

## The Emergence of Grammar from Perspective

Brian MacWhinney

Carnegie Mellon University

[macw@cmu.edu](mailto:macw@cmu.edu)

In: The Grounding of Cognition: The role of perception and action in memory, Diane Pecher and Rolf A. Zwaan (Eds.)

Successful communication rests not just on shared knowledge and reference (H. Clark & 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; O' Grady, in press) 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 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 mental acts.
2. Language uses perspective taking to bind together these five imagery subsystems.

3. Grammar emerges from conversation as a method for supporting accurate tracking and switching of perspective.
4. By tracing perspective shifts in language, children are able to learn the cognitive pathways and mental models sanctioned by their culture.

This hypothesis builds on recent advances in cognitive linguistics, embodied cognition, cognitive neuroscience, anthropology, and developmental psychology. The Perspective Hypothesis represents a particular case of a more general type of analysis called “emergentism” (MacWhinney, 1999b, 2002). The general emergentist framework views emergence as operative on five time levels. The current article focuses on the online emergence of perspective during conversational processing, as highlighted in point (3). Developmental aspects of the emergence of perspective marking are discussed elsewhere (MacWhinney, 1999a).

The articulation of a theory of perspective is central to 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 cognitive systems governing direct experience, space-time deixis, causal plans, social roles, and mental acts. Because perspective interacts with imagery constructed on each of these levels, it provides a general rubric for knitting together all of cognition. Consider a sentence, such as “Last night, my sister’s friend reminded me I had dropped my keys under the table behind the garage.” Here, we see how a single utterance integrates information about time (last night), social relations (sister’s friend), mental acts (remind), space (under, behind), objects (keys, table, garage), and events (drop). Within each of these informational levels, there are perspectives, including those of the sister, the friend, the speaker, and the various locations. Although this information may be elaborated in different regions of the brain, language achieves an integration of this information across all of these domains. According to the Perspective Hypothesis, it does this by specifying shifts of perspective across each of the five levels.

Consider how it might that we would understand the sentence, “the cat licked herself.” In the basic superficial mode of processing, which I call the depictive mode, we

see a movie of the cat raising her paw to her mouth and licking the fur with her tongue. In a deeper, more embodied level of processing, which I call the enactive mode, we take the stance of the cat by mapping her paw to our hand and her tongue to our tongue. Most people would say that they are unlikely to employ the enactive mode, when 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.

### **Level 1 – Direct Experience**

Language codes direct experience through words for objects such as “banana” or “finger.” Our basic mode of interaction with objects is through direct experience and direct perception. Direct perceptions involve vision, touch, smell, taste, kinesthesia, and proprioception. These interactions take advantage of the affordances (Gibson, 1977) that objects provide for both perception and 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 facilitates quick adaptive reactions. Because this system does not need to rely on memory, imagery, perspective, or other cognition systems (Gibson, 1977), it can remain fully grounded.

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.”

### **Imagery and Decoupling**

The direct grounding of perception is limited and elaborated in three important ways. First, it is clear that we know things that do not derive from direct perception. Consider again the case of the banana. We know that bananas are rich in potassium and Vitamin E, that they are grown in Central America by United Fruit cooperatives, and so on. These are declarative facts (Paivio, 1971; Tabachneck-Schijf, Leonardo, & Simon, 1997) that elaborate the primary notion of a banana that we derive from direct embodied perception.

Second, direct perception is revised and redirected by decoupling through mental imagery. Direct experience can be captured and replayed through the system of mental imagery. When we imagine a banana, we call up images of its shape, taste, and feel, even when it is not physically present. Imagery serves a variety of cognitive functions in the areas of planning, memory, dreams, and perspective taking. 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) and Martin, Wiggs,

Ungerleider, and Haxby (1996) and Cohen et al. (1996) have shown that, when subjects are asked to engage in mental imagery, they use modality-specific 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. The imagery system relies on the cognitive creation of a body map (Damasio, 1999; Kakei, Hoffman, & Strick, 1999). This body map then functions as an internal homunculus that can be projected to track the actions of others through the system of motor neurons (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996).

Imagery works together with memory, planning, dreaming, and projection to allow us to decouple thought 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. In effect, imagery facilitates the partial decoupling of cognition from direct perception. At the same time, the decomposable nature of perceptual symbol systems (Barsalou, 1999) allows us to reestablish partial grounding for the purposes of comprehension, planning, and memory. However, the fact that cognition can become partially ungrounded through imagery should not be construed as meaning that it is fully ungrounded (Burgess & Lund, 1997).

Finally, there is evidence that top-down influences from memory can redirect the nature of direct perception. Phenomena such as the McGurk effect, apparent motion, amodal completion, and repetition priming all indicate that our intake of new perceptual material can be biased by concurrent perceptions or images derived from memory.

### **The Language of Direct Experience**

The mapping of direct experiences onto linguistic form is confined in English to the level of individual words, including nouns, verbs, and adjectives. The noun “banana” packages together all our experiences with this object into a single unanalyzed whole. 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. Adjectives such as “green” or “round” encode largely perceptual dimensions of direct experiences.

Polysynthetic languages can express more complex direct perceptions in single words. For example, in Navajo, a chair is “bikáá’dah’asdáhi” or “on-it-one-sits.” To take a more familiar example, many languages refer to a corkscrew as a “cork puller.” 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. Adjectives encode images of direct perceptions for attributes such as weight, color, or smell.

Much of the current work on embodied cognition focuses on the interpretation of language referring to direct perceptions. For example, Stanfield and Zwaan (2001) found that, when given sentences such as “John pounded the nail into the floor,” subjects construct interpretations with a nail pointing downwards. In these tests, subjects must develop full images of the relevant sentences. However, the actual dependent measures in the studies are limited to the imagined direct perceptions linked to the position of specific objects, such as the “nail” in this example.

## **Level 2 – Space and Time**

Perspective taking in space and time depends on a different set 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 mapping onto the position of another agent. Deictic centers (Duchan, Bruder, & Hewitt, 1995) can be constructed in three frameworks: egocentric, allocentric, and geocentric. Positions in these frames are coded through locative adverbs and prepositions.

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”.

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 (H. H. Clark, 1973; Bryant, Tversky, & Franklin, 1992).

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.

Shifts in spatial perspective can lead to strange alternations of allocentric reference. 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.

### **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 event times (ET) that have a location in relation 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.

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 is relatively more difficult (Zwaan, 1996). However, in practice, 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.

Formal methods for calculating time also allow us to construct the temporal analog to the geocentric frame. For example, we can use moments such as New Year's Day, the birth of Christ, noon, and midnight as absolute reference points from which we compute time forward and backward. At with the geocentric spatial frame, we can shift between these calendrocentric frames by telescoping from minutes, to hours, days, months, years, and centuries.

### **Level 3 – Events**

The basic unit of cognition on Level 3 is the event. Events are chained together to encode long event sequences or plans. Events and plans involve the linkage of a variety of actions on objects. For example, we might devise a plan to clean up the house that will involve a variety of operations using brooms, vacuums, sponges, and fluids. In addition, our plans may involve other people with whom we work in parallel and in cooperation.

The execution and tracking of these complex plans requires not only perspective taking, but also perspective shifting. These shifts involve new combinations of actors, actions, and 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).

Complex plans are composed of individual events, each organized from a particular perspective. Even when we maintain a single overall causal perspective across a series of events, we still make brief shifts to secondary perspectives. When describing how to assemble a bench, we might say, “Take the long segment of the rear fence guard and insert the black plastic guide screw until it is parallel to the bottom of the guard; then align the guard perpendicular to the right edge of the moveable brace.” Here, we first take the perspective of the person doing the assembly, while shifting secondary attention first to the rear fence guard and then the guide screw. Then the guide screw itself becomes the perspective for a moment, until we then shift back to the perspective of the person doing the assembly, for the verb “align.” Note that the shift of perspective to the guide screw was prepared by its receipt of secondary attention as the object of “insert.” In other contexts, we may make even stronger shifts between actors, as when a football announcer describes a play by first taking the perspective of the quarterback, then the rusher, then the receiver, and finally the defense tackler.

In order to segment reality into separate events, language and cognition provide us with a system that orders nouns into role slots constellated 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 (Hudson, 1984; MacWhinney, 1988; Hausser, 1999; Kay & Fillmore, 1999) 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 slot-filling mechanism is used recursively 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).

Item-based patterns are the building blocks of larger clauses. Using item-based patterns, adjectives and other modifiers combine with nouns to form noun phrases. These phrases then attach to each other and to verbs using prepositions and other operators. Conjunctions, complementizers, and relativizers then combine clauses into complex sentences. In order to track shifts and flows of perspective through these complex structures, language provides us with a wide array of grammatical structures and cues including passivization, clefting, dislocation, coreference, reflexivity, obviation, possession, quantification, scope, ergativity, relativization, subordination, ellipsis, coordination, agreement, case marking, and word order placement. These systems are primarily sensitive to Level 3 causal chains, but they also encode Level 2 space-time structures, as well as some of the role and mental act structures we will discuss in the final two sections. The next five subsections focus on examining how five specific syntactic processes are shaped by the impact of perspective shifting. The five processes are ambiguity, relativization, pronominal co-reference, reflexivization, and clitic assimilation. I am selecting these five processes as illustrations because they are easily accessible and have figured heavily in both the linguistic and psycholinguistic literature. However, the analyses I offer here can be developed equally well for all major grammatical constructions.

### **Ambiguity**

Syntactic ambiguities and garden paths arise from competition (MacWhinney, 1987; MacDonald, Pearlmutter, & Seidenberg, 1994) between alternative perspectives. Consider the example of sentence (4) below. In this sentence, we tend to assume the first noun is the perspective of the participial “visiting”, yielding the interpretation that “if relatives visit you, they can be a nuisance.” At the same time, we are also able to imagine that some unspecified person is the perspective of “visit” with the relatives as the object, yielding the interpretation that “it can be a nuisance to pay a visit to one’s relatives.” In (5), on the other hand, the pronoun “they” prepares us to adopt the shifted perspective. In

example (6), because the verb “cry” is intransitive, the only possible interpretation is the one with “babies” as the perspective of “crying.”

4. Visiting relatives can be a nuisance.
5. If they arrive in the middle of a workday, visiting relatives can be a nuisance.
6. Crying babies can be a nuisance.

In (7), 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 (8), although again we can maintain the perspective of “Brendan” if we wish.

7. Brendan saw the Grand Canyon flying to New York.
8. Brendan saw the dogs running to the beach.

In cases of prepositional phrase attachment competitions, such as (9), 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.

9. The women discussed the dogs on the beach.
10. The women discussed the dogs chasing the cats.

In (10), on the other hand, we have a harder time imagining that the women, instead of the dogs, are chasing the cats.

The starting point or initial nominal phrase (if there is one) is always the default perspective. In most English sentences, this means that the perspective is the subject of the verb. 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 (8) and (10). 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. Examples (11)-(13) illustrate this effect:

11. Although John frequently jogs, a mile is a long distance for him.
12. Although John frequently jogs a mile, the marathon is too much for him.
13. Although John frequently smokes, a mile is a long distance for him.

Detailed self-paced reading and eye-movement studies of sentences like (11), 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 reflecting what happens during the time spent in shifting to a new perspective when the cues preparing the processor for the shift are weak. Examples, such as (5) and (11-13), show that perspectival shifting is an integral part of online, incremental sentence processing (Marslen-Wilson & Tyler, 1980).

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

14. Someone loves everyone.

15. Everyone is loved by someone.

If we take the perspective of “someone” in (14), 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 (15), because in that sentence “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. Further examples of this type include perspective shifts in numerical quantification, such as (16) and (17):

16. Two students read three books.

17. Three books are read by two students.

In (16) assumption of the perspective of the starting point allows us to imagine that the two students are reading the same three books. If, on the other hand, we process the quantifier scoping by assuming the perspective of the books, then we can imagine that there would be a total of six students reading the books.

Perspective shift theory also allows us to understand why (18) is acceptable and (19) is not. In (18) the perspective of every farmer is distributed so that each of the farmers ends up owning a well-fed donkey. In this perspective, there are many donkeys. Sentence

(19), on the other hand, forces us to break this distributive scoping and to think suddenly in terms of a single donkey, which violates the mental model set up in the main clause.

18. Every farmer who owns a donkey feeds it.

19. \*Every farmer who owns a donkey feeds it, but will it grow?

### Relativization

Restrictive relative clauses provide further evidence of the impact of perspective shifting on sentence processing difficulty. Processing these structures can require us to compute multiple shifts of perspective. Consider these four types of restrictive relative clauses:

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 true perspective switches in the SS relative type. In the OS type, perspective flows from the main clause subject (dog) to the main clause object (cat) in accord with the general principle of partial shift of perspective to the object. At the word “that” perspective then flows further to “the cat” as the subject of the relative clause. This perspective shift is made less abrupt by the fact that “cat” had already received secondary focus before the shift was 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.

The perspective account predicts this order of difficulty:  $SS > OO = OS > SO$ . Studies of both acquisition (MacWhinney, 1982) and adult processing (MacWhinney & Pléh, 1988) have provided support for these predictions. For example, a reaction time study of Hungarian relative clause processing by MacWhinney and Pléh (1988) shows

how the processing of perspective operates in a language with highly variable word order. 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 the 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 opposite order of difficulty arises: 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:

- 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. This processing 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.

Sentences with multiple center 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. But note that the mere stacking of nouns by itself 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 sloppy 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

### Pronominal Co-reference

Perspective taking also plays a central role in shaping the grammar of pronominal co-reference. Consider sentences (20) and (21). Coreference between “he” and “Bill” is possible in (21), but blocked in (20).

20. \* He<sub>i</sub> says Bill<sub>i</sub> came early.

21. Bill<sub>i</sub> says he<sub>i</sub> came early.

Note that the pronoun “he” in (20) and (21) can refer to someone mentioned outside of the sentence such as “Tom.” What is specifically blocked in (20) is coreference between “he” and “Bill” as indicated by their subscripts. The theory of Government and Binding (Reinhart, 1981; Chomsky, 1982; Grodzinsky & Reinhart, 1993) seeks to explain this phenomenon and a wide variety of related phenomena in pronominal co-reference in terms of structural relations in a phrase-marker tree. Principle C of the binding theory holds that a pronoun can only be bound to a referent in the clause through a c-command relationship. An element is said to c-command another element if it stands in a direct chain above it in a phrase tree. In (20), “Bill” is too low in the tree to c-command the pronoun. As a result, Principle C excludes a co-referential reading for (20), but not for (21). In (21) “Bill” c-commands the pronoun because it stands in a direct chain of dominance above it in the tree. As a result, the pronoun can be bound to the referent “Bill” in (21).

The Perspective Hypothesis attributes the unavailability of the co-referential reading of (20) to a very different set of forces. The initial claim of the perspective hypothesis 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. In (20), the listener must relate the initial pronoun to some already established discourse entity. Since “Bill” is not yet available, the listener is forced to



assume that “he” refers to some previously mentioned actor. In the case of (21) on the other hand, “Bill” is available as a referent and therefore “he” can co-refer to “Bill.”

The blockage of coreference in (20) is not a simple matter of linear order, since co-reference between a pronoun and a following noun is possible, when the pronoun is in an initial subordinate clause. Consider the contrasts between these four sentences, where the asterisk on (24) indicates that “he” cannot be co-referential with “Lester.”

- 22. Lester<sub>i</sub> started to feel dizzy, when he<sub>i</sub> drank the vodka.
- 23. When he<sub>i</sub> drank the vodka, Lester<sub>i</sub> started to feel dizzy.
- 24. \*He<sub>i</sub> started to feel dizzy, when Lester<sub>i</sub> drank the vodka.
- 25. When Lester<sub>i</sub> drank the vodka, he<sub>i</sub> started to feel dizzy.

In (22) and (23), “Lester” c-commands the pronoun, since the subordinate clause is a branch of the VP. As a result, coreference is possible, even if the subordinate clause occurs at the beginning of the sentence, as in (23). In (24) and (25), on the other hand, “Lester” no longer c-commands the pronoun, and coreference should be blocked. However, the acceptability of (25) is a problem for this version of binding theory. Reinhart (1983) explains the anomaly by arguing that coreference in (25) is supported by discourse constraints.

The Perspective Hypothesis offers a somewhat different account for this pattern. It attributes the acceptability of coreference in (22) and (25) to the fact that the reference “Lester” has already been mentioned before the pronoun is encountered. It attributes the acceptability of coreference in (23) to the fact that the subordinating conjunction “when” gives the processor instructions that a subsequent NP can be used for co-reference to “he.” In (24), no such instructions are available and coreference is blocked by the fact that the pronoun appears in initial position, as in (20). We can state these two principles in the following form:

- 26. Perspective Referentiality: Each clause needs to be organized from the viewpoint of a perspective. In English, the perspective is given by the first nominal. If that nominal is a pronoun, it must be bound to a noun previously mentioned in the sentence or discourse and is not available for binding to following nominal referents. This requirement applies in somewhat weaker form to direct and indirect objects.

27. Cues for Cataphora. Cues that emphasize backgrounding or ongoing relevance allow an perspectival pronoun to maintain its candidacy for cataphoric binding (i. e. binding to following nominal referents).

The Perspective 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:

28. \*He<sub>i</sub> often said that Bill<sub>i</sub> was crazy.  
 29. ? John often told him<sub>i</sub> that Bill<sub>i</sub> was crazy.  
 30. ? John often said to him<sub>i</sub> that Bill<sub>i</sub> was crazy.  
 31. His<sub>i</sub> new dog licked Bill<sub>i</sub>.  
 32. The students who studied with him<sub>i</sub> enjoyed John<sub>i</sub>

By the time we reach elements that are no longer in the main clause, as in (31) and (32), cataphora 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.

Principle C of the binding theory can account for some of these patterns. For example, the acceptability of (32) above is in conformity with the fact that there is no c-command relation between “him” and “John.” However, the acceptability of (31) is not. Because both the binding theory and the Perspective Hypothesis provide a central role for the perspective/subject, it is not surprising to find that their predictions are often similar. The two accounts differ most clearly for patterns that are outside of the scope of core syntactic patterning. Consider this pair:

33. She<sub>i</sub> had just come back from vacation, when Mary<sub>i</sub> saw the stack of unopened mail piled up at her front door.  
 34. \*She<sub>i</sub> came back from vacation, when Mary<sub>i</sub> saw the stack of unopened mail piled up at her front door.

The presence of “had just” in (33) 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.” This is a further instance of the principle of Cues for Cataphora.

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

- 35. \*Near John<sub>i</sub>, he<sub>i</sub> keeps a laser printer.
- 36. Near John's<sub>i</sub> computer desk, he<sub>i</sub> keeps a laser printer.
- 37. \*He<sub>i</sub> keeps a laser printer near John<sub>i</sub>.
- 38. \*He<sub>i</sub> keeps a laser printer near John's<sub>i</sub> computer desk.

In (36) we have enough conceptual material in the prepositional phrase to enactively construct a temporary perspective for “John.” In (35) 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 (37) and (38) above. Co-reference is clearly blocked in (37) and (38), despite the fact that it is possible in (36). This indicates that linear order is important for the establishment of perspective and that (36) does not “derive” in any direct sense from (38). These examples motivate a third principle of the Perspective Hypothesis account for coreference.

- 39. Perspective Promotion: A nominal in a backgrounded prepositional phrase is so low in perspective that it cannot be a coreferent. However, additional cues of current relevance and perspectival action can elevate its status to allow it to become a candidate coreferent.

These sentences from Reinhart (1983) provide further examples of aspectual effects on perspective taking.

- 40. In Carter's<sub>i</sub> hometown, he<sub>i</sub> is still considered a genius.
- 41. ? In Carter's<sub>i</sub> hometown, he<sub>i</sub> is considered a genius.

Although both of these sentences can be given co-referential readings, it is relatively easier to do so for (40), because “still” serves as a cue for cataphora that forces perspective promotion in the preposed prepositional phrase.

Just as markers of ongoing relevance such as “had just” or “still” can promote the candidacy of a pronoun in a main clause for cataphora, so indefinite marking on a

nominal can decrease its candidacy for co-reference, as indicated by the comparison of (42) with (43).

42. While Ruth argued with the man<sub>i</sub>, he<sub>i</sub> cooked dinner.

43. ? While Ruth argued with a man<sub>i</sub>, he<sub>i</sub> cooked dinner.

44. While Ruth was arguing with a man<sub>i</sub>, he<sub>i</sub> was cooking dinner.

The functionalist literature has long recognized the fact that when a new nominal is first introduced as indefinite, it is a poor candidate for coreference. The addition of an aspectual marker of current relevance in (44) overcomes the effect of indefiniteness in (43), 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. These effects illustrate the following two principles:

45. Givenness: Indefinite nominals are relatively poor candidates for coreference.

However, their candidacy can be promoted by cues for ongoing relevance and perspectival focusing.

46. Cue Summation: In accord with the Competition Model (McDonald & MacWhinney, 1989) account, the candidacy of a noun for coreference is the product of the cues in favor of its candidacy over the product of all cues present.

Strong crossover (Postal, 1971) sentences provide further illustrations of these basic principles. In these sentences, the initial wh-word (who) indicates the presence of information that needs to be identified. In (47) 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. However, when the pronoun is not in the crucial subject role, co-reference or crossover between the wh-word and the pronoun is possible, as in (49) and (50).

47. \*Who(m)<sub>i</sub> does he<sub>i</sub> like most?

48. Who<sub>i</sub> likes himself<sub>i</sub>/\*him<sub>i</sub> most?

49. Who<sub>i</sub> thought that Mary loved him<sub>i</sub>?

50. Who<sub>i</sub> likes his<sub>i</sub> mother most?

In (48) coreference is possible for “himself” but not “him” in accord with the principles of perspective flow in reflexives discussed in the next section.

### **Reflexivization**

The original claim of the binding theory was that non-reflexive personal pronouns such as “he” or “him” are bound primarily to referents in higher clauses, whereas reflexive pronouns such as “himself” are bound to other elements within the same clause. Two decades of further research (Zribi-Hertz, 1989; Tenny & Speas, 2002) have called this initial characterization into question without providing a satisfactory alternative. However, the Perspective Hypothesis provides a promising way of understanding the wide range of phenomena related to reflexivization. The analysis I will present here, as well as my selection of example sentences, depends heavily on input from my colleague Carol Tenny whose analysis in Tenny (2002) agrees in most regards with what I am presenting here.

We should begin by noting that the core claim of the binding theory – that clausemates much be reflexivized, does a good job of accounting for the contrasts such as (51) and (52).

51. \*John<sub>i</sub> kicked him<sub>i</sub>.

52. John<sub>i</sub> kicked himself<sub>i</sub>.

Despite the success of this basic principle, there are many structures, even within a single clause, that permit coreference without reflexivization. Consider the following examples:

53. Phil hid the book behind him/himself.

54. Phil ignored the oil on him/himself\*.

In (53), both anaphoric and reflexive coreference are possible. In (54) anaphoric reference is possible, but reflexive reference is more difficult. The Perspective Hypothesis accounts for this difference in terms of the principle of Perspective Flow. In (53), once the act of hiding is completed, our perspective shifts back to “Phil” allowing us to view him still as the perspective and a candidate for reflexivization. In (54), on the other hand, once our attention has shifted to the “oil,” we have no particular reason to

refocus on “Phil.” The effect of perspective flow on reflexives can be summarized in the following two principles:

55. Perspective Flow and Short Distance Reflexives: Within a clause, coreference to the principle perspective must be marked by a reflexive. Anaphoric reference is possible if intervening introduction of a secondary perspective blocks the domination of the initial perspective.

56. Perspective Flow and Long Distance Reflexives: Outside a clause, reflexive coreference is possible if the clause-external referent is still highly foregrounded and perspectival.

Let us consider some further examples of these effects. One clause-internal domain that permits both reflexive and anaphoric coreference is the domain of complex noun phrases with representational nouns such as “picture,” “story,” or “report.” These nouns shift perspective from the main clause to the representation of the referent in the embedded noun phrase, as in (57). Because perspective has been shifted, anaphoric coreference with “him” becomes possible, since “John” is no longer a totally commanding perspective. However, (58) illustrates how this shift of perspective also depends on the shape of the activity in the main clause. In (58) the action of telling is more dynamic than the action of hearing in (57). As a result, anaphoric reference is blocked in (58).

57. John<sub>i</sub> heard a story about him<sub>i</sub>/himself<sub>i</sub>.

58. Max<sub>i</sub> told a story about \*him<sub>i</sub>/himself<sub>i</sub>.

The facilitation of anaphoric reference in (57) is not simply a function of placement of the pronoun in a prepositional phrase, as (59) illustrates. In this example, anaphoric coreference is blocked by the fact that the head of the noun phrase “Mary” has itself become perspectival.

59. John talked to Mary<sub>i</sub> about \*her<sub>i</sub>/herself<sub>i</sub>.

The presence of intervening perspectives facilitates the use of short distance pronouns that would otherwise be blocked by reflexives. Consider some further examples:

60. John<sub>i</sub> saw a snake near him<sub>i</sub>/himself<sub>i</sub>.

61. Jessie<sub>i</sub> stole a photo of her<sub>i</sub>/herself<sub>i</sub> out of the archives.

In the classic example (60) from Lakoff (1974) , the shift of perspective to the “snake” is enough to permit anaphoric coreference, although reflexive coreference is also possible. In (61) a similar shift is induced by the representational noun “photo.” However, it is not always the case that an intervening noun will shift perspective enough to permit anaphoric reference, as examples (62) - (64) illustrate.

62. Bill<sub>i</sub> dragged the box behind him<sub>i</sub>/himself<sub>i</sub>.

63. Bill<sub>i</sub> dragged the box toward him<sub>i</sub>/himself<sub>i</sub>.

64. Bill<sub>i</sub> dragged the box to \*him<sub>i</sub>/himself<sub>i</sub>.

65. Bill dragged the box on \*him/himself.

In (62) the preposition “behind” identifies Bill as a location, thereby causing enough perspective shift to license the short distance pronoun. In (63), the preposition “toward” activates the role of Bill as goal, again establishing a new perspective. In (64) and (65), on the other hand, the prepositions “to” and “on” simply specify the shape of the action and fail to refocus perspective enough to license the short distance pronouns.

In examples (57-65) there is an intervening noun that can facilitate the shift of perspective. However, in (66) and (67) the licensing of anaphoric coreference occurs without this shift.

66. John<sub>i</sub> signaled behind him<sub>i</sub>/himself<sub>i</sub> to the pedestrians.

67. Bill<sub>i</sub> pointed next to him<sub>i</sub>/himself<sub>i</sub> at the mildew on the roses.

In these sentences, the verbs themselves trigger a strong shift of perspective away from the subject, drawing attention to other objects through the acts of signaling and pointing.

We also need to consider another group of predicates that, like (58) and (64) fail to license anaphoric coreference. These are illustrated in sentences (68) and (69).

68. Max twisted the knife into \*him/himself.

69. Margaret pinned the nametag to \*her/herself.

70. Mary painted a portrait of \*her/herself.

In these examples, the perspective continues to maintain active control of the action, despite the presence of an intervening object. Because of this, there is not enough shift in perspective to permit anaphoric coreference. However, if attention is shifted away from the causer to the path itself, as in (71), anaphoric coreference is possible.

71. Max twisted the knife partway into him/himself.

Finally, let us consider a set of constructions in which perspective shift is induced by the presence of other refocusing devices. These devices can include evaluative adjectives such as “beloved” or “silly” as illustrated in (72) and (73), definite markers as in (74) and (75), and further specification of the complex noun phrase as in (76) and (77).

72. Jessie stole a photo of \*her/herself out of the archives.

73. Jessie stole a silly photo of her/herself out of the archives.

74. Anna hid a snapshot of \*her/herself under the linoleum.

75. Anna hid the snapshot of \*her/herself under the linoleum.

76. Lucie talked about the operation on \*her/herself.

77. Lucie talked about the operation on her/herself that Dr. Edward performed.

### **Clitic assimilation**

As a final example of the impact of perspective taking on grammar, let us consider the process of clitic assimilation. In English, the infinitive “to” often assimilates with a preceding modal verb to produce contractions such as “wanna” from “want to” in cases such as (79). However, this assimilation is blocked in environments like the one in (80), making (81) unacceptable.

78. Why do you want to go?

79. Why do you wanna go?

80. Who(m) do you want to go?

81. \* Who(m) do you wanna go?

According the binding theory (Chomsky, 1981), the blocking of the assimilation in (80-81) 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 perspective flow. According the Perspective Hypothesis, cliticization is possible in (79) because the perspective of the higher clause is maintained as the perspective of the complement. In (81), on the other hand, perspective undergoes a forced processing shift from “who(m)” to “you” and then back to “who(m).” These perspective shifts block cliticization.

Perspective can also shift to implicit external controllers. Compare examples (82) and (83) below in which the infinitive does not cliticize with (84) where it does.



82. I get ta go. (Privilege)

83. I got ta go. (Privilege)

84. I gotta go. (Obligation)

In the case of (84), the first person subject has an immediate obligation to fulfill, whereas in (82) and (83), 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 (82) and (83), than it is in (84). According to the Perspective Hypothesis, cliticization occurs when a motivated subject engages directly in an action. When there is a shift to another actor, or a conflict of perspectives, cliticization is blocked.

### **Levels 4 and 5 – Social roles and mental acts**

Our analysis in this paper has focused on the grammar of the clause, as reflected in basic structures governing perspective identification in direct experience, space-time deixis, and clausal action. Examples such as (78-84) show that even core grammatical structures can reflect social role relations and aspects of the theory of mind. Because of space limitations, we cannot analyze the effects of these higher levels on grammar in detail here. However, it may be useful to draw a bit of attention to some of the more obvious ways in which social roles and mental acts impact grammar and discourse.

First, it is import to note that single lexical items characterize many complex social roles and mental acts. Items like “libel,” “Internet,” or “solidarity,” encode social scenarios organized about the perspective of social actors (Barsalou & Wiemer-Hastings, this volume). Let us take the noun “libel” as an example. When we speak of some communication as being “libelous,” we mean something like the following. The person using the word “libel” is taking the perspective of an “accused” person who declares to some general audience that the (purported) libeler has asserted that the accused has engaged in some illegal or immoral activity. Moreover, the accused wishes to convince the general audience that the libeler’s claims are false and designed to make the audience think poorly of the accused in ways that influence the his or her ability to function in public life involving the general audience. In fact, the full legal characterization of libel is more complex than this, but the everyday use of the word “libel” has roughly this basic form. This single word conveys a complex set of interacting and shifting social

perspectives. To evaluate whether or not a statement is libelous, we have to assume the perspective of the accused, the purported libeler, and the audience to evaluate the various claims and possible counterclaims. All of this requires continual integration and shifting of social roles and mental acts.

Second, language makes extensive use of kinship terms, appellations, and pronouns to characterize social roles. The decision about whether to call someone “you,” “your Honor,” “Mary,” or “Mrs. Smith” depends on a complex system of role evaluation. In other languages, these distinctions can extend to influencing a wide range of grammatical structures, such as in the Japanese verb sets that mark three levels of honorific relations.

Third, verbs like “promise,” “forgive,” “admire,” and “persuade” encode multiple relations of expectation, benefit, evaluation, and prediction between social actors. To evaluate the uses of these verbs requires flexible perspective taking and coordination. Within this larger group of mental state verbs, one dimension of contrast is known as “explicit causality.” Sentence (85) illustrates the use of the experiencer-stimulus verb “admire”; whereas sentence (86) illustrates the use of a stimulus-experiencer verb like “apologize.”

85. John admired Mary, because she was calm under stress.

86. John apologized to Mary, because he had cracked under stress.

McDonald and MacWhinney (1995) asked subjects to listen to sentences like (85-86), while making a crossmodal probe recognition judgment. Probe targets included old nouns (John, Mary) new nouns (Frank, Jill), old verbs (admire, apologize), and new verbs (criticize, resemble). The probes were placed at various points before and after the pronoun (“he” and “she”). The task was to judge whether the probe was old or new. McDonald and MacWhinney found that stimulus-experiencer verbs like “apologize” in (86) tend to preserve the reaction time advantage for the first noun (John) as a probe throughout the sentence. In terms of the perspective hypothesis, this means that perspective is not shifted away from the starting point in these sentences. However, experiencer-stimulus verbs like “admired” in (85) tend to force a shift in perspective away from the starting point (John) to the stimulus (Mary) right at pronoun. This leads to a period of time around the pronoun during which “Mary” has relatively faster probe recognition times. However, by the end of the sentence in (86), the advantage of the first

noun reappears. The fact that these shifts are being processed immediately on-line is evidence in support of the perspective taking account of sentence processing.

The implicit perspectives in verbs also influence the grammar of complementation. Smyth (1995) found that children in the age range between 5 and 8 have problems understanding co-reference in sentences like (87-91).

87. Minnie told Dorothy that she knew Superman.

88. Minnie told Dorothy that Superman knew her.

89. Minnie asked Dorothy if she knew Superman.

90. Minnie reminded Dorothy that she knew Superman.

91. Minnie told Dorothy that she made Superman cry.

Adults are able to maintain the viewpoint of the initial subject (Gernsbacher, 1990) even in the complement clause. However, children (Franks & Connell, 1996) process (87-91) in a very different way, being more likely to shift to the perspective of Dorothy. Adults reason that it makes little sense for Minnie to tell Dorothy about what she knows, since Dorothy should already have a pretty good view of the contents of her own mind. These social perspectives are nicely encoded in verbs such as “tell,” “ask,” or “remind.” For example, it does make sense to remind Dorothy about her knowledge, since reminding implies the possibility of forgetting. These various speech act verbs thus serve as models to the child of ways of structuring social interactions and theories of mind (Bartsch & Wellman, 1995).

## **Conclusion**

In this paper we have examined the ways in which the Perspective Hypothesis can offer new explanations for a variety of patterns in grammar and sentence processing. Elsewhere (MacWhinney, 1999a, 2002, 2003), I have discussed how this 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. Language allows us to integrate information from the domains of direct experience, space/time, plans, roles, and

mental acts. Across each of these dimensions, we assume and shift between perspectives in order to construct a fully human, unified conscious awareness.

## References

- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-660.
- Bartsch, K., & Wellman, H. (1995). *Children talk about the mind*. New York: Oxford University Press.
- Bryant, D. J., Tversky, B., & Franklin, N. (1992). Internal and external spatial frameworks for representing described scenes. *Journal of Memory and Language*, 31, 74-98.
- Burgess, C., & Lund, K. (1997). Modelling parsing constraints with high-dimension context space. *Language and Cognitive Processes*, 12, 177-210.
- 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., & Marshall, C. (1981). Definite reference and mutual knowledge. In B. W. A. Joshi & 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, 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.
- Damasio, A. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York: Harcourt Brace.
- Duchan, J. F., Bruder, G. A., & Hewitt, L. E. (1995). *Deixis in narrative: A cognitive science perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Franks, S. L., & Connell, P. J. (1996). Knowledge of binding in normal and SLI children. *Journal of Child Language*, 23, 431-464.
- 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 & 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.
- Grodzinsky, J., & Reinhart, T. (1993). The innateness of binding and coreference. *Linguistic Inquiry*, 24, 187-222.

- 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 & 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.
- Hudson, R. (1984). *Word grammar*. Oxford: Blackwell.
- Jeannerod, M. (1997). *The cognitive neuroscience of action*. Cambridge, MA: Blackwell.
- Takei, S., Hoffman, D. S., & Strick, P. L. (1999). Muscle and movement representations in the primary motor cortex. *Science*, 285, 2136-2139.
- Kay, P., & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalization: The "what's X doing Y?" construction. *Language*, 75, 1-33.
- Kuno, S. (1986). *Functional syntax*. Chicago: University of Chicago Press.
- Lakoff, G. (1974). Syntactic amalgams. In R. F. M. LaGaly & A. Bruck (Eds.), *Papers from the Tenth Regional Meeting*. Chicago: Chicago Linguistic Society.
- Langacker, R. (1995). Viewing in grammar and cognition. In P. W. Davis (Ed.), *Alternative linguistics: Descriptive and theoretical models* (pp. 153-212). Amsterdam: John Benjamins.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676-703.
- MacWhinney, B. (1977). Starting points. *Language*, 53, 152-168.
- MacWhinney, B. (1982). Basic syntactic processes. In S. Kuczaj (Ed.), *Language acquisition: Vol. 1. Syntax and semantics* (pp. 73-136). Hillsdale, NJ: Lawrence Erlbaum.
- 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 & M. Rice (Eds.), *The teachability of language* (pp. 63-104). New York: Cambridge University Press.
- MacWhinney, B. (1999a). The emergence of language from embodiment. In B. MacWhinney (Ed.), *The emergence of language* (pp. 213-256). Mahwah, NJ: Lawrence Erlbaum.
- MacWhinney, B. (2002). Language emergence. In P. Burmeister, T. Piske & A. Rohde (Eds.), *An integrated view of language development - Papers in honor of Henning Wode* (pp. 17-42). Trier: Wissenschaftlicher Verlag Trier.
- MacWhinney, B. (2003). The gradual evolution of language. In B. Malle & T. Givón (Eds.), *The evolution of language*. Philadelphia: Benjamins.
- MacWhinney, B. (Ed.). (1999b). *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum Associates.
- MacWhinney, B., & Pléh, C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, 29, 95-141.
- Marslen-Wilson, W. D., & Tyler, L. K. T. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1-71.

- Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, *379*, 649-652.
- McDonald, J. L., & MacWhinney, B. (1989). Maximum likelihood models for sentence processing research. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 397-421). New York: Cambridge University Press.
- McDonald, J. L., & MacWhinney, B. J. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language*, *34*, 543-566.
- 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. (in press). *Syntactic carpentry*. Mahwah, NJ: Lawrence Erlbaum Associates.
- 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.
- Postal, P. (1971). *Cross-over phenomena*. New York: Holt, Rinehart, and Winston.
- 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.
- Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131-141.
- Sacerdoti, E. (1977). *A structure for plans and behavior*. New York: Elsevier Computer Science Library.
- Smyth, R. (1995). Conceptual perspective-taking and children's interpretation of pronouns in reported speech. *Journal of Child Language*, *22*, 171-187.
- Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*, 153-156.
- Tabachneck-Schijf, H. J. M., Leonardo, A. M., & Simon, H. A. (1997). CaMeRa: A computational model of multiple representations. *Cognitive Science*, *21*, 305-350.
- Tenny, C., & Speas, P. (2002). Configurational properties of point of view roles. In A. DiSciullo (Ed.), *Asymmetry in Grammar*. Amsterdam: John Benjamins.
- van Hoek, K. (1997). *Anaphora and conceptual structure*. Chicago: University of Chicago Press.
- Vendler, Z. (1957). Verbs and times. *Philosophical Review*, *56*, 143-160.
- Zribi-Hertz, A. (1989). Anaphor vinding and narrative point of view: English reflexive pronouns in sentence and discourse. *Language*, *65*(4), 695-727.
- Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1196-1207.