Cambridge University Press.

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.

Marslen-Wilson, W. (1975). Sentence perception as an interactive parallel process. *Science, 189,* 226–227.

Marslen-Wilson, W. D., & Tyler, L. K. T. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1–71.

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., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes, 4,* 287–335.

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 neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology, 28,* 597–613.

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

Montessori, M. (1913). *Pedagogical anthropology.* New York: Frederick A. Stokes.

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

Murray, J., Klin, C., & Myers, J. (1993). Forward inferences in narrative text. *Journal of Memory and Language, 32,* 464–473.

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

O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 777–784.

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.

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. (1959). *The language and thought of the child.* London: Routledge & Kegan Paul.

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.
Regier, T. (1996). *The human semantic potential.* Cambridge, MA: MIT 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.
Rumelhart, D., Smolensky, P., McClelland, J., & Hinton, G. (1986). Schemata and sequential thought processes in PDP models. In J. McClelland & D. Rumelhart (Eds.), *Parallel distributed processing: Vol. 2* (pp. 7–57). Cambridge, MA: MIT Press.
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.* Philadelphia: John Benjamins.
Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Lawrence Erlbaum Associates.
Shimamura, A. P., Janowsky, J. S., & Squire, L. R. (1990). Memory for the temporal order of events in patients with frontal lobe lesions and amnesic patients. *Neuropsychologia, 28*(8), 803–813.
Silverstein, M. (1976). Hierarchy of features and ergativity. In R. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 121–171). Canberra: Australian Institute of Aboriginal Studies.
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 & Winston.
Sokolov, A. (1972). *Inner speech and thought.* New York: Plenum.
Sowa, J. F. (1984). *Conceptual structures: Information processing in mind and machine.* Reading, MA: Addison-Wesley.
Sridhar, S. N. (1988). *Cognition and sentence production: Cross-linguistic study.* New York: Springer Verlag.
Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science, 12,* 59–100.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634.
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.
Vendler, Z. (1957). Verbs and times. *Philosophical Review, 56,* 143–160.
Vygotsky, L. (1962). *Thought and language.* Cambridge, MA: MIT Press.
Werner, H., & Kaplan, B. (1963). *Symbol formation: An organismic-developmental approach to language and the expression of thought.* New York: Wiley.
Wilkins, W. K., & Wakefield, J. (1995). Brain evolution and neurolinguistic preconditions. *Behavioral and Brain Sciences, 18,* 161–226.
Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1196–1207.

Chapter 9

Social Perspectives on the Emergence of Language

Catherine E. Snow
Harvard Graduate School of Education

The term *social* in the title of this chapter is being used in two senses. In its first sense, *social* refers to the capacities of the infant and young child, which I argue are the source out of which language emerges. In its second sense, *social* refers to the context in which language acquisition occurs. Of course, the social support that adults are inclined to provide to language learners is partially a consequence of the social capacities that infants possess—the degree to which infants are socially appealing creatures. In the chapter, then, I present first an argument about the infant's social capabilities and how they provide a context for the emergence of language, and second about the adult's provision of social support to the learner constructing an emergent language system.

THE INFANT'S SOCIAL CAPACITIES

Discussions of the emergence of language must deal with at least two basic questions: "Emergence from what?" and "Emergence with what help?" Any argument that language emerges as a natural product of development, that is, as a phenomenon without a particular and separate developmental path, must specify what the preconditions are for that emergence, or the domains of accomplishment that make the emergence of this powerful new system possible. Once we have settled what language emerges from, we have a stronger basis for hypothesizing under what conditions of environmental support the emergence is likely to be relatively trouble free.