4

The Emergence of Grammar from Perspective

Brian MacWhinney
Carnegie Mellon University

Successful communication rests not just on shared knowledge and reference (H. Clark and Marshall, 1981), but also on a process of mutual perspective taking. By giving clear cues to our listeners about which perspectives they should assume and how they should move from one perspective to the next, we maximize the extent to which they can share our perceptions and ideas. When language is rich in cues for perspective taking and perspective shifting, it awakens the imagination of the listener and leads to successful sharing of ideas, impressions, attitudes, and narratives. When the process of perspective sharing is disrupted by interruptions, monotony, excessive complexity, or lack of shared knowledge, communication can break down.

Although we understand intuitively that perspective taking is central to communication, few psycholinguistic or cognitive models assign it more than a peripheral role. Linguistic theory typically views perspective as a secondary pragmatic filter (Kuno, 1986) that operates only after hard linguistic constraints have been fulfilled. This paper explores the hypothesis that, far from being peripheral or secondary, perspective taking is at the very core of language structure and higher-level cognition. This approach, which I will call the perspective hypothesis, makes the following basic claims:

1. Perspective taking operates online using images created in five systems: direct experience, space/time deixis, plans, social roles, and belief. In this paper, we will explore the first three of these systems, leaving a discussion of the last two to further work.
2. Language uses perspective taking to bind together these five imagery subsystems.
3. Grammar arose as a social convenience to support accurate tracking and switching of perspective.
4. Language comprehension and production use both depictive and enactive imagery. Depictive imagery relies on the ventral image processing system, whereas enactive imagery relies on the dorsal system for perception-action linkages. Perspective taking depends primarily on processing in the dorsal stream.
5. On the level of direct experience, perspective shifting depends on imagery grounded directly on body maps.
6. On the level of deixis in space and time, perspective shifting depends on the projection of the body image across egocentric, allocentric, and geocentric frames.
7. On the level of plans, perspective shifting in the transitivity system assigns roles to referents through the transitivity system. Premotor working memory areas and inferior frontal action planning areas provide the processing capacity to control perspective shifts in action chains.
8. By tracing perspective shifts in language, children are able to learn the cognitive pathways and mental models sanctioned by their culture.
9. The emergence of language as a species-specific human skill depends on a series of gradual evolutionary adaptations (MacWhinney, 2003) that supported perspective taking in the four subsystems, as well as additional adaptations for vocal control.

The perspective hypothesis relies heavily on a series of recent advances in cognitive psychology, cognitive neuroscience, and

cognitive linguistics. In particular, it builds on the following theoretical positions and empirical advances:

1. As Miller and Johnson-Laird (1976) and Fauconnier (1994) have shown, language allows us to construct and describe mental models and mental spaces.
2. As Shank and Abelson (1977) and Rumelhart (1975) have shown, we use mental models to elaborate schemata, frames, and stories in which people have specified social roles.
3. As Zwaan and Radvansky (1998) and Glenberg (1997) have demonstrated, discourse comprehension produces an embodied situational model that instantiates mental spaces and social frames.
4. As Lakoff (1987) and Lakoff and Johnson (1980) have shown, language uses metaphor and extension to represent the body in the mind. In the terms of Feldman et al. (1996), we can say that language produces a cognitive simulation of reality.
5. As Barsalou (1999) and Langacker (1987) have demonstrated, cognition manipulates a system of perceptually grounded symbols. These symbols derive their expressive power from a retrievable (Ballard, Hayhoe, Pook, and Rao, 1997) mapping to direct experience.
6. As Harnad (1990) has argued, the grounding of cognition in a body can solve the symbol-grounding problem (Searle, 1980).
7. As Talmy (2000) has shown, clausal packaging, conflation, and structuring express the ways in which we map our human understanding of force and causation onto the physical and social world.
8. As Holloway (1995), Deacon (1997), Donald (1991), and Dunbar (2000) have suggested, language and cognition have co-evolved across the full six million years of human evolution.
9. As Damasio (1999), Donald (1998), Shallice and Burgess

(1996), MacNeilage (1998) and others have argued, the most recent evolutionary changes have allowed language to link all aspects of cognition through functional neural circuitry.

10. As demonstrated in neuroimaging work (Jeannerod, 1997; Kosslyn, Thompson, Kim, and Alpert, 1995; Osman, Albert, and Heit, 1999; Pulvermüller, 1999), the construction of mental images relies on the same neural pathways that produce direct action and perception.
11. As Vygotsky (1962) and Tomasello (1999) have noted, language facilitates socialization of the child in accord with culturally specific frameworks for cognition.

Each of these positions is supported by a wide range of linguistic, psychological, and biological evidence. Together, these views have yielded a rich picture of the ways in which embodied perceptual symbol systems and situation models support language and cognition. By way of shorthand, I will refer to this emergent consensus as the theory of embodied cognition. Although the theory of embodied cognition has led to important advances in our understanding of the relation between language and cognition, it has not yet provided an account of real-time processes in language comprehension and production. Without such an account, it will be difficult to analyze the grammatical systems of human languages from the viewpoint of embodied cognition. In this chapter, I argue that bridging this gap requires us to extend current situational model theory to deal with the construct of perspective.

The articulation of a theory of perspective is not a minor afterthought in the formulation of the theory of embodied cognition. It forces a fundamental rethinking of the dynamics of mental models, the nature of sentence processing, the functional grounding of grammatical structures, the shape of language acquisition, and the co-evolution of language and cognition. This rethinking is fundamental because perspective serves as a common thread that links together the four semi-modular cognitive systems governing direct experience, space-time deixis, plans, and social roles. Because perspective interacts with imagery on each of these four levels, it

provides a general rubric for knitting together all of cognition. By codifying ways of making these links for perspective taking, language provides us with smooth, controllable access to all of the objects of imagery and cognition.

Because perspective operates at the level of the sentence and not the word, it has little impact on processing or development on the auditory, articulatory, and lexical levels. Perspective does not provide a new way of understanding lexical processing mechanisms such as spreading activation, inhibition, and interference. On the contrary, the operation of perspective is itself an outgrowth of basic learning and processing mechanisms such as induction, self-organization, imagery, and generalization. In this sense, it makes little sense to propose a theory of embodied cognition grounded on perspective as a replacement for standard cognitive psychology. Instead, perspective can be viewed as an elaboration of more well-understood, basic cognitive mechanisms.

1. Empirical Demonstrations

There is now a voluminous experimental literature documenting the impact of embodied situation models on discourse processing. Glenberg (1997), Zwaan and Radvansky (1998), and Zwaan, Kaup, Stanfield, and Madden (in press) have reviewed this work in detail. The findings of this work are extremely consistent. When we listen to sentences, even in laboratory experimental contexts, we actively generate images of the situation models described by these sentences.

One method for demonstrating this effect involves giving subjects probes that are either consistent with these mental models or not. When the probes are consistent, they respond quickly, when they are not, responses are slower. For example, when we think about "aiming a dart" we imagine pinching together our fingers (Klatzky et al., 1989). To take a more complex example (Bransford, Barclay, and Franks, 1972), when subjects read (1), as opposed to (2), they are likely to false alarm when tested with (3).

1. Three turtles rested on a floating log, and a fish swam

beneath them.
2. Three turtles rested beside a floating log, and a fish swam beneath it.
3. Three turtles rested on a floating log, and a fish swam beneath it.

A second method involves giving subjects passages that produce coherent situation models such as (1) and ones that do not, such as (2).

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

The prediction here is simply that (1) is easier to read and recall than (2). This method can also be used to show that situation models function online. For example Hess, Foss, and Carroll (1995) showed that reading of the final word in a sentence was faster if it matched up with the situation model generated by the sentence.

Other methods involve showing that graphs that are consistent with situation models facilitate processing (Glenberg and Langston, 1992), checking for the updating of situation models using new information (Ehrlich and Johnson-Laird, 1982), and checking at various points for the availability of a protagonist (Carreiras, Carriedo, Alonso, and Fernández, 1997). This work provides clear experimental evidence that we do indeed construct situation models as we process discourse and that these constructions involve the assumption of perspectives in terms of direct experience, spatial position, temporal location, causal action, and social roles.

2. Depictive and Enactive Modes

The perspective hypothesis holds that we can construct mental models in either depictive or enactive modes. When we constructed

images depictively, they appear as images on a visual screen and we watch them as a spectator. Depiction relies primarily on processing in the ventral visual processing stream (Ungerleider and Haxby, 1994) that runs from the primary visual areas in the occipital lobe through the object recognition areas of the temporal lobe. Processing in the depictive mode allows only a minimal amount of perspective taking, perhaps just enough to focus attention on a figure over a ground, but not enough to become involved with the actions of that figure.

When we construct mental images in the enactive mode, they involve us not just as spectators, but also as participants. Processing in the enactive mode involves perspective taking, since we can adopt the enactive viewpoint of specific objects or participants. Processing in this mode relies on the dorsal visual stream (Goodale, 1993) that runs through the parietal and eventually projects to supplementary eye field areas in the premotor cortex. This dorsal stream processes images and models in terms of links between perception and action. The ventral stream is older (Holloway, 1995) and processing in the depictive mode is relatively easier and more automatic. As Landau and Jackendoff (1993) have noted, image processing in the ventral stream provides more detail than spatial processing in the dorsal stream. Although processing in the dorsal stream is less precise and slower, it allows us to link perception to action in a way that will support perspective taking.

As an example of how processing differs in these two modes, consider this sentence: "The skateboarder vaulted over the railing." In the depictive modality, we see the skateboarder vaulting over the railing, as if he were a figure in a video. However, if we process this sentence enactively, 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. Enactive processing allows us to construct a fuller and deeper (Craik and Lockhart, 1972) elaboration of the

mental model.

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

> The cat spotted a mockingbird perched on the feeder. She crouched down low in the grass, inching closer and closer with all her muscles tensed. Just as she pounced, the bird escaped. Disappointed, she lept up to a garden chair, raised her paw to her tongue, and began licking it.

Here, each clause links to the previous one through the perspective of the cat as the protagonist. As we chain these references together, we induce the listener to assume a single enactive perspective. The longer and more vivid our descriptions, the more they stimulate enactive processes in comprehension. Depictive and enactive processes run in parallel. As we go through the work of constructing a basic depictive mental model, we elaborate it enactively, as much as time and energy permit. Because our enactive interpretations may fail to reach completion, it is often the case that they are not fully available to our consciousness.

When we look at sentence production, as opposed to sentence comprehension, the situation is different. In production, we often have direct access to memories that are encoded in the enactive mode. Unless we are describing events as we see them, we are usually retrieving events and referents from memory. According to the perspective-taking hypothesis, these events are most likely to be encoded enactively. Sometimes we have failed to construct stored mental models in a fully coherent enactive framework. For example, when telling a joke, we might forget the punch line. This type of

failure indicates that, even within our own embodied mental models, enactive processing can be incomplete.

Perspective taking operates on five component subsystems. These subsystems process information in terms of (1) direct experience, (2) deictic spatio-temporal reference frames, (3) plans, (4) social roles, and (5) belief. Each of these subsystems can function rapidly and accurately without perspective taking. However, without perspective taking, their output is highly stimulus-bound (Hermer-Vazquez, Moffet, and Munkholm, 2001), depictive, limited, and modular. Our primate relatives display some basic abilities to perform perspective taking on each of these four levels (MacWhinney, 2003). Even without language, perspective taking partially liberates primate cognition from a complete dependence on stimulus input and permits the construction of fragmentary and limited mental models. However, with language, we can use perspective taking across these five systems to build up a single, unified embodied situation model.

3. Direct Experience

Our basic mode of interaction with objects is through direct experience. Direct perception arises immediately as we interact with objects. We use vision, touch, smell, taste, kinesthesia, and proprioception to estimate the affordances (Gibson, 1977) that objects provide for action. As we use our arms, legs, and bodies to act upon objects, we derive direct feedback from these objects. This feedback loop between action and perception does not rely on symbols, perspective, or any other form of cognitive distancing. Instead, it is designed to give us immediate contact with the world in a way that leads to full embodiment and quick adaptive reactions. Because this system does not rely on memory, imagery, perspective, or other cognition systems (Gibson, 1977), it remains fully grounded on the direct relation between the organism and the environment.

Consider the ways in which we perceive a banana. When we see a banana, we receive nothing more than an image of a yellow curved

object. However, as we interact directly with the banana, additional perceptions start to unfold. When we grab a banana, our hands experience the texture of the peel, the ridges along the peel, the smooth extensions between the ridges, and the rougher edges where the banana connects with other bananas into a bunch. When we hold or throw a banana, we appreciate its weight and balance. When we peel a banana, we encounter still further sensations involving the action of peeling, as well as the peel itself. With the peel removed, we can access new sensations from the meat of the banana. An overripe banana can assault us with its pungent smell. When we eat a banana, our whole body becomes involved in chewing, swallowing, and digestion. All of these direct interactions in vision, smell, taste, touch, skeletal postures, kinesthesia, proprioception, and locomotor feedback arise from a single object that we categorize as a "banana." It is this rich and diverse set of sensations and motor plans that constitutes the fullest grounding for our understanding of the word "banana." Of course, we know other things about bananas. We know that they are rich in potassium and Vitamin E, that they are grown in Central America by United Fruit cooperatives, and so on, but these are secondary, declarative facts (Paivio, 1971; Tabachneck-Schijf, Leonardo, and Simon, 1997) that rely on the primary notion of a banana that we derive from direct embodied perception.

3.1 Direct imagery

The development of mental imagery produces cognitive ungrounding or distancing. When we imagine a banana, we call up images of the shape, taste, and feel of a banana even when it is not physically present. This imagery does not depend exclusively on language. We might be hungry and think of a banana as a possible food source, or we might detect a smell that would lead us to construct a visual image of a banana. Recent research in neurophysiology has shown that, when we imagine objects and actions in this way, we typically activate the same neuronal

pathways that are used for direct perception and direct action. For example, when we imagine performing bicep curls, there are discharges to the biceps (Jeannerod, 1997). When a trained marksman imagines shooting a gun, the discharges to the muscles mimic those found in real target practice. When we imagine eating, there is an increase in salivation. Neuroimaging studies by Parsons et al. (1995), Martin, Wiggs, Ungerleider, and Haxby (1996), and Cohen et al. (1996) have shown that, when subjects are asked to engage in mental imagery, they use modality-specific sensorimotor cortical systems. For example, in the study by Martin et al., the naming of tool words specifically activated the areas of the left premotor cortex that control hand movements. Experimental work has shown repeatedly that switching between these sensorimotor systems during imagery tasks exacts a clear cost in reaction time (Klatzky, Pellegrino, McCloskey, and Doherty, 1989; Solomon and Barsalou, 2001).

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

Although imagery relies on the pathways used by direct perception and direct action (Decety and Grèzes, 1999; Decety et al., 1994), it differs from direct experience in four important ways:

1. Temporal lag. Using ERP and EMG measurements, Osman (1999) has shown that the image of a trigger release comes 100 msec later than the actual release.
2. Partial independence. Many patients with visual agnosia display good object recognition, but some damage to mental imagery. This shows that, although imagery

depends on the pathways used by direct perception, it also involves additional central resources that can be separately damaged. However, Behrmann, Moskovitch, and Winocur (1996) report findings from a patient who shows good imagery, but damaged object recognition. Similarly, Caplan and Waters (Caplan and Waters, 1995) have shown that patients with motor apraxia can use the phonological loop to remember word strings, even without being able to articulate normally. In order to explain patterns like these, we have to assume that imagery is partially independent of the final pathways for direct perception and action. Rather than being entirely linked to direct experience, imagery constructs internal fictive processes that are potentially separable from direct perception. In this sense, the homunculus that is used to produce imagery simulates reality, but it is not reality.

3. Decomposition. Imagery also differs from direct perception in the ways in which it can access the pieces of stored images. Barsalou (1999) has argued that our perception of an object such as an automobile allows us to enactively decompose the auto into its various pieces, such as the doors, the windows, the windshield, and the parts of the motor. The relation of each part to the others is traced through a top-down reenactment of direct experience, both perceptual and motoric. In this way, the full construct of an automobile relies on the vestiges of direct experience that are stored away as enactive images.
4. Generation. Imagery also differs from direct experience in the fact that the generation of images requires some form of active retrieval from memory. Studies using the verb generation task have pointed to an important role for frontal cortex in supporting strategic aspects of meaning access and generation (Petersen, Fox, Posner, Mintun, and Raichle, 1988; Posner, Petersen, Fox, and Raichle, 1988). In this task, subjects are shown pictures of objects and asked to think of actions they might

perform on these objects. In addition, lesion studies (Gainotti, Silveri, Daniele, and Giustolisi, 1995), PET studies (Posner et al., 1988), and fMRI analyses (Menard, Kosslyn, Thompson, Alpert, and Rauch, 1996) have shown that right frontal areas are involved in the generation or retrieval of action terms. Together, these studies point to an important role for frontal cortex in generating access cues for specific actions and the words that express those actions.

3.2 Partial ungrounding

Imagery works together with memory, planning, dreaming, and projection to allow us to move away from direct experience. Together, these processes allow us to move beyond a direct linkage to object and actions and to imagine potential actions and their possible results. These processes lead to a partial ungrounding of cognition. However, the decomposable nature of perceptual symbol systems (Barsalou, 1999) allows us to recreate full grounding when needed for fuller comprehension. The fact that cognition can become partially ungrounded through imagery should not be construed as meaning that it is fully ungrounded (Burgess and Lund, 1997).

3.3 Direct experience and language

Words provide convenient methods for mapping direct experiences onto linguistic form. In most cases, words afford little opportunity for perspective switching. The word "banana" packages together all our experiences with this object into a single unanalyzed whole. In some cases, however, our experiences are decomposed, even on the level of the word. For example, in Navajo, a chair is "bikáá'dah'asdáhí" or "on-it-one-sits." To take a more familiar example, many languages refer to a corkscrew as a "cork puller." In such examples, objects are being characterized in terms of our action

upon them. Miller and Johnson-Laird (1976) showed that definitions of nouns in terms of criterial attributes were often not as predictive as definitions in terms of imagined affordances. For example, they found that attempts to define a "table" in terms of the number or the placement of its legs or the shape of the top often failed to capture the possible variation in the shape of what counts as a table. It works better to define a table instead as an object that provides a space upon which we can place work. In this way, Miller and Johnson-Laird eventually came to the same conclusion that the Navajo reached when they called a table "bikáá'dání" or "at-it-one-works."

Languages can also capture aspects of direct experience through the projection of the body image. 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 direct experiences we discussed earlier in the case of "banana" to flesh out the meanings of words, even before they are placed into syntactic combination. The 18th century philosopher Giovanni Batista Vico understood this, when he noted that:

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

Plato attributes an even earlier statement of this type to the first philosopher, Protagoras, who declared, "Man is the measure of all things."

Adjectives encode images of direct perceptions for attributes such as weight, color, or smell. Verbs encode images of direct

action, often in relation to movements of the body. When we hear the word "walk," we immediately activate the basic elements of the physical components of walking (Narayanan, 1997). These include alternating motions of the legs, counterbalanced swinging of the arms, pressures on the knees and other joints, and the sense of our weight coming down on the earth. Because we have good access to the components of motor plans, these images can be decomposed. However, in practice they often function as unanalyzed images of integrated plans. More generally, although individual words can construct meaning by reference to the body and direct experience, they do not in themselves allow for shifts in perspective. For that, we need to move to the higher levels of the phrase and the clause.

4. Space and time

Perspective taking requires different sets of cognition mechanisms. For direct experience, perspective taking involves the projection of the body image onto the body and motions of other agents. For space, perspective taking involves the projection of a deictic center and map onto the position of another agent. Deictic centers can be constructed in three frameworks: egocentric, allocentric, and geocentric.

4.1 The egocentric frame

Egocentric deixis directly encodes the perspective of the speaker. The spatial position of the speaker becomes the deictic center or "here." Locations away from this deictic center are "there." In face-to-face conversation, the deictic center can include both speaker and listener as a single deictic center. In this case, "here" can refer to the general position of the speaker and 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".

To map our local environment around a deictic center, we create a series of deictic codes to mark the locations of objects with respect to previous body postures and eye fixations (Ballard et al., 1997). By accessing these stored deictic codes, we avoid the many computations that would be involved in having to worry repeatedly about the locations of all of the possible objects in the world around us. Ballard argues that the brain does this by establishing an internal deictic code for each object in working memory. These codes are stored with reference to our images of our body and eye positions and movements.

The establishment of deictic codes depends on a set of mechanisms for the neuronal encoding of eye movements, body image, and body maps. In an early study on this topic, Bossom (1965) gave monkeys 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 supplementary eye fields resulted in damage to the readapted visual field. This finding suggests that these frontal structures support the construction of a dynamic and adaptable visual field. Using single-cell recording techniques with macaque monkeys, Olson and Gettner (1995) located cells in the supplementary eye field of prefrontal cortex that respond not to positions in the actual visual field, but to positions on objects in visual memory. These results suggest that the prefrontal visual area works together with parietal areas to facilitate the processing of spatial representations.

Connections between posterior and frontal areas (Goldman-Rakic, 1987) provide a method for the temporary storage of deictic codes in premotor working memory areas and the accessing of previous attentional foci. Permanent traces may be stored by offline hippocampal processing and cortical downloading (McClelland, McNaughton, and O'Reilly, 1995; Redish and Touretzky, 1997). The fact that primates have a short-term memory capacity equal to that of humans (Levine and Prueitt, 1989) suggests that the basic deictic memory system for egocentric perspective can

operate smoothly without additional reliance on verbal memory systems such as the phonological loop (Baddeley, 1990; Gathercole and Baddeley, 1993).

4.2 The allocentric frame

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

Shifts in spatial perspective can lead to strange alternations of the perspectival field. For example, if we are lying down on our backs in a hospital bed, we might refer to the area beyond our feet as "in front of me," even though the area beyond the feet is usually referred to as "under me." To do this, we may even imagine raising our head a bit to correct the reference field, so that at least our head is still upright. We may also override the normal shape of the allocentric field by our own egocentric perspective. For example,

when having a party in the back yard of a house, we may refer to the area on the other side of the house as "in back of the house," thereby overriding the usual reference to this area as "the front of the house." In this case, we are maintaining our current egocentric position and perspective as basic and locating the external object within that egocentric perspective.

Prepositions often reflect the perspectival nature of allocentric reference. For example, the preposition "in back of" is based on taking the point of view of an object and locating what would correspond to its back, if it were viewed as having a body. Body parts such as the face, the stomach, the buttocks, the feet, and the head all serve as the grounding for prepositions in many languages. In fact, Heine (1993) found in a survey of African languages that over three-quarters of the relational terms derive from body parts. Historically, parts are first projected to regions of inanimate objects, such as the "back" of a car. Next, they come to refer to regions in contact with these parts, such as "back of the car." Finally, they come to refer to areas detached from the objects, as in "in back of the car." The projection from the body can also support the development of words for "to" or "in front" based on the human eye, since the eye glances toward things. Even abstract case-marking systems can be shown to derive historically from simple deictic markers such as "to" and "from" (Anderson, 1971).

The computation of allocentric reference required an evolutionary adaptation to basic primate spatial processing. We know that the parietal cortex in primates maintains separate maps for body-referenced and world-referenced positions (Snyder, Grieve, Brothcie, and Anderson, 1998). Body-referenced positions are adequate for egocentric spatial representations. However, world-referenced positions must be elaborated by perspective-taking to form allocentric representations. This could be achieved by linking frontal mechanisms to these parietal mechanisms. In addition, hippocampal mechanisms are used in spatial computations (McClelland et al., 1995). These mechanisms would not need to be modified, since they would simply store codes from a revised deictic center. It is possible that the expansion of parietal cortex and

processing in the dorsal stream that occurred about 4MYA (Holloway, 1995) could have provided our hominid ancestors with the ability to construct fully shiftable allocentric deictic centers.

4.3 The geocentric frame

The third deictic reference system, the geocentric frame, enforces a perspective based on fixed external landmarks, such as the position of a mountain range, the sun, the North Star, the North Pole, or a river. These landmarks must dominate a large part of the relevant spatial world, since they are used as the basis for a full-blown Cartesian coordinate system. The Guugu Yimithirr language in northeast Queensland (Haviland, 1996) makes extensive use of this form of spatial reference. In Guugu Yimithirr, rather than asking someone to "move back from the table," one might say, "move a bit to the mountain." We can use this type of geocentric reference in English too when we locate objects in terms of compass points. However, our uncertainty about whether our listener shares our judgments about which way is "west" in a given microenvironment makes use of this system far less common. On the other hand, we often make use of Cartesian grids centered on specific local landmarks in English. For example, we can describe a position as being "fifty yards behind the school." In this case, we are adopting an initial perspective that is determined either by our own location (e.g., facing the school) or by the allocentric perspective of the school for which the entry door is the front. If we are facing the school, these two reference frames pick out the same location. When we describe the position as being located "fifty yards toward the mountain from the school," we are taking the perspective of the mountain, rather than that of the speaker or the school. We then construct a temporary Cartesian grid based on the mountain and perform allocentric projection to the school. Then we compute a distance of 50 yards from the school in the direction of the mountain.

As we have already noted, language uses a variety of closed-class forms to express basic spatial relations. In English,

much of this work is done through prepositions, pronouns, and tense markers. In other languages, there may be a greater reliance on expressions of topological relations, contact, shape, and enclosure. However, all languages provide a rich set of expressions for egocentric and allocentric construction of space and time. These devices can be chained together in expressions, such as "in the pond under the log across the stream." Processing of these chains of spatial expressions requires the same perspective shifting mechanisms needed to process plans, as we will see in the next section.

4.4 Temporal perspective

In many ways, we conceive of time as analogous to space. Like space, time has an extent through which we track events and objects in terms of their relation to particular reference moments. Just as spatial objects have positions and extents, events have locations in time and durations. Time can also be organized egocentrically, allocentrically, or globally. When we use the egocentric frame, we relate events to our current speaking time (ST) (Vendler, 1957). Just as there is an ego-centered "here" in space, there is an ego-centered "now" in time. Just as we can project a deictic center onto another object spatially, we can also project a temporal center onto another time in the past or future. In this case, the central referent is not speaking time, but another reference time (RT). We can track the position of events in relation to either ST or RT or both using linguistic markings for tense. We can also encode various other properties of events such as completion, repetition, duration, and so on, using aspectual markers. When we come to depicting the duration of events, we can view them either as having an extent in a single dimension ("a long time") or a relative size ("mucho tiempo").

Just as we tend to view events as occurring in front of us, rather than behind us, we also tend to view time as moving forwards from past to future. As a result, it is easier to process sentences like (1)

with an iconic temporal order than ones like (2) with a reversed order. However, sentences like (3) which require no foreshadowing of an upcoming event, are the most natural of all.

1. After we ate our dinner, we went to the movie.
2. Before we went to the movie, we ate our dinner.
3. We ate our dinner and then we went to the movie.

Temporal reference in narrative assumes a strict iconic relation between the flow of the discourse and the flow of time. Processing of sequences that violate temporal iconicity by placing the consequent before the antecedent are relatively more difficult. However, in reality, it is difficult to describe events in a fully linear fashion and we need to mark flashbacks and other diversions through tense, aspect, and temporal adverbials.

5. Plans

Primates can make sophisticated use of tools for single operations on objects, but they cannot form lengthier plans that combine these actions (Byrne, 1999). Donald (1998) has argued that the ability to formulate and execute plans was the centerpiece of the mimetic revolution that accompanied the geographical expansion of *homo erectus* after 2MYA. During this period, brain mass doubled in allometric terms. Much of this expansion benefited prefrontal areas that support attention, memory, and plan organization. The expansion also benefited frontal areas that control vocal processes and temporal areas for auditory and lexical memory.

Plans require not only perspective taking, but also perspective shifting. Shifts involve new combinations of actions and objects. For example, a plan for making an arrow will involve climbing a hill to find a suitable stone, returning to a work area, locating a chipping stone, chipping the point, plucking a branch from a tree, shaping the branch into a straight stick, slicing sinew, and tying the point. Although the self remains the protagonist throughout this plan, there are continual shifts through direct experience, space, and

causal action on objects. Representing perspective shifts requires a method for representing and accessing competing plans, resolving the competition, and developing optimal sequences of the components (Sacerdoti, 1977).

It appears that the frontal lobes are uniquely adapted to construct plans in a manner that facilitates perspective tracking and switching. Dorsolateral prefrontal cortex plays a fundamental role in the storing of alternative representations in working memory (Barch et al., 1997; Braver et al., 1997; D. D. Cohen et al., 1997; Goldman-Rakic, 1987; Owen, Downes, Sahakian, Polkay, and Robbins, 1990). The ability to shift between perspectives requires a neural system for representing alternative perspectives, as well as a method for inhibiting one or more of the competing perspectives. For example, in the Stroop task (J. Cohen, Dunbar, and McClelland, 1990), the reader must inhibit the perception of the color of the word in order to quickly read the name of the word. In processing SO relative clauses, we need to move quickly from the viewpoint of the subject of the main clause to the viewpoint of the subject of the subordinate clause. In processing social relations, we need to quickly assess the viewpoints of other people, particularly as they conflict with our own views. Right frontal cortex supports memory for events that occur in discourse. The ability to store the traces of recent events in working memory is crucial for the construction of connected discourse. If we could not recall previously mentioned characters and actions, we would be unable to follow even the most basic descriptions and narratives.

The complex interconnectivity between frontal, thalamic, and cingulate areas (Fuster, 1989; Kolb and Whishaw, 1995) suggests that the frontal system integrates a variety of mental facilities, all in the service of perspective taking and shifting. Mesulam (1990) asks, "Why does (prefrontal) area PG project to so many different patches of prefrontal cortex? Why are the various areas of prefrontal cortex interconnected in such intricate patterns?" The perspective hypothesis claims that frontal cortex is attempting to integrate perspective taking and shifting in verbally represented plans.

5.1 Dissecting events

To formulate plans (Sacerdoti, 1977), we must have a way of representing the individual components of events. The ability to dissect events into their components involves a form of representation that is only incompletely attained in primates. Chimpanzees have no problems representing and naming individual objects or simple actions (Savage-Rumbaugh and Taglialatela, 2001). However, they are not able to combine these representations into fuller predicates (Terrace, Petitto, Sanders, and Bever, 1980). Some (Greenfield, 1991) have suggested that this failure arises from an inability to combine elements. However, others (Merlin Donald, 1998; Tomasello, 1999) see the deficit as involving an inability to segment events into their components. According to this view, events such as "shaking salt" are initially encoded as a single merged experience in which the shaking, the saltshaker, and the salt all form a single perceptual-action Gestalt. Similarly, in the act of cutting wood, there is no fundamental gap separating the wood, the axe, the chopping, the lifting, and the self as agent.

In order to segment reality into separate events, language and cognition provide us with a system that orders nouns into role slots constellate around verbs. We use verbs to segment the flow of reality into bite-size actions and events. Then we flesh out the nature of the events by linking actors and objects to the verbs, as fillers of role slots. Item-based grammars (Hausser, 1999; Hudson, 1984; Kay and Fillmore, 1999; MacWhinney, 1988) derive syntactic structure from the ways in which individual words or groups of words combine with others. For example, the verb "fall" can combine with the perspective of "glass" to produce "the glass fell." In this combination, we say that "fall" has an open slot or valency for the role of the perspective and that the nominal phrase "the glass" is able to fill that slot and thereby play the role of the perspective. In item-based grammars, this basic mechanism is used to produce the full range of human language. The specific phrasal structures of various languages emerge as a response to the process of combining words into appropriate role slots as we listen to sentences in real time (Hawkins, 1999).

5.2 Competing perspectives

Much of the variation we find between languages involves alternative methods for marking perspective in syntactic combinations. These variations arise because of competition between alternative participants for case marking, agreement marking, and word order positioning (Bates and MacWhinney, 1989). When we describe an event such as "the farmer grew the corn," there are two competing perspectives. The perspective of the corn is directly involved with the growing. If we wish to understand the changes that occur in the corn, we would have to assume this perspective. On the other hand, the perspective of the farmer is also relevant, since he cares for the corn in ways that make it grow. When dissecting events that have more than one participant, languages typically make a default commitment to one of these two perspectives. Accusative-nominative languages, like English, place focus on the actor by treating it as the default perspective for the clause. They then treat the activity of the patient as a secondary perspective contained within the scope of the larger perspective of the subject. In ergative-absolutive languages, like Basque or Djirbal, the default primary focus is typically on the participant undergoing the change, rather than on the participant causing the change. In the sentence "The farmer grew the corn," the farmer is placed into the ergative case and the corn is in the absolutive case. The absolutive is also the case that is used for the word "corn" in the intransitive sentence "The corn grew." This means that ergative 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.

Variations in the marking of ergativity demonstrate three clear effects of perspective taking on event construction. These effects arise because we are more likely to assume the perspective of the agent when the action is immediately present and when the agent is closer to ego.

1. Tense. Gujarati (Delancey, 1981) uses ergative-absolutive marking in the perfective, but not in the imperfective

tense. Because of the ongoing nature of the imperfective ("was buying"), we tend to become involved in the action and therefore assume the perspective of the causor. Because of the completive nature of the perfective ("bought"), we are less involved in the action and more willing to treat the agent as the secondary or ergative perspective.

2. Person. In languages like Kham (Delancey, 1981), the choice of ergative or nominative marking can also depend on the person of the agent. When the actor is in third person, many languages use ergative marking. However, when the actor is in first or second person, these same languages often shift to using 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.
3. Intentionality. Ergative marking can also be used to mark intentionality. Delancey (1981) describes this for the Caucasian language Batsbi, which uses ergative case for the subject of verb like "fall" when the falling is intentional and absolutive marking for the subject when the falling is unintentional. Sentences with the absolutive could be read like "falling happened to me."

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

5.3 Constructions marking perspective

The choice of absolutive or accusative marking is only one of many linguistic choices influenced by perspective. Other constructions (Kay and Fillmore, 1999), structures (Chomsky, 1981), or options (Halliday and Hasan, 1976) shaped by perspective include:

1. Passives. The choice of a passive over an active can be induced by the fact that the perspective is indefinite, unknown, or unidentifiable. Often we select the passive when we wish to avoid attributing responsibility (Seliger, 1989).
2. Double object. The decision to say either "I sent John the book" or "I sent the book to John" reflects the extent to which we wish to focus on the secondary perspective of the recipient (Zubin, 1979).
3. Inverse. Languages may require that nouns be placed in an order of relative animacy. When this happens, the verb can mark perspective inversion (Whistler, 1985).
4. Obviative. The marking of possession often involves expressions for equation, description, existence, and location (Heine, 1997). In the obviative, possession can be shifted to a non-perspectival owner, as in split ergative person marking.
5. Fictive agency. In "the library boasts three major collections," we are taking the perspective of an inanimate object and treating it fictively (Talmy, 1988) as an agent. Other examples include "the path runs down to the river" and "the screws hold the legs onto the table."
6. Conflation. We can join together "the car knocked the bicycle" and the "bicycle fell over" into the single clause "the car knocked over the bicycle." When we do this, we subordinate the perspective of the bicycle in the second clause to the overall controlling perspective of the car.
7. Comparison. The directionality of comparison is governed by the projection of features from a perspective. Saying that "Bill speaks like my chimpanzee" is much different from saying that "My chimpanzee speaks like Bill."
8. Complementation and control. The unmarked complement structure is one which maintains the perspective of the main clause, as in "Bill wanted to go" where "Bill" is the subject of the main verb "want" and

the complement verb "go." In addition. The contrast between "the doll is easy to see" and "the doll is eager to see" reflects alternative perspectival configurations of the verbal adjectives "eager" and "easy."

9. Relativization. In "the cat the dog chased hissed," the "cat" shifts from the role of agent to the object and back again to agent. These shifts of perspective must be clearly marked in the grammar.
10. Binding. Pronouns mark both co-reference and perspective. When we say "He said Bill won" we know that "Bill" is not co-referent with "he." These facts about the grammar are determined by the system for marking perspective shifts.

This is only a partial list. Other syntactic processes influenced by perspective include adverbialization, phrasal attachment, dislocation, clefting, topicalization, possession, ellipsis, coordination, and reflexivization. In fact, it is difficult to find any syntactic process that is not at least partially impacted by perspective marking.

Generalizing from this observation, we can say that syntax has two basic functions. The first function is attachment. The syntactic processor uses surface cues to link words together into a relatedness structure (Hausser, 1999; O'Grady, 2002) during online processing. The second function is perspective. The syntactic processor uses these same surface cues to assume a series of perspectives. For example, in the sentence, "The cat licked herself," the processor links the subject and object into the slots required by the verb "lick." At the same time the perspectival processor encourages us to assume the role of "the cat" to interpret this process enactively, much as we would interpret fictive agency as in "the screws hold the legs to the table."

In the following sections, we will examine these effects in detail for three selected areas: binding, relativization, and ambiguity marking. Given limitations in space, we will focus on these three areas because of their centrality in both linguistic and psycholinguistic work of the last 20 years. However, similar analyses apply to each of the syntactic domains we have mentioned.

5.4 Co-reference and c-command

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

When the first noun is low in referentiality (Ariel, 1990), the foundation is unclear and the process of comprehension through structure building is thwarted. If the starting point is a full nominal, referentiality is seldom at issue. However, if the starting point is a pronoun, then there must be a procedure for making it referential by finding an antecedent. One way of doing this is to link up the pronoun to an entity mentioned in the previous discourse. In a sequence like (1), it is easy to link up "he" with "John," since John has already been established as an available discourse referent. However, in (2), the pronoun has no antecedent, and the sequence seems awkward and unlinked.

(1) John_i was trying to list the Ten Commandments. He_i was unable to get past the first six.
(2) 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 an embodied situation model.

The theory of Government and Binding (Chomsky, 1982; Grodzinsky and Reinhart, 1993; Reinhart, 1981) treats this phenomenon in terms of structural relations in a phrase-marker tree.

Principle C of the binding theory holds that a pronoun cannot c-command its referent. An element is said to c-command another element if it stands in a direct chain above it in a phrase tree. As a result, Principle C excludes a co-referential reading for (1), but not for (2).

(1) He_i says $Bill_i$ came early.
(2) $Bill_i$ says he_i came early.

In (1) the pronoun c-commands its referent because it stands in a direct chain of dominance above it in the tree. In (2) the pronoun is down below its referent in the tree and therefore does not c-command "Bill." The perspective hypothesis attributes the unavailability of the co-referential reading of (1) to the fact that starting points must be referential. Without further cues, the processor cannot wait for a subsequent identifying co-referent and chooses instead to force co-reference with some entity from previous discourse. In (2), on the other hand, "Bill" is available as a referent and therefore "he" can co-refer to "Bill."

This effect is not a simple matter of linear order, since co-reference between a pronoun and a following noun is perfectly good when the pronoun is in an initial subordinate clause. Consider this contrast, where the asterisk on (3) indicates that "he" cannot be co-referential with "Lester."

(1) When he_i drank the vodka, $Lester_i$ started to feel dizzy.
(2) $Lester_i$ started to feel dizzy, when he_i drank the vodka.
(3) *He_i started to feel dizzy, when $Lester_i$ drank the vodka.
(4) When $Lester_i$ drank the vodka, he_i started to feel dizzy.

Binding theory views the subordinate clause as generated within the VP where it is available for logical interpretation. In (1) and (2), "Lester" c-commands the pronoun and coreference is possible, even after the movement. In (3) and (4), it does not and coreference should be blocked. However, the acceptability of (4) is a problem for this version of binding theory. Reinhart (1983) explains the anomaly by arguing that coreference in (4) is supported by discourse constraints. In a sense, Reinhart's discourse account is akin to the

discourse account developed within the perspective hypothesis. The perspective hypothesis attributes the acceptability of (1) to the presence of the subordinating conjunction "when" which gives the processor instructions that a subsequent NP can be used for co-reference to "he." In (3), no such instructions are available.

The referentiality requirement also applies in a somewhat weakened form to the direct and indirect objects of verbs. Van Hoek (1997) shows how availability for co-reference is determined by position in the argument chain (Givón, 1976). Although attention is first focused on the subject or trajector, it then moves secondarily to the object or other complements of the verb that are next in the "line of sight" (Langacker, 1995). This gradation of the perspectival effect as we move through the roles of subject, direct object, adjunct, and possessor is illustrated here:

(1) He_i often said that $Bill_i$ was crazy.
(2) ? John often told him_i that $Bill_i$ was crazy.
(3) ? John often said to him_i that $Bill_i$ was crazy.
(4) John often said to his_i mother that $Bill_i$ was crazy.
(5) The students who studied with him_i enjoyed $John_i$.

By the time we reach elements that are no longer in the main clause, as in (5), co-reference back to the main clause is not blocked, since elements in a subordinate clause are not crucial perspectives for the structure building process. This gradient pattern of acceptability for increasingly peripheral clausal participants matches up with the view that the process of perspective taking during structure building requires core participants to be referential. Solan (1983) has shown that even 4-year-olds prefer sentences like (2) to (1).

Principle C can account for some of these patterns. For example, the acceptability of (5) above is in conformity with the fact that there is no c-command relation between "him" and "John." It is often true that both the binding theory and the perspective hypothesis provide good parallel accounts of particular anaphoric patterns. In this sense, formal theory and the perspective account complement each other.

The perspective hypothesis also provides an account of the

acceptability of certain types of forward co-reference that are not explained by the binding theory. Consider this pair:

(1) She$_i$ had just come back from vacation, when Mary$_i$ saw the stack of unopened mail piled up at her front door.
(2) *She$_i$ came back from vacation, when Mary$_i$ saw the stack of unopened mail piled up at her front door.

The presence of "had just" in (1) works to generate a sense of ongoing relevance that keeps the first clause in discourse focus long enough to permit co-reference between "she" and "Mary." These sentences from Reinhart (1983) provide further examples of aspectual effects on perspective taking.

(1) In Carter's$_i$ hometown, he$_i$ is still considered a genius.
(2) ? In Carter's$_i$ hometown, he$_i$ is considered a genius.

Although both of these sentences can be given co-referential readings, it is relatively easier to do so for (1), because of the presence of the "still" which keeps the co-referent active in memory.

Preposed prepositional phrases have often presented problems for binding theory accounts (Kuno, 1986). Consider these examples:

(1)Near John$_i$, he$_i$ keeps a laser printer.
(2)Near John's$_i$ computer desk, he$_i$ keeps a laser printer.
(3)*He$_i$ keeps a laser printer near John$_i$.
(4)*He$_i$ keeps a laser printer near John's$_i$ computer desk.

In (2) we have enough conceptual material in the prepositional phrase to enactively construct a temporary perspective for "John." In (1) this is not true, and therefore "John" is not active enough to link to "he." The binding theory attempts to explain patterns of this type by referring to the "unmoved" versions of the sentences in (3) and (4) above. Co-reference is clearly blocked in (3) and (4), despite the fact that it is possible in (2). This indicates that linear order is crucial for the establishment of perspective and that (2) does not derive either online or offline from (4).

Two further examples from van Hoek (1997) illustrate a related point.

(1) In Tim's$_i$ play, he$_i$ offers Mary a mansion.
(2) In Tim's$_i$ play, he$_i$ promised Mary a role.

In (1), we take the role of an outside observer describing a creative act inside the frame of the play. In (2), on the other hand, we are less involved in Tim's perspective. Here, the structural account would view the preposed phrase of (1) as an adverb and therefore not subject to blocked co-reference. Here, again, both the formal and functional accounts both work descriptively. However, the functional account is more useful in terms of explanation.

Just as markers of ongoing relevance such as "had just" or "still" can increase the openness of a pronoun in a main clause to co-reference, so indefinite marking can decrease the openness of a noun in a preposed subordinate clause noun for co-reference, as indicated by the comparison of (1) with (2).

(1) While Ruth argued with the man$_i$, he$_i$ cooked dinner.
(2) ? While Ruth argued with a man$_i$, he$_i$ cooked dinner.
(3) While Ruth was arguing with a man$_i$, he$_i$ was cooking dinner.

The addition of an aspectual marker of current relevance in (3) overcomes the effect of indefiniteness in (2), making "man" available as a co-referent for "he". Gradient patterning of this type provides further evidence that pronominal co-reference is under the control of pragmatic factors (Kuno, 1986). In this case, the specific pragmatic factors involve interactions between definiteness and perspective. The more definite the referent, the easier it is to assume its perspective.

Wh-words introduce a further uncertainty into the process of structure building. In strong crossover (Postal, 1971) sentences like (1), the initial wh-word "who" indicates the presence of information that needs to be identified.

(1) *Who$_i$ does he$_i$ like most?
(2) Who$_i$ does he$_j$ like most?
(3) Who$_i$ is hated by his$_i$ brother most?
(4) Who$_i$ thought that Mary loved him$_i$?

(5) Who$_i$ likes his$_i$ mother most?
(6) Who$_i$ said Mary kissed him$_i$?
(7) Who$_i$ likes himself$_i$ most

In (1) the listener has to set up "who" as an item that must be eventually bound to some argument slot. At the same time, the listener has to use "he" as the perspective for structure building. The wh-word is not a possible candidate for the binding of the crucial subject pronoun, so it must be bound to some other referent as in (2). However, when there is a pronoun that is not in the crucial subject role, co-reference between the wh-word and the pronoun is possible, as in (2) through (7). In these examples, the wh-word can co-refer to non-central components, such as objects and elements from embedded clauses. Only co-reference with subjects, as in (1), is blocked.

This brief discussion has only sampled only a few of the way in which the perspective hypothesis can illuminate the grammar of co-reference. Other areas of the binding theory in which the perspective hypothesis provides direct accounts include the contrast between strong and weak crossover, the binding of reflexives (Kuno, 1986), and the assignment of quantifier scopes.

5.5 Clitic assimilation

The English infinitive "to" typically assimilates with a preceding model verb to produce contractions such as "wanna" from "want to" in cases such as (1). However, this assimilation is blocked in some environments, such as (2), leaving us with (3) instead.

(1) Why do you wanna go?
(2) *Who do you wanna to go?
(3) Who do you want to go?

Chomsky (1981) and others have argued that the blocking of the assimilation in (3) is due to the presence of the trace of an empty category in the syntactic tree. However, there is reason to believe that the environment in which assimilation is favored is determined

not by syntactic forces, but by perspectival forces. In particular, we can contrast (1) and (2) below in which the infinitive does not cliticize with the verb with (3) where it does. In the case of (3), the subject has an immediate obligation to fulfill, whereas in (1) and (2), the fact that the subject receives the privilege of going is due presumably to the intercession of an outside party. Thus, the perspective continuation is less direct in (1) and (2), than it is in (3).

(1) I get ta go. (Privilege)
(2) I got ta go. (Privilege)
(3) I gotta go. (Obligation)

According to this account, cliticization occurs when a motivated subject engages in an action. When there is a shift to another actor, or a conflict of perspectives, as in "Who do you want ta go?", cliticization is blocked.

5.6 Relativization

Restrictive relative clauses can require us to compute multiple shifts of perspective. Consider these four types:

SS:	The dog that chased the cat kicked the horse.	0 switches
OS:	The dog chased the cat that kicked the horse.	1- switch
OO:	The dog chased the cat the horse kicked.	1+ switch
SO:	The dog the cat chased kicked the horse.	2 switches

In the SS 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 OS type, perspective switches from the main clause subject (dog) to the relative clause subject (cat). However, this perspective shift is made less abrupt by the fact that "cat" is the object of the main clause and receives some secondary focus before the shift is made. In the OO type, perspective also switches once. However, in this case, it switches more abruptly to the subject 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, "the dog the cat the boy liked chased snarled" has four difficult perspective switches (dog -> cat -> boy -> cat -> dog). Sentences that have as much perspective shifting as this without additional lexical or pragmatic support are incomprehensible, at least at first hearing.[1]

The perspective account predicts this order of difficulty: SS > OO = OS > SO. Studies of the acquisition (MacWhinney, 1982) and adult processing (MacWhinney and Pléh, 1988) have provided support for these predictions. A reaction time study of Hungarian relative clause processing by MacWhinney and Pléh (1988) shows how perspective processing integrates topicalization and subject-ivalization. In Hungarian, all six orders of 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 object is the topic, the order of difficulty is OO > OS = SO > SS. These sentences illustrate this contrast in Hungarian, using English words and with the relative clause in parentheses and NOM and ACC to mark the nominative subject and the accusative object:

---

1 The mere stacking of nouns is not enough to trigger perspective-shift overload. Consider the sentence, "My mother's brother's wife's sister's doctor's friend had a heart attack." Here, we do not really succeed in taking each perspective and switching to the next, but some form of minimalist comprehension is still possible. This is because we just allow ourselves to skip over each perspective and land on the last one mentioned. In the end, we just know that someone's friend had a heart attack.

S(SV)OV: The boy-NOM (he chased car-ACC) liked girl-ACC.
"The boy who chased the car liked the girl."

O(OV)SV: The boy-ACC (car-NOM chased him) girl-NOM liked.
"The girl like the boy the car chased."

The S(SV)OV pattern is the easiest type for processing in the SOV word order. It follows the English pattern observed above. The O(OV)SV pattern is the easiest type to process in the OSV word order. Here the consistent maintenance of an object perspective through the shift from the main to the relative clause is easy, since the processor can then smoothly shift later to the overall sentence perspective. This contrast illustrates the fundamental difference in the way topic-centered languages manage the processing of perspective.

5.7 Ambiguity

Syntactic ambiguities and garden paths typically arise from competition (MacDonald, Pearlmutter, and Seidenberg, 1994; MacWhinney, 1987) between alternative perspectives. With a preposed participial as in (1), we assume the default perspective of a speech act participant ("you" or "me"), although we can also entertain the perspective of the "relatives." In (2), the preposed subordinate clause prepares us to quickly accept the perspective of the "visiting relatives." However, even in this case, we can still shift, if we wish, to the perspective of a speech act participant.

(1) Visiting relatives can be a nuisance.
(2) If they arrive in the middle of a workday, visiting relatives can be a nuisance.
(3) Brendan saw the Grand Canyon flying to New York.
(4) Brendan saw the dogs running to the beach.
(5) The women discussed the dogs on the beach.
(6) The women discussed the dogs chasing the cats.

In (3), the initial perspective resides with "Brendan" and the shift to the perspective of "Grand Canyon" is difficult because it is

inanimate and immobile. The shift to the perspective of "the dogs" is easier in (4), although again we can maintain the perspective of "Brendan" if we wish. In cases of prepositional phrase attachment competitions, such as (5), we can maintain the perspective of the starting point or shift to the direct object. If we identify with "the women," then we have to use the beach as the location of their discussion. If we shift perspective to "the dogs" then we can imagine the women looking out their kitchen window and talking about the dogs as they run around on the beach. In (6), we have a harder time imagining that the women, instead of the dogs, are chasing the cats.

As these examples illustrate, the starting point is always the default perspective. In transitive sentences, there is always some attentional shift to the object, but this shift can be amplified, if there are additional cues, as in (6). In some syntactic contexts in English, it is possible to shift perspective even more abruptly by treating the verb as intransitive and the following noun as a new subject. These examples illustrate this effect:

(1) Although John frequently jogs, a mile is a long distance for him.
(2) Although John frequently jogs a mile, the marathon is too much for him.
(3) Although John frequently smokes, a mile is a long distance for him.

Detailed self-paced reading and eye-movement studies of sentences like (1), with the comma removed, show that subjects often slow down just after reading "a mile." This slow down has been taken as evidence for the garden-path theory of sentence processing (Mitchell, 1994). However, it can also be interpreted as indicating time spent in shifting to a new perspective when the cues preparing the processor for the shift are weak. Examples of this type show that perspective interpretation is an integral part of online, incremental sentence processing (Marslen-Wilson and Tyler, 1980).

Perspectival ambiguities also arise from competitions between alternative interpretations of quantifier scopes. Consider these two examples:

(1) Someone loves everyone.
(2) Everyone is loved by someone.

If we take the perspective of "someone" in (1), we derive an interpretation in which it is true of some person that that person loves all other people. However, if we take the perspective of "everyone," we derive an interpretation in which everyone is loved by at least one person. This second interpretation is much more likely in (2), because there "everyone" is the starting point. However, both interpretations are potentially available in both cases, because it is always possible to switch perspective away from the starting point to subsequent referents in a sentence, given additional processing time and resources.

6. Perspective and Language Acquisition

The perspective hypothesis provides us with a new way of integrating old insights regarding processes in language acquisition. The view of language learning as an enactive process was articulated by Plato, Augustine, Vico, Dewey, Kant, Montesorri, Piaget, Vygotsky, and others.

6.1 Direct experience and word learning

In the framework of the current account, we can say that language development begins with highly grounded mimetic symbols. Diary studies by Lewis (1936) and Halliday (1975) have shown how children express themselves through gestures, cries and other motions in the prelinguistic period. Others (Bates, 1976; Bower, 1974; Piaget, 1952) have noticed that the pointing gesture develops out of the attempt to grasp an object. Similarly, expressive sighs develops from the act of relaxing the muscles of the chest. Through symbolic distancing (Werner and Kaplan, 1963), fully grounded actions and perceptions slowing become ungrounded. When these early gestures and prosodies match up with established norms

sanctioned by the community, they are reinforced and retained. When they do not clearly match community norms, they are modified or dropped. In this way, even these highly grounded forms of reference become codified.

The perspective hypothesis holds that the child must construct an enactive relation between meaning and sound. Saussurean doctrine holds that this relation is arbitrary (de Saussure, 1966). However, this arbitrariness may hold only in terms of the larger community. For the individual learner, learning is facilitated by the formation of covert, private links (Atkinson, 1975) between sound and meaning based on properties such as sound symbolism (Brown, Black, and Horowitz, 1955; Hinton, Nichols, and Ohala, 1994) or enactive matches (Meltzoff, 1988; Werner and Kaplan, 1963). When language fails to provide these matches, children simply construct their own. Once the links have generated strong reciprocal connections (Van Orden, Holden, Podgornik, and Aitchison, 1999) and once lexical access becomes automated (Keenan and MacWhinney, 1987), these ad hoc links can fade away.

Sometimes, children will overtly display the internal enactive cues they have used to acquire new meanings and concepts. For example, Jon Fincham (personal communication) reports that, when his son Adam was just past 3, he had his first experience using scissors to cut down lines drawn on paper. Accompanying each full stroke of the scissors was a perfectly synchronized, corresponding mouth/jaw movement. When the scissors opened, so would his mouth. When the scissors closed, again so would his mouth. He would do this repeatedly with each cut of the scissors whenever he used them for several weeks.

Recently, cognitive psychologists have explored the possibility that word meanings are acquired simply from the statistics of co-occurrence (Burgess and Lund, 1997; Landauer and Dumais, 1997). These models have been remarkably successful in capturing a variety of effects. The semantic vectors acquired in these models nicely mirror the distribution of patterns in human associative memory. As a result, it is plausible to imagine that these systems provide supplements to the basic processes of grounded learning.

They may also help the child in the solution of aspects of the bootstrapping problem for both word meaning (Li, Burgess, and Lund, 2001) and syntax (Gleitman, 1990). For adults, this type of learning can support development of the highly ungrounded use of language that predominates in areas such as academic or legal discourse. However, by itself, co-occurrence learning of this type cannot provide a satisfactory account of basic word meaning (Kaschak and Glenberg, 2000).

6.2 Marking spatial and temporal perspectives

Children first learn to make spatial reference by developing a basic egocentric understanding of the positions of objects in space. Piaget (1952) has described this development in terms of the development of the object concept and procedures for dealing with invisible displacements. In learning to remember the positions of objects, the preverbal child relies on each of the three spatial reference systems. However, as Piaget has observed, the egocentric frame is primary. At the end of the second year, when the child comes to the task of learning language, the first locative terms are primarily egocentric and deictic. Allocentric terms such as "in" or "on" are initially processed in terms of affordances and topological relations, rather than through a complete shift of perspective to the distal object. Slowly, the use of the allocentric frame takes on an independent existence and children learn to shift reference between these frames. Geocentric reference is acquired much later (de Leon, 1994).

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

Work on the development of spatial perspective shifting has tended to focus on the comprehension of instructions and maps. The work has shown that the ability to shift perspectives emerges gradually during the school years (Hardwick, McIntyre, and Pick, 1976; Rieser, Garing, and Young, 1994).

6.3 Perspective and item-based patterns

Children's first sentences are produced through the use of what MacWhinney (1975; 1982) called item-based predicate patterns. These patterns are grounded on the individual syntax and conceptual structure of operators such as "my," "give," "more," and "with." Before age 3, there is little generalization over these argument structures and much of grammar is tightly grounded on the action schema underlying each of these predicates. Recent research (Goldberg, 1999; Lieven, Pine, and Baldwin, 1997; Tomasello, 1992) has emphasized the role of individual verbs in constructing syntax from the bottom up. In all of these accounts, the child begins with separated "verb islands" that are later linked together in larger construction types. Each verb encodes a slightly different pattern of muscle control, attentional movement, iteration, and goal direction; and each verb involves slightly different action perspectives. After age 3 (Tomasello, 2000), children begin to relate the various verb types into loosely coherent constructions. However, all aspects of this learning are still closely linked to the underlying physical realization of the verb.

Researchers using the NTL framework (Bailey, Chang, Feldman, and Narayanan, 1998; Maia and Chang, 2001) have shown how one can construct detailed models of the components of verbs such as "walk." "stumble," "grab," or "push." An NTL model of "pushing" would refer in detail to the actions of the hands, back, and legs. If a rather small object were to be pushed across a short space, then only the hands would be involved, as when we push a salt shaker across the table. However, if we have to push a table against the wall, we will need to use our legs, our back and specific postures of our hands. Moreover, pushing is a process that has a beginning, duration, and possible end. All of these elements must be tightly specified in an embodied model of the verb.

Slobin (1985) argues that children use causal roles to express a perspective that he calls the manipulative activity scene. In this scene, children distinguish the role of the initial perspective from the role of the object of the action. For each verb, the nature of these

actions and changes is different. Some involve movements; others involve experiences; still others involve various forms of causation. As a result, children work within each individual verb frame to distinguish the initial perspective or starting point (arg1 or the first argument) from the final object of the action (arg2 or the second argument). In verbs with a single argument, there is only one perspective. In verbs with three arguments, there can be an additional secondary perspective (arg3 or the third argument). The specific semantic value of these three roles (arg1, arg2, and arg3) must be characterized separately for each verb. The NTL framework shows how this characterization can be grounded on the specific action schemas associated with body movements and intentional shifts for each verb.

Children learning a language with clear accusative marking such as Russian (Gvozdev, 1949) or Hungarian (MacWhinney, 1974) first learn to mark the accusative on verbs that have clear manipulative activities, such as "break" or "hit". Similarly, children learning Kaluli (Schieffelin, 1985) first mark the ergative when it occurs with high transitivity verbs. Because verb frame generalization is limited before age 3, there is little overgeneralization of these early markings to intransitives or verbs with low transitivity.

Early on, children's perspectives on individual verbs can lie between those of accusative and ergative languages. For example, when a child says "picky up," we may initially assume that this means, "You pick me up." However, the actual early meaning is probably more focused on the child than on the agent who does the picking up. In this sense, it is more like "me experience picking up."

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

6.4 Perspective and the development of binding

There have also been many studies of children's learning of anaphoric relations, particularly in the context of the binding theory (Chomsky, 1981). This research shows that children are sensitive early on to violations of Principle C, which block coreference in "He said Bill won." This fact has been used to argue that Principle C is an innate component of Universal Grammar (UG). However, these facts can also be interpreted as evidence for the cognitive centrality of perspective taking.

In the area of reflexives, the developmental results have been more problematic for proponents of the binding theory. For example, in a sentence such as (1), children tend to interpret "him" as co-referential with "horse," as if it were (2).

(1) The dog said that the horse hit him.
(2) The dog said that the horse hit himself.

Sentence (2) obeys Principle A of the binding theory that a reflexive pronoun must have a more prominent antecedent in its minimal domain. Children have no trouble learning this rule, since it involves a clear cue and a local syntactic structure. However, Principle B, which requires that a pronominal must not have a more prominent antecedent in its minimal domain, causes children more problems. To get around this empirical failure, theorists (Chien and Wexler, 1990; Grodzinsky and Reinhart, 1993) have introduced a partition in the binding theory between referring expressions that trigger binding and co-reference and non-referring expressions that only trigger co-reference. However, evidence in support of this two-process account is incomplete and inconsistent (O'Grady, 1997).

The perspective hypothesis provides a rather more direct account of children's processing of these sentences. According to this account, the child starts processing (1) from the perspective of "the dog." Perspective then shifts to "the horse" and does not return to the overall subject in time to bind "him" to "the dog." In order to master the perspective shifting required by Principle B, children must

improve their methods for holding two subjects in mind and switching quickly between them[2]. Children with Specific Language Impairment (SLI) have a particularly difficult time mastering this switching (Franks and Connell, 1996; van der Lely, 1994). This suggests that the syntactic impairment in at least some children with SLI may well emerge from a deeper impairment of core processes in perspective taking and switching. Processing of (2) is less problematic, because there is a clear local cue that forces coreference to the current perspective.

6.5 Perspective and coordination

Perspective maintenance plays an important role in children's imitations and productions of conjoined sentences (Ardery, 1979; Lust and Mervis, 1980; D. I. Slobin and Welsh, 1973). These studies have shown that young children find it easier to imitate a sentence like (1), as opposed to ones like (2).

(1) Mary cooked the meal and ate the bread.
(2) Mary cooked and John ate the bread.

In (1) there is no perspective shift, since the perspective of Mary is maintained throughout. In (2), on the other hand, perspective shifts from Mary to John. Moreover, in order to find out what Mary is

---

2 The perspective account also helps us to understand some aspects of reflexives that are problematic for the c-command account. In the perspective account, co-reference is blocked in (b) because reflexives need to bind to referents that are already mentioned or for which there is a cue that promises that they will be mentioned.
The journey exposed Tom to himself far more than he had hoped.
* The journey exposed himself to Tom far more than he had hoped.
Bill told John that pictures of himself were on display in the Post Office.
Alfred thinks he is a great cook, and Felix does too.
In (c) the ability of "himself" to refer to either "Bill" or "John" is a reflection of the fact that both are perspectival. A similar effect arises in the very different structure of (d) in which both Alfred and Felix are possible perspectives.

cooking, we have to maintain the perspective of both Mary and John until the end of the sentence.

7. Conclusion

The perspective hypothesis offers a new way of understanding the linkage between language, society, and the brain. In this new formulation, communication is viewed as a social interaction that activates mental processes of perspective taking. Because perspective taking and shifting are fundamental to communication, language provides a wide array of grammatical devices for specifically marking perspective and perspective shift. The process of perspective shifting relies on at least four major neuronal systems that involve large areas of the cortex. Together, these systems allow us to store and produce images of previous direct experiences, spatial positions, plans, and social roles. Perspective allows us to thread together information from these three semimodular sources into a coherent integrated cognitive view.

The perspective hypothesis generates a broad series of empirically testable claims about cognitive processing, language processing, language structure, and neuronal processing. However, the hypothesis must still be clarified in several ways:

1. The conditions governing the movement of attention during online processing need to be fully specified and simulated in the form of a processing model for a variety of languages, This work can build on cross-linguistic studies of sentence processing (MacWhinney and Bates, 1989) and the analyses of cognitive grammar (Langacker, 1987).
2. The management of perspective taking through grammatical devices needs to be specified for a wider variety of grammatical structures in a wider variety of languages.
3. The perspective hypothesis needs to be systematically

applied to the sentence-processing literature to evaluate the extent to which it can provide alternative accounts to theories such as the garden-path model (Frazier, 1987) or capacity limitations (R. Lewis, 1998).

4. The implications of the hypothesis for sentence production need to be more fully specified.
5. The specific functional neural circuits that support perspective switching on the four proposed levels need to be more fully characterized and documented.
6. The development of ungrounded cognition through the growth of perspective, memory, and imagery needs to be documented in developmental terms.
7. We need more information about the emergence of perspective taking during language evolution.

This is a lengthy agenda. However, if examination of these issues helps us to better understand language, cognition, and the brain, then exploration of the perspective hypothesis will have been worthwhile.

References

Anderson, J. M. (1971). *The grammar of case: Towards a localist theory*. London: Cambridge University Press.

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

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

Aristotle. (1932). *The Rhetoric*. New York: Appleton-Century-Crofts, Inc.

Atkinson, R. (1975). Mnemotechnics in second-language learning. *American Psychologist, 30*, 821–828.

Austin, J. L. (1962). *How to Do Things with Words*. Cambridge, MA: Harvard University Press.

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

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

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

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

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

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

Bates, E. (1976). *Language and context: The acquisition of pragmatics*. New York: Academic Press.

Bates, E., and MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney and E. Bates (Eds.), *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

Black, J., Turner, T., and Bower, G. (1979). Point of view in narrative comprehension, memory, and production. *Journal of Verbal Learning and Verbal Behavior, 18*, 187–198.

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

Bower, T. G. R. (1974). *Development In Infancy*. San Francisco: Freeman.

Bransford, J., Barclay, R., and Franks, J. (1972). Sentence memory: A constructive vs. interpretive approach. *Cognitive Psychology, 3*, 193–209.

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

Brown, R. N., Black, A. H., and Horowitz, A. E. (1955). Phonetic symbolism in four languages. *Journal of Abnormal and Social Psychology, 50*, 388–393.

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

Burgess, C., and Lund, K. (1997). Modelling parsing constraints with high-dimension context space. *Language and Cognitive Processes, 12*, 177–210.

Byrne, M. (1999). Human cognitive evolution. In M. C. Corballis and S. E. G. Lea (Eds.), *The descent of mind: Psychological perspectives on hominid evolution* (pp. 71–87). Oxford: Oxford University Press.

Caplan, D., and Waters, G. S. (1995). On the nature of the phonological output planning processes involved in verbal rehearsal: Evidence from aphasia. *Brain and Language, 48*, 191–220.

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

Chien, Y., and Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition, 1*, 225–295.

Chomsky, N. (1981). *Lectures on government and binding*. Cinnaminson, NJ: Foris.

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

Clark, H., and Marshall, C. (1981). Definite reference and mutual knowledge. In B. W. A. Joshi and I. Sag (Eds.), *Elements of discourse understanding*. Cambridge, MA: Cambridge University Press.

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

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

Cohen, J., Dunbar, K., and McClelland, J. (1990). On the control of automatic

processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332–361.

Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A. K., et al. (1996). Changes in cortical activity during mental rotation. A mapping study using functional MRI. *Brain, 119*, 89–100.

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

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

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

de Saussure, F. (1966). *Course in general linguistics*. New York: McGraw-Hill.

Deacon, T. (1997). *The symbolic species: The co-evolution of language and the brain*. New York: Norton.

Decety, J., and Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences, 3*, 172–241.

Decety, J., Perani, D., Jeannerod, M., Bettinardi, V., Tadary, B., Woods, R., et al. (1994). Mapping motor representations with positron emission tomography. *Nature, 371*, 600–602.

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

Donald, M. (1991). *Origins of the Modern Mind*. Cambridge, MA: Harvard University Press.

Donald, M. (1998). Mimesis and the Executive Suite: missing links in language evolution. In J. R. Hurford, M. G. Studdert-Kennedy and C. Knight (Eds.), *Approaches to the evolution of language*. New York: Cambridge University Press.

Dunbar, R. (2000). Causal reasoning, mental rehearsal, and the evolution of primate cognition. In C. Heyes and L. Huber (Eds.), *The evolution of cognition*. Cambridge, MA: MIT Press.

Ehrlich, K., and Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior, 21*, 296–306.

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

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

Feldman, J., Lakoff, G., Bailey, D., Narayanan, S., Regier, T., and Stolcke, A. (1996). L_o — The first five years of an automated language acquisition project. *AI Review, 10*, 103–129.

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

Franks, S. L., and Connell, P. J. (1996). Knowledge of binding in normal and SLI children. *Journal of Child Language, 23*, 431–464.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 601–681). London, UK: Lawrence Erlbaum Associates.

Frith, C. D., and Frith, U. (1999). Interacting minds — a biological basis. *Science, 286*, 1692–1695.

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

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

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

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

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

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

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*, 3–55.

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

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

Goffman, E. (1955). On face-work: An analysis of ritual elements in social interaction. *Psychiatry, 18*, 213–231.

Goldberg, A. E. (1999). The emergence of the semantics of argument structure constructions. In B. MacWhinney (Ed.), *The emergence of language* (pp. 197–213). Mahwah, NJ: Lawrence Erlbaum Associates.

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

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

Goodall, J. (1979). Life and death at Gombe. *National Geographic, 155*, 592–620.

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

Grodzinsky, J., and Reinhart, T. (1993). The innateness of binding and coreference. *Linguistic Inquiry, 24*, 187–222.

Gvozdev, A. N. (1949). *Formirovaniye u rebenka grammaticheskogo stroya*. Moscow: Akademija Pedagogika Nauk RSFSR.

Halliday, M. (1975). *Learning to mean: explorations in the development of language*. London: Edward Arnold.

Halliday, M., and Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hardwick, D., McIntyre, C., and Pick, H. (1976). The content and manipulation of cognitive maps in children and adults. *Monographs of the Society for Research in Child Development, 41*, Whole No. 3.

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

Hausser, R. (1999). *Foundations of computational linguistics: Man-machine communication in natural language*. Berlin: Springer.

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

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language, 75*, 244–285.

Heine, B. (1997). *Cognitive foundations of grammar*. New York: Oxford University Press.

Heine, B., Güldemann, T., Kilian-Hatz, C., Lessau, D., Roberg, H., Schladt, M., et al. (1993). Conceptual shift: a lexicon of grammaticalization processes in African languages. *Afrikanistische Arbeitspapier Köln, 34*, 1–112.

Hermer-Vazquez, L., Moffet, A., and Munkholm, P. (2001). Language, space, and the development of cognitive flexibility in humans: The case of two spatial memory tasks. *Cognition, 79*, 263–299.

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

Hinton, L., Nichols, J., and Ohala, J. (Eds.). (1994). *Sound symbolism*. Cambridge: Cambridge University Press.

Holloway, R. (1995). Toward a synthetic theory of human brain evolution. In J.-P. Changeux and J. Chavaillon (Eds.), *Origins of the human brain* (pp. 42–60). Oxford: Clarendon Press.

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

Hudson, R. (1984). *Word grammar*. Oxford: Blackwell.

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

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

Kaschak, M. P., and Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language, 43*, 508–529.

Kay, P., and Fillmore, C. J. (1999). Grammatical constructions and linguistic generalization: The "what's X doing Y?" construction. *Language, 75*, 1–33.

Keenan, J., and MacWhinney, B. (1987). Understanding the relation between comprehension and production. In H. W. Dechert and M. Raupach (Eds.), *Psycholinguistic models of production*. Norwood, N.J.: ABLEX.

Klatzky, R. L., Pellegrino, J. W., McCloskey, B. P., and 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., and Whishaw, I. Q. (1995). *Fundamentals of Human Neuropsychology. Fourth Edition*. New York: W. H. Freeman.

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

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

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

Lakoff, G., and Johnson, M. (1980). *Metaphors we live by*. Chicago: Chicago University Press.

Landau, B., and Jackendoff, R. (1993). "What" and "where" in spatial language and spatial cognition. *Behavioral and Brain Sciences, 16*, 217–265.

Landauer, T., and Dumais, S. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211–240.

Langacker, R. (1987). *Foundations of cognitive grammar: Vol. 1*. Stanford, CA: Stanford University Press.

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

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

Lewis, M. M. (1936). *Infant speech: A study of the beginnings of language*. New York: Harcourt, Brace and Co.

Lewis, R. (1998). Reanalysis and limited repair parsing: Leaping off the Garden Path. In J. D. Fodor and F. Ferreira (Eds.), *Reanalysis in sentence processing*. Boston: Kluwer.

Li, P., Burgess, C., and Lund, K. (2001). The acquisition of word meaning through global lexical co-occurrences. *Proceedings of the 23rd Annual Meeting of the Cognitive Science Society*, 221–244.

Lieven, E. V. M., Pine, J. M., and Baldwin, G. (1997). Positional learning and early grammatical development. *Journal of Child Language, 24*, 187–219.

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

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

Lust, B., and 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., and Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676–703.

MacNeilage, P. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21*, 499–546.

MacWhinney, B. (1974). *How Hungarian children learn to speak*. University of California, Berkeley.

MacWhinney, B. (1975). Pragmatic patterns in child syntax. *Stanford Papers And Reports on Child Language Development, 10*, 153–165.

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

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

MacWhinney, B. (1987). Toward a psycholinguistically plausible parser. In S.

Thomason (Ed.), *Proceedings of the Eastern States Conference on Linguistics*. Columbus, Ohio: Ohio State University.

MacWhinney, B. (1988). Competition and teachability. In R. Schiefelbusch and M. Rice (Eds.), *The teachability of language* (pp. 63–104). New York: Cambridge University Press.

MacWhinney, B. (2003). The gradual evolution of language. In B. Malle and T. Givón (Eds.), *The evolution of language*. Philadelphia: Benjamins.

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

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

Maia, T., and Chang. (2001). Grounded learning of grammatical constructions. *2001 AAAI Spring Symposium on learning grounded representations*.

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

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

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

McDonald, J. L., and 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.

Meltzoff, A. N. (1988). Infant Imitation and Memory: Nine-Month-Olds in Immediate and Deffered Tests. *Child Development, 59*, 217–225.

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

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

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

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

Mitchell, D. C. (1994). Sentence parsing. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego, CA: Academic Press.

Narayanan, S. (1997). Talking the talk is like walking the walk. *Proceedings of the 19th Meeting of the Cognitive Science Society*, 55–59.

O' Grady, W. (2002). An emergentist approach to syntax.

O'Grady, W. (1997). *Syntactic development*. Chicago: Chicago University Press.

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

Osman, A., Albert, R., and Heit, M. (1999). *Motor cortex activation during overt, inhibited, and imagined movement*. Paper presented at the Psychonomics, Los Angeles.

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

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

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

Passingham, R. (1993). *The frontal lobes and voluntary action*. Oxford: Oxford University Press.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., and 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.

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

Postal, P. (1971). *Cross-over phenomena*. New York: Holt, Rinehart, and Winston.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences, 22*, 253–336.

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

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

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

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

Rizzolatti, G., Fadiga, L., Gallese, V., and Fogassi, L. (1996). Premotor cortex

and the recognition of motor actions. *Cognitive Brain Research, 3*, 131–141.

Rumelhart, D. E. (1975). Notes on a schema for stories. In Bobrow and A. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press.

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

Savage-Rumbaugh, E., and Taglialatela, J. (2001). Language, apes, and understanding speech. In T. Givón (Ed.), *The evolution of language*.

Schank, R., and Abelson, R. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, N. J.: Lawrence Erlbaum.

Schieffelin, B. (1985). The acquisition of Kaluli. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition. Volume 1: The data*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Searle, J. R. (1970). *Speech acts: An essay in the philosophy of language*. Cambridge: University Press.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*, 417–424.

Seliger, H. (1989). Semantic transfer constraints on the production of English passives by Hebrew-English bilinguals. In H. Dechert and M. Raupach (Eds.), *Transfer in language production* (pp. 21–34). Norwood: NJ: Ablex.

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

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

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

Smyth, R. (1995). Conceptual perspective-taking and children's interpretation of pronouns in reported speech. *Journal of Child Language, 22*, 171–187.

Snyder, L. H., Grieve, K. L., Brothcie, P., and Anderson, R. A. (1998). Separate body- and world-referenced representations of visual space in parietal cortex. *Nature, 394*, 887–891.

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

Solan, L. (1983). *Pronominal reference: Child language and the theory of grammar*. Boston: Reidel.

Solomon, K., and Barsalou, L. W. (2001). Representing properties locally. *Cognitive Psychology, 43*, 129–169.

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

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

Talmy, L. (2000). *Toward a cognitive semantics. Vol. 1: The concept structuring system*. Cambridge, MA: MIT Press.

Terrace, H. S., Petitto, L. A., Sanders, R. J., and Bever, T. G. (1980). On the grammatical capacity of apes. In K. Nelson (Ed.), *Children's language: vol. 2*. New York: Gardner.

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

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

Tomasello, M. (1999). *The cultural origins of human communication*. New York: Cambridge University Press.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*, 209–253.

Tucker, D. (2001). Embodied meaning: An evolutionary-developmental analysis of adaptive semantics. In B. Malle and T. Givón (Eds.), *The evolution of language*. Philadelphia: Benjamins.

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

van der Lely, H. (1994). Canonical linking rules: Forward vs. reverse linking in normally developing and Specifically Language Impaired children. *Cognition, 51*, 29–72.

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

Van Orden, G., Holden, J. G., Podgornik, M., and Aitchison, C. S. (1999). What swimming says about reading: Coordination, context, and homophone errors. *Ecological Psychology, 11*, 45–79.

Vendler, Z. (1957). Verbs and times. *Philosophical Review, 56*, 143–160.

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

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

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

Whistler, K. (1985). Focus, perspective, and inverse person marking in Nootkan. In J. Nichols and T. Woodbury (Eds.), *Grammar inside and outside the clause* (pp. 227–265). New York: Cambridge University Press.

Zubin, D. A. (1979). Discourse function of morphology: The focus system in German. In T. Givón (Ed.), *Syntax and semantics: Discourse and syntax* (Vol. 12). New York: Academic Press.

Zwaan, R. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*, 162–185.

Zwaan, R. A., Kaup, B., Stanfield, R. A., and Madden, C. J. (in press). Language comprehension as guided experience. *Behavioral and Brain Sciences*.