

11

How Mental Models Encode Embodied Linguistic Perspectives

Brian MacWhinney

Humans demonstrate a remarkable ability to take other people's perspectives. When we watch movies, we find ourselves identifying with the actors, sensing their joys, hopes, fears, and sorrows. As viewers, we can be moved to exhilaration as we watch our heroes overcome obstacles; or we can be moved to tears when they suffer losses and defeats. This process of identification does not always have to be linked to intense emotional involvement. At a soccer match, we can follow the movements of a player moving in to shoot for a goal. We can identify with the player's position, stance, and maneuvers against the challenges offered by the defenders. We can track the actions, as the player drives toward the goal and kicks the ball into the net. This ability to take the perspective of another person is very general. Just as we follow the movements of dancers, actors, and athletes, we can also follow the thoughts and emotions expressed by others in language. In this paper, we will explore the ways in which language builds upon our basic system for projecting the body image to support a rich system of perspective tracking and mental model construction.

Projection

It is useful to think of projection as relying on the interaction of four systems, each of which is fundamental to psychological functioning. These four systems involve body image, localization, empathy, and perspective tracking.

Body Image Matching

In order to assume the perspective of another actor, one must first be able to construct a full image for one's own body. This image expresses not only the positions of our organs and limbs, but also their movements and configurations. For most of us, the notion of a body schema is something natural and inescapable. However, there are various neural disorders and injuries (Ramachandran, 2000; Ramachandran & Hubbard, 2001) that can lead to the disruption of the body image. The construction of the body image is based on processing in a wide variety of sites including the cerebellum (Middleton & Strick, 1998), medial prefrontal cortex (Macrae, Heatherton, & Kelley, 2004), primary motor cortex (Kakei, Hoffman, & Strick, 1999), and insula (Adolphs, 2003).

In order to achieve perspective taking, we need a system that can project our body image to other agents. The first step in this process must involve body part mapping. We have to identify specific organs or body parts on others that we then map onto parallel parts in our own body image. Meltzoff (1995) and others have traced the roots of body part matching back to infancy, when the baby can already demonstrate matching by imitating actions such as tongue protrusion or head motions. We have learned that, when a subject in an fMRI experiment tracks the movements of a particular limb or organ, there is corresponding activation of the same neural pathways the subject would use to produce those actions. 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 imag-

ery, 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. Similarly, when we observe pain in a particular finger in another person, transcranial magnetic stimulation shows activation of the precise motor area that controls the muscle for that same finger (Avenanti, Buetti, Galati, & Aglioti, 2005). The general conclusion from this research is that, once we have achieved body part matching, we can perceive the actions of others by projecting them onto the mechanisms we use for creating and perceiving these actions ourselves. In effect, we run a cognitive simulation (Bailey, Chang, Feldman, & Narayanan, 1998) and we match the products of this cognitive simulation to our perception of the actions of others. This ability to compute a cognitive simulation is at the heart of both projection and imagery.

From recent single-cell work with monkeys, we know that perceptual-motor matching is a very general aspect of the primate brain. Apart from the mirror neurons (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996) that were originally identified in the monkey counterpart to Broca's area, there are also perceptual-motor matching systems in other areas of both frontal and parietal cortex. The facility with which we are able to project perspective onto others suggests that the body matching system is probably quite extensive, but the exact extent of this system is not yet known.

Localization

In order to assume the perspective of another actor, one must also position that actor correctly in mental space. To do this we rely on the same system that allows us to locate our own bodies in space. This system must be robust against eye movements (Colby, this volume) and turns of the head (Ballard, Hayhoe, Pook, & Rao, 1997); it should allow us to maintain facing and update our position vis-à-vis landmarks and other spatial configurations (Klatzky & Wu; Proffitt, this volume). Once we have succeeded in projecting our body image to the other, we can next begin to position this projected image in the distal physical location (Loomis & Philbeck, this volume). One difficult part of this projection is the requirement that we rotate our body image 180° if the other person is facing us.

Empathy

In order to assume the emotional and social perspective of another, we must first be able to have access to our own emotions. We can then project or transfer these feelings to others. Of course, there is no doubt that humans have a well-organized system of emotions that map into physiological expression through the face, body, and autonomic nervous system. Once we have achieved projection of body image and spatial location, we can begin to identify with the emotions of the other person. Studies of patients with lesions (Adolphs & Spezio, 2006) have pointed to a role for the amygdala in the processing of emotional expression. Studies with normals using fMRI (Meltzoff & Decety, 2003) have shown activation in the amygdala and temporal areas (Pelphrey et al., 2003; Pelphrey, Morris, & McCarthy, 2005; Pelphrey, Viola, & McCarthy, 2004) for the processing of social expressions and actions.

Perspective Tracking

To obtain smooth projection, we must be able to constantly update, shift, and track the other actor on each of these dimensions (body image, spatial location, empathy). To do this, we have to treat the other body as obeying the same kinematic principles that govern our own body (Knoblich, this volume). At the same time, we must be able to distinguish the projected or imagined image from our own body image. This means that the tracking system must allow us to treat the projected image as “fictive” or “simulated” and our own body image as real. Maintaining this type of clear separation between simulation and reality would seem to require special neuronal adaptations. These adaptations include mirror neurons as one component, but they also require additional support from frontal areas for attention and regulation. This additional support is required to allow us to shift perspective between multiple external agents. For this system to function smoothly, we must store a coherent fictive representation of the other person. In order to switch smoothly between perspectives, it is crucial that the projected agent be easily accessed and constructed as a clearly separate fictive representation.

Researchers such as Donald (1991) and Givon (2002) have proposed that, two million years ago, *homo erectus* relied on a system of

gestural and mimetic communication to further social goals. Adaptations that supported this tracking of visual perspectives would have formed an evolutionary substrate for the later elaboration of perspective tracking in language (MacWhinney, 2005). There is still a close temporal and conceptual linkage between gesture and language (McNeill, 1992), suggesting that both systems continue to rely on a core underlying set of abilities to track perspective.

Subjective evidence from drama, movies, and literature suggests that identification is a dynamic and integrated process. This process produces a flowing match between our own motoric and emotional functions and the motoric and emotional functions of others. However, neurological evidence regarding the dynamic functioning of this system is still missing. We know that there is neuronal support for each of the four projection systems. However, we cannot yet follow perspective tracking in real time. In large part, this gap in our knowledge is a function of the limitations of measurements taken through methods such as fMRI and ERP.

The Perspective Hypothesis

Fortunately, there is evidence of a very different type that can help us understand the nature of ongoing projection and perspective taking. This is the evidence that is encoded in the structure of grammatical systems. Close analysis shows how grammar reflects the cognitive operations underlying perspective tracking. I refer to this general view of mental projection as the perspective hypothesis. We can decompose the overall perspective hypothesis into six more specific assumptions or claims:

1. Language functions by promoting the sharing of mental models between the speaker and the listener.
2. To build up complex mental models, referents and actions must be connected dynamically through temporal, spatial, and causal linkages.
3. The links in mental models are structured by perspective tracking.
4. Perspective tracking is supported by specific neuronal projection systems that keep these cognitive simulations separate from direct perception and action.

5. The primary function of grammatical devices is to mark perspective tracking, as it emerges in conversation and narrative.
6. Perspective tracking, as realized in language, facilitates the linking of mental systems and the cultural transmission of linked structures.

For psychologists and cognitive linguists, the first two assumptions are largely uncontroversial. Since the beginning of the Cognitive Revolution, researchers have assumed that comprehension involves the construction of linkages in complex mental models. In regard to these first two claims, the perspective hypothesis is simply building on traditional, well-supported, assumptions.

The traditional form of assumption 3 is that that discourse linkages involve coreference between nodes in semantic structure. For example, we could have a discourse composed of two simple propositions in a sentence such as: *the boy kicked the ball, and the ball rolled into the gutter*. In the mental model or semantic memory constructed from this sentence, the “ball” of the first clause is linked through coreference to the “ball” of the second clause. In classic models of sentence interpretation (Budi & Anderson, 2004; Kintsch & Van Dijk, 1978), the main mechanism for discourse linkage was coreference. Even models that decomposed semantic structure (Miller & Johnson-Laird, 1976) still maintained a reliance on coreference as the method for linking propositions in mental models.

The perspective hypothesis introduces a fundamental shift in our understanding of the construction of mental models. In this new view, which is encoded as assumption 3, propositions are embodied representations constructed from a specific perspective, which is initially the perspective of the sentential subject. We can refer to the combined interaction of perspective taking and perspective shifting as perspective tracking. Although coreference plays a secondary role in licensing linkage between propositions, mental models use perspective tracking as their primary integrating mechanism. In the case of our simple example sentence, this means that the perspective of the “boy” in the first sentence is shifted to that of the “ball” in the second clause and we then track the motion of the ball as it rolls into the gutter. In other words, we construct mental models by taking and shifting perspectives. Within this larger process of perspective tracking, deixis, or verbal pointing, plays a role of bringing new referents to our attention or locating referents in either working memory or long-term memory. However, the structuring of proposi-

tions into mental models depends primarily on taking the perspective of the entities discussed in a discourse and tracking this flow of perspective between the various discourse participants.

Assumption 4 is grounded on the growing evidence from cognitive neuroscience regarding specific neuronal systems that manage the projection of body image, spatial location, and emotion. Without making this assumption, it would be very difficult to imagine how one could believe that language processing relies intimately and continually on neuronal support for perspective tracking.

Assumption 5 is the centerpiece of the perspective hypothesis. When this assumption is linked to assumption 3, it takes on a particularly strong form. The combination of these two assumptions represents a novel position in cognitive science. Researchers who emphasize the formal determination of linguistic structure (Chomsky, 1975) have repeatedly rejected links between grammatical structure and “pragmatic” factors such as perspective taking. Although cognitive linguistics provides a role for perspective in the theory of subjectivisation (Stein & Wright, 1995), there is little acceptance of assumption 3. Despite the importance of embodied cognition to cognitive linguistics, there are no current models in this tradition that rely on perspective tracking as the major force integrating mental models. Within experimental psychology, recent work has focused on demonstrating the embodied nature of mental models. For example, Stanfield and Zwaan (2001) found that, when given sentences such as “John pounded the nail into the floor,” subjects are faster to name pictures of nails pointing downward than nails pointing sideways. This indicates that they construct interpretations with a nail pointing downward. Results of this type, summarized in the chapters in Pecher and Zwaan (2005), provide clear and important evidence for the embodied nature of mental models, but they tell us little about perspective tracking. To examine the course of perspective tracking experimentally, we will need to use online measures of interpretation, such as cross-modal naming, which will allow us to probe ongoing changes in mental models as they are being constructed.

Assumption 6 focuses on the consequences of perspectival mental models for promoting conceptual integration and cultural transmission. Perspectival mental models provide 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). The sentence produces an integrated tracking across the perspectives of the sister, the friend, the speaker, and the various locations. Although this information may be initially activated in different regions of the brain, language achieves an integration of this information across all of these domains.

The primary focus of this paper is on assumption 5. We will conduct this exploration in three parts. We will begin with a psycholinguistic account of how perspectives are shifted through sentence structures. Second, we will conduct a linguistic examination of a wide range of grammatical constructions to understand how they mark perspective tracking. Finally, we will consider the consequences of this analysis for theories of cognition and development in accord with assumption 6.

Perspective Tracking

Modern psycholinguistic research has tended to focus on the process of sentence comprehension or interpretation, rather than sentence production. It is much easier to achieve experimental control over comprehension than over production. Because our models of sentence interpretation are more detailed, it is easiest to explain the function of perspective tracking first for comprehension and to extend this account later to perspective marking in production. On the comprehension side, there are five important principles that have a direct bearing on perspective tracking.

1. *Incremental Interpretation.* The first principle, which has been widely supported in recent research, is the principle of incrementalism. According to this principle, the listener attempts to go from words to mental models as soon as material is unambiguously recognized. In processing terms, incrementalism is equivalent to the notion of cascading (McClelland, 1979) in which processes feed into each other as soon as they reach a point where data can be passed.
2. *Load Reduction.* In some cases, processing may involve placing phrases into working memory without yet committing to their grammatical role or without attaching them to other phrases.

However, if the listener can attach material to a fully interpreted mental model, then this will reduce processing load.

3. *Starting Points*. This principle of load reduction through attachment (Gibson, 1998) interacts with a third principle that governs the centrality of the starting point (MacWhinney, 1977) or the Advantage of First Mention (Gernsbacher, 1990). When we begin a sentence, we use the first nominal phrase as the basis for structure building (Gernsbacher, 1990). As we move through the sentence from left to right incrementally, we add to this structure through attachments. When a phrase or word cannot be attached, it increases the load. So, we are motivated to attach phrases as soon as possible to reduce these costs (O'Grady, 2005).
4. *Role Slots*. The process of phrasal attachment is driven by role slot assignments. In different models, this principle has very different names, varying from thematic role assignment to theta-binding. The basic idea (MacWhinney, 1987a) is that predicates (verbs, adjectives, prepositions) expect to attach to nominal arguments that fill various thematic or case roles, such as agent, object, or recipient. For languages that rely on SVO, SOV, or VSO orders, the first noun is placed tentatively into the role of the perspective. This perspective then actively searches for a verb that will allow it to assume a dynamic perspective. For example, in the sentence "the runner fell," we begin with the perspective of the runner as the starting point. We then move on incrementally to the verb *fall*. At that point, the linkage of "runner" to the role slot for a perspective for "falling" allows us to build a mental model in which the runner engages in the action of falling.
5. *Competition and Cues*. Role slot filling is a competitive process (MacWhinney, 1987a). In comprehension, several nominal phrases or "slot fillers" may compete for assignment to a given slot and role. Only one of the fillers can win this competition, and the losers must then compete for other slots. The outcome of the competition is determined by the presence of grammatical, lexical, and semantic cues that favor one or the other competitor. The process of cue summation obeys basic Bayesian principles.

At this point, it would perhaps be useful to think about how these five principles lay the groundwork for the Perspective Hypothesis. In a sense all of these principles are integrated by the basic need during comprehension to construct a mental model. The principles of incremental interpretation and load reduction are direct responses to the fact that there is a cost associated with maintaining unattached verbal chunks in short-term memory. In order to reduce this cost, we

adopt the initial hypothesis that the first nominal is the perspective. Because the construction of mental models is perspectival (assumption 3 above), we are able to take this starting point as the foundation for a larger mental model. The status of the starting point as the perspective is further supported if it can compete successfully for the perspective slot on the verb. In some marked constructions and word orders, it may lose out in this competition. But to understand how this happens, we need to look more closely at the dynamics of perspective tracking through grammatical constructions. Some of the forces that work to either maintain or shift this initial perspective include:

1. *Shift*. If the verb takes multiple thematic roles, then it may shift perspective from the starting point to secondary perspectives. This can occur for transitives, passives, clefts, and a variety of other constructions. In most cases, this shift is not complete, and the perspective of the starting point is at least partially maintained.
2. *Modification*. Individual referents may be further modified or elaborated by attached phrases and clauses. This occurs in relativization, complex NP formation, appositives, and other constructions.
3. *Maintenance*. Perspective can be maintained across clauses by devices (or cues) such as anaphoric pronouns, gerunds, resumptives, and conjunctions.

Structure vs. Function

Having now listed the general assumptions of the perspective hypothesis, we are in a position to explore the application of this hypothesis to a wide variety of syntactic constructions. However, before beginning that exploration, we need to consider a competing approach that accounts for some, but not all of the phenomena to be discussed. This approach, developed within generative linguistics over the last 50 years (Chomsky, 1957), which attempts to account for syntactic patterns in terms of structural relationships. Of the various structural relationships explored during this half century, perhaps the most prominent is the relation called *c-command*. We will therefore begin our exploration by comparing two different accounts of coreference—one based on the constraints imposed by *c-command* and another based on the constraints imposed by perspective tracking. This analysis is not intended to reject a possible role for *c-com-*

mand in linguistic description. Rather, the goal here is illustrate the impact of perspective tracking on grammatical constructions.

Coreference

We will begin our explorations with a consideration of a few selected aspects of the grammar of coreference. Consider sentences 1 and 2: Coreference between “he” and “Bill” is possible in (1), but blocked in 2.

1. *He_i says Bill_i came early.
2. Bill_i says he_i came early.

The coreferential reading of these sentences is marked by the presence of subscripts on the nominals. Without these subscripts and the coreference they require, the pronoun *he* in 1 can refer to someone mentioned outside of the sentence, such as “Tom.” What is specifically blocked in 1 is coreference between “he” and “Bill” as indicated by their subscripts. The perspective hypothesis accounts for the ungrammaticality of 1 by invoking the principle of referential commitment. When we hear “he” in 1, we need to take it as the starting point for the construction of a mental model for the clause. This means that we must make a referential commitment. We cannot wait until later to identify “he.” In (2), on the other hand, “Bill” is already available for coreference and so it is easy to link “he” to “Bill.”

The theory of government and binding (Chomsky, 1982; Grodzinsky & Reinhart, 1993; Reinhart, 1981) seeks to explain this phenomenon (and many others) in terms of structural relations in a phrase-marker tree. The backbone of this account is a relation known as *c-command*. Each element in the tree is said to “*c-command*” its siblings and their descendants. Principle C of the binding theory requires that lexical NPs such as *Bill* or *the man* be “free”; that is, not coindexed with a *c-commanding* element. This principle excludes a coreferential reading for 1 in which “Bill” is coindexed with the *c-commanding* pronoun, but not for 2 in which the intended antecedent *c-commands* the pronoun rather than vice versa.

Grammatical subjects are always in a syntactic position that allows them to *c-command* other elements in the same clause. This means that there is a very close parallel between the patterns expressed in *c-command* and the principle of referential commitment from the

perspective hypothesis. Because initial nominals serve as the bases for structure building and perspective propagation, it is initially difficult to distinguish the structural account from the functionalist account. However, if we look closely, we will see that there are a variety of phenomena that can be understood better in terms of mental model construction than in terms of c-command.

Noncentrality

One interesting contrast is the relative increase in acceptability of coreference that occurs as one moves from central to peripheral arguments. C-command blocks coreference in 3 and 4. This prediction of c-command is correct for 3. However, for many speakers, 4 is possible, although c-command disallows it.

3. * He_i said Bill_i was crazy.
4. * John told him_i that Bill_i was crazy.
5. John said to him_i that Bill_i was crazy.
6. John told his_i mother that Bill_i was crazy.

Although c-command makes the wrong predictions for 4, it correctly allows for coreference in 5 and 6, since the pronouns here do not c-command “Bill.”

The perspective hypothesis views this contrast in a very different way. As a further corollary of assumption 3 and its corollary of referential commitment, we have the following:

Principle of Noncentrality: If an element is not central to the building of a mental model, then it need not be referentially committed and is therefore open for backward anaphora. The less centrally involved the element, the more it is open for backward anaphora.

The most central argument is the subject (“he” in 3), followed by the object (“him” in 4), the oblique or prepositional object (“him” in 5), and finally a possessor in a complex NP (“his” in 6). As these roles become less and less central to the process of structure building, they become more and more open to backward anaphora.

Delaying Commitment

C-command provides no account for the grammaticality of sentences such as (7–9). This is because pronouns in subordinate clauses are too low in the structure to having NPs in other clauses as siblings.

7. After Lester_i drank the third vodka, he_i dropped his cup.
8. After he_i drank the third vodka, Lester_i dropped his cup.
9. *He_i drank the third vodka, and Lester_i dropped his cup.

Although both 8 and 9 have a pronoun that precedes the target noun “Lester,” the coreferential reading of 8 is easier to get than the coreferential reading of 10. Although *c-command* provides no account for this contrast, it can be explained by reference to the perspective hypothesis. To do this, we can rely on the following corollary derived from assumption 5.

Principle of Cue Marking: The default process of perspective tracking can be modified by grammatical cues that signal delays in referential commitment, clausal backgrounding, or shifts in perspective.

In the case of 8, the crucial grammatical cue is the subordinating conjunction *after*, which signals the beginning of a background clause. In the construction of a mental model, background material is placed “on hold” for later attachment to foreground material. However, the storage of the backgrounded initial clause of 8 does not incur a large processing cost, since its component pieces are fully structured. Moreover, as long as it remains in the background, the pronoun can be involved in backward coreference. Because there is no cue in 9 to protect the pronoun, it must be committed referentially and backward anaphora is blocked.

As a further illustration of the effect of grammatical cues in the delaying of commitment, consider the contrast between 10 and 11.

10. *She_i jumped inside the car, when Debra_i saw a large man lurking in the shadows.
11. She_i was jumping inside the car, when Debra_i saw a large man lurking in the shadows.

Here the presence of progressive aspect in 11 places the information in the main clause “on hold” in the mental model, because this information is being judged as relevant to the interpretation of the subsequent clause. In a series of experiments, Harris and Bates (2002) show that progressive marking leads to greater acceptance of backward anaphora. Thus, the progressive functions like the subordinating conjunction as a cue to delaying of commitment. A similar effect is illustrated by 12 and 13 from Reinhart (1983).

12. In Carter's_i hometown, he_i is still considered a genius.
 13. In Carter's_i hometown, he_i is considered a genius.

Here, it is easier to get a coreferential reading for 12 than for 13. This is because “still” serves as a cue that forces perspective promotion in the preposed prepositional phrase.

The opposite side of this coin involves the way in which indefinite reference blocks forward anaphora (i.e., the establishment of coreference between a full nominal and a following pronoun). It is somewhat easier to achieve forward anaphora in 14 than in 15.

14. While Ruth argued with the man_i, he_i cooked dinner.
 15. *While Ruth argued with a man_i, he_i cooked dinner.

In the case of 14, once we shift perspective from Ruth to “the man,” we now have a definite person in mind and it makes good sense to continue that perspective with “he.” In 15, on the other hand, we have no clear commitment to the identity of “a man” and using this unclear referent as the binder of “he” seems strange.

Reflexivization

The c-command relation is also used to account for patterns of grammaticality in the use of reflexive pronouns, such as “herself” or “myself.” The most common use of these pronouns is to mark coreference to a “clausemate,” which is often the subject of the current clause, as in 16 and 17.

16. *Mary_i pinched her_i.
 17. Mary_i pinched herself_i.

The perspective of the reflexive is a rather remarkable one, since it forces the actor to look back on herself as both the cause of the action and the recipient of the action at the same time.

When both referents are central arguments, reflexivization is mandatory. Sentences like 16 are impossible if the two nominals are coreferential. However, if one of the nominals is central to the process of sentence building, and if the other material in the sentence serves to shift perspective away from the starting point, then a

clausemate coreferent can use a nonreflexive pronoun. Sentences 18 and 19 illustrate this.

18. Phil_i hid the book behind him_i/himself_i.
19. Phil_i ignored the oil on him_i/himself_i*.

In 18, nonreflexive coreference and reflexive coreference are both possible. In 19 only anaphoric coreference is possible. This is because the act of hiding tends to maintain the causal perspective of “Phil” more than the act of ignoring. When Phil hides the book, it is still very much in his mind and so its position vis-à-vis his body still triggers self-reference. However, when Phil ignores the oil, it is no longer in his mind. At this point, the observation of the oil is dependent on an outside viewer and no longer subject to reflexivity.

Nouns such as *story* or *picture* can also trigger perspective shifting within clauses. In 20 and 21, reflexives are required if there is coreference, because there is no intervening material that shifts perspective away from either “John” or “Mary.” In 22, however, the nonreflexive is possible, since the story sets up a new perspective from which John is viewed as an actor in the story and not a listener to the story. In 23, the action of telling involves Max so deeply in the story itself that the full perspective shift is impossible and the non-reflexive cannot be used.

20. John talked to Mary_i about *her_i/herself_i.
21. John talked to Mary_i about *him_i/himself_i.
22. John_i heard a story about him_i/himself_i.
23. Max_i told a story about *him_i/himself_i.

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

24. John_i saw a snake near him_i/himself_i.
25. Jessie_i stole a photo of her_i/herself_i out of the archives.

The material that detracts from the reflexive perspective may also follow the pronoun, as in these examples from Tenny and Speas (2002).

26. John_i signaled behind him_i/himself_i to the pedestrians.
27. Bill_i pointed next to him_i/himself_i at the mildew on the roses.
28. Lucie_i talked about the operation on her_i/herself_i that Dr. Edward performed.

In these sentences, use of the nonreflexive prepares the listener for a shift of perspective following the pronoun. Without that additional material, the nonreflexive would be strange. Finally, perspective shift can also be induced by evaluative adjectives such as “beloved” or “silly,” as in these examples from Tenny and Speas (2002):

29. Jessie_i stole a photo of *her_i/herself_i out of the archives.
30. Jessie_i stole a silly photo of her_i/herself_i out of the archives.

In all of these cases, creation of an additional perspective can serve to shift attention away from the core reflexive relation, licensing use of a nonreflexive pronoun.

Ambiguity

Syntactic ambiguities and garden paths are typically described in terms of the construction of alternative structural trees. However, we can also view ambiguities as arising from the competition (MacDonald, Pearlmutter, & Seidenberg, 1994; MacWhinney, 1987b) between alternative perspectives. Moreover, if we look closely at the processing of these ambiguities, there is evidence for perspective tracking effects that go beyond simple structural competition. Consider the examples in sentences 31 to 34.

31. Visiting relatives can be a nuisance.
32. Crying babies can be a nuisance.
33. Teasing babies is unfair.
34. If they arrive in the middle of a workday, visiting relatives can be a nuisance.

Looking at each of these ambiguities in terms of the principle of incremental interpretation, we can see how alternative perspectives organize alternative interpretations. In each case, there is a competition between the overtly expressed noun following the participle and an unexpressed subject. In each case, the participle is looking to fill

the subject/perspective role. In example 31, it is plausible that relatives could be “visiting.” With “relatives” filling the role of the perspective, the interpretation is that “if relatives visit you, they can be a nuisance.” At the same time, we are also able to imagine that some unspecified person serves as the omitted perspective of “visit.” In this case, the “relatives” fill the role of the object, yielding the interpretation that “it can be a nuisance to pay a visit to one’s relatives.”

In example 32, on the other hand, the verb is intransitive. If we were to associate the perspective with an unexpressed subject, then we would have no role for “babies.” So, here, only one interpretation is possible, and it involves the “babies” as the initial perspective. That initial perspective is eventually shifted at the word *nuisance*, since we have to take the perspective of the person being annoyed to understand how the babies become a nuisance. In 33, the “babies” are unlikely to be doing the teasing and they serve as good objects, so the unexpressed subject wins the role of perspective. In 35, the foregrounding of “they” in the first clause prepares the way for perspective continuation to the second clause which promotes “relatives” as the subject of visiting. Although, we can still find an ambiguity in 34, we are less likely to notice it than in 31. As we trace through these various competitions, we see that the demands for incremental construction of a perspectival mental model work to shape the extent to which we can maintain alternative ambiguous perspectives.

Perspectival ambiguity can also arise from the competition between alternative phrasal attachments. In example 35, the initial perspective resides with “Brendan.” Although the verb *fly* would prefer to have a preverbal noun serve as its perspective, the implausibility of seeing a canyon fly through the air tends to force us away from this syntactically preferred option.

- 35. Brendan saw the Grand Canyon flying to New York.
- 36. Brendan saw the dogs running to the beach.
- 37. The women discussed the dogs on the beach.
- 38. The women discussed the dogs chasing the cats.

However, the shift to the perspective of “the dogs” is easier in 36, although again we can maintain the perspective of “Brendan” if we wish. In cases of prepositional phrase attachment competitions, such as 37, 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 38, on the other hand, we have a harder time imagining that the women, instead of the dogs, are chasing the cats. Sentences such as 37 and 38 have motivated a variety of formal accounts of sentence processing within the framework of the Garden Path account (Frazier, 1987). For these sentences, the perspective hypothesis provides an account that focuses on conceptual competitions, rather than recovery from nonconceptual parsing decisions.

It is possible to shift perspective abruptly between clauses by treating the verb of the first clause as intransitive and the following noun as a new subject. A shift of this type is most likely to occur with a verb like *jog* that is biased toward an intransitive reading, although it can also function as a transitive. Examples 39 to 41 illustrate this effect:

39. Although John frequently jogs, a mile is a long distance for him.
40. Although John frequently jogs a mile, the marathon is too much for him.
41. Although John frequently smokes, a mile is a short distance for him.

Detailed self-paced reading and eye-movement studies of sentences like 39, 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 of this type show that perspectival shifting is an integral part of online, incremental sentence processing (Marslen-Wilson & Tyler, 1980).

This description of the processing of these ambiguities has relied on the six assumptions stated earlier. In particular, these perspective shifts are triggered by alternative activations of role fillers and phrasal attachments, as specified in greater detail in descriptive accounts such as MacWhinney (1987a) or O’Grady (2005) or in computationally explicit models such as Kempen and Hoenkamp (1987) or Hausser (1999). These competitions play out as we add incrementally to the ongoing perspectival mental models we are creating. Syntactic

processes play out their role through the competitive operation of role filling and attachment, but the actual shift of perspective occurs within the mental model that is being constructed in a simulated perspectival space.

Scope

Ambiguities in quantifier scope provide a very different way of understanding the operation of perspective tracking. To understand what is at issue here, consider this example:

42. Every boy climbed a tree.

For sentences that begin with quantified nominals like “every boy,” the construction of an initial perspective is more complex than in the case of sentences that begin with a simple nominal like “my dog” or “Bill.” For simple nominals, we only have to create a single unified imagined agent in perspectival space. For quantified nominals, we have to create a perspective that allows for multiple agents. Moreover, right from the beginning, we have to take into account the nature of the quantifier. We can think of quantified perspectives in terms of Venn diagrams (Johnson-Laird, 1983). For the phrase “every boy,” we set up a Venn diagram that includes several nodes characterized as “boys.” We do not need to actually count these nodes in our imagination. Instead, we use an automated procedure for perspective activation that sets up enough imagined nodes to satisfy us that there are several boys. We then link “every” to “climb” by duplicating the acts of climbing across the multiple perspectives. We do not have to actually make each boy engage in climbing in our mind. Instead, we can rely on an automated procedure that makes one boy climb and then assumes that the others will “do the same.”

The ambiguity in sentence 42 arises when we come to “a tree.” At this point, we have the option of imagining climbing either a single tree or multiple trees (O’Grady, 2006). We can think of this ambiguity as involving a single unified multiple perspective with many boys and one tree versus a divided multiple perspective with many boys and many trees. If we imagine that the initial perspective constitutes a unified group, we are relatively less likely to imagine multiple trees. In fact, the nature of the verb determines the extent to which we keep

a unified or divided multiple perspective. Consider these examples from O'Grady (2006):

43. Everyone gathered at a restaurant.
44. Everyone surrounded a dog.

Here, the activity of gathering or surrounding requires the individuals in the initial multiple perspective to act in concert. When they act this way, the components of the multiple perspective are more likely to focus on a single object, rather than multiple objects. Thus, the shift to a unified single representation for the object is determined by the embodied representation of the subject's perspective as it combines with the activity of the verb.

Scope ambiguities also display interesting interactions with grammatical constructions that shift the order of sentence elements. In example 45, we can imagine either that the students are all reading the same books or that each student is reading a different set of three books. This contrast is much like the contrast in 42 between a multiple perspective that remains unified when processing the object and a multiple perspective that divides when processing the object.

45. Two students read three books.
46. Three books are read by two students.

In 46, on the other hand, it is difficult to imagine more than one set of three books. As a result, 46 does not permit the ambiguity that we find in 45. This effect illustrates the operation of assumption 3 regarding the conceptual centrality of the starting point. According to assumption 3, starting points constitute the foundation stone for the construction of the rest of the edifice of the sentence. Because the remaining edifice rests on this foundation stone, we need to make as full a commitment as we possibly can to the referential clarity of the starting point. We can express this particular corollary of assumption 3 in these terms:

Principle of Referential Commitment: Nominals that are being used as starting points should be linked to unambiguous referents in mental models. If full definite coreference cannot be achieved, then nominal starting points are assumed to be uniquely identifiable new referents.

In the case of 46, this means that, once “three books” are postulated in mental space they cannot be multiplied into two sets of three books as in 45.

This same principle is involved in the construction of interpretations for sentences like 47 to 50:

- 47. The devoted environmentalist tracked every mountain goat.
- 48. A devoted environmentalist tracked every mountain goat.
- 49. The boy ate every peanut.
- 50. A boy ate every peanut.

The contrast in these sentences is between starting points that are fully referential, as in 47 and 49, and those that are indefinite, as in 48 and 50. When the starting point is fully referential, it cannot later be divided by backwards multiplication of perspectives. As a result, 47 and 49 are not ambiguous. However, in 48 we can imagine each mountain goat being tracked by a different environmentalist. To do this, we return to the perspective of the starting point and multiply that perspective. This multiplication of perspectives is possible for 48, because indefinite perspectives are not as fully committed referentially as definite perspectives. In 50, we can imagine a similar ambiguity, although the idea of many boys each eating only one peanut is perhaps a bit silly.

Perspective tracking theory also explains why 51 and 52 are acceptable, whereas 53 is questionable. In 51 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 being fed. This means that we can continue in (52) by asking whether or not all of these donkeys will grow. Sentence (53), 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.

- 51. Every farmer who owns a donkey feeds it.
- 52. Every farmer who owns a donkey feeds it, but will they grow?
- 53. 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:

- 54. SS: The dog that chased the cat kicked the horse. 0 switches
- 55. OS: The dog chased the cat that kicked the horse. 1- switch
- 56. OO: The dog chased the cat the horse kicked. 1+ switch
- 57. 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.

Sentences with multiple center embeddings have even more switches. Consider an example like (58) which has four difficult perspective switches (dog -> cat -> boy -> cat -> dog).

- 58. The dog the cat the boy liked chased snarled.
- 59. My mother’s brother’s wife’s sister’s doctor’s friend had a heart attack.

Sentences that have as much perspective shifting as 58, 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 example 59. In that example, we do not really succeed in taking each perspec-

tive and switching to the next. Instead, 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 and fail to track the relation of that friend to the mother's brother. In the terms of a processing model, we can say that we continue to push words onto a stack and end up losing track, in terms of our mental model, of the items that were pushed down to the bottom of the stack. In the case of 59, we still end up with an interpretable sentence. In the case of 60, the result makes little sense.

Researchers have often tried to account for these processing difficulties in structural terms. However, fMRI work (Booth et al., 2001; Just, Carpenter, Keller, Eddy, & Thulborn, 1996) has contrasted the processing of object relative sentences like 57 with the processing of subject relatives like 54. These studies have shown that 57 produces greater activation in a wide variety of left-hemisphere areas. Models such as that proposed by Grodzinsky and Amunts (2006) have attempted to link structural complexity to processing in Broca's area. But the fMRI results show that complexity leads to activation across a far wider area. This wider profile of activation is consistent with the idea that the complexity involved is not structural, but rather involves the fuller construction of switched perspectives in a full mental model.

Perhaps no sentence has figured more heavily in discussions of sentence processing than example 60.

60. The horse raced past the barn fell.

The competition model account of processing in reduced relative sentences of this type emphasizes the dual morphological function of verbal suffix *-ed*. This suffix can mark either the past tense or the past participle. When *-ed* marks the past tense, the verb *raced* is a simple intransitive with "horse" as its subject or perspective. However, when *-ed* marks the participle, then the verb *raced* allows for a nonexpressed subject and takes the "horse" as the object, much as in the shift of sentence 32 above with the "visiting relatives." Because the resting activation of the past tense interpretation of the suffix is higher than that of the participle, listeners may not pick up the participle interpretation until they realize that the sentence will not parse with the past tense interpretation. In this case, we sense a garden path because we only active the weak perspective configuration

when the strong configuration fails. Other similar configurations will behave very differently. For example, in 61, we sense no garden path because there is no noun following “kept” that would allow the transitive reading and we therefore have to rely on the reduced relative reading. In 62, there is no ambiguity, because the irregular participle cannot be confused with the past tense.

- 61. The bird kept in the cage sang.
- 62. The car driven past the barn honked.

Clitic Assimilation

As a further, detailed example of the impact of perspective taking on grammar, let us consider the process of clitic assimilation. In English, the infinitive “to” often fuses with a preceding modal verb to produce contractions such as “wanna” from “want to” in cases such as 64. However, this assimilation is blocked in environments like the one in 65, making 66 unacceptable.

- 63. Who do you want to see?
- 64. Who do you wanna see?
- 65. Who do you want to go?
- 66. *Who do you wanna go?

The perspective hypothesis views the reduced infinitive in 64 as a cue that marks perspective maintenance. In 65, this reduction is impossible, because there is a forced processing shift from “who” to “you” and then back to “who(m).” Infinitive reduction also marks perspective maintenance in examples 67 to 69.

- 67. I get ta go. (Privilege)
- 68. I got ta go. (Privilege, past tense)
- 69. I gotta go. (Obligation)

In 67 and 68, the privilege of going is due presumably to the intercession of an outside party. The perspective of this outside party interrupts perspective maintenance. In 69, on the other hand, the obligation is internal to the speaker and perspective is maintained across the reduced infinitive.

Implicit Causality

Much of our analysis here has focused on perspective marking by highly grammaticalized forms like pronouns, participles, gerundives, relativizers, and infinitives. However, perspective marking extends far beyond grammatical forms, appearing widely inside adjectives, verbs, nouns, and prepositions. Individual lexical items can characterize complex social roles and mental acts. Items like “libel,” “Internet,” or “solidarity,” encode social scenarios organized about the perspective of social actors. Let us take the noun *libel* as an example. When we speak of some communication as being “libelous,” we are 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 his or her ability to function in public life. 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.

Verbs like *promise*, *forgive*, *admire*, and *persuade* also 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 “implicit causality.” Sentence 70 illustrates the implicit causality configuration of the experiencer-stimulus verb *admire*. The causal configuration is revealed in the second clause where the subject (“she”) is the cause of the admiration. In sentence 71, with the stimulus-experiencer verb *apologize*, causality remains with the subject of the first clause (“John”).

70. John admired Mary_i, because she_i was calm under stress.

71. John_i apologized to Mary, because he_i had cracked under stress.

According to the perspective hypothesis, shifts in causality should lead to shifts in perspective. To track these shifts experimentally,

McDonald and MacWhinney (1995) asked subjects to listen to sentences like 70 and 71, while making a cross-modal 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 71 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 69 tend to force a shift in perspective away from the starting point (“John”) to the stimulus (“Mary”) right at the 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 70, 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 hypothesis.

In addition to encoding implicit causality, verbs can also encode information regarding implicit source of knowledge.

72. Minnie told Dorothy that she knew Superman.
73. Minnie asked Dorothy if she knew Superman.
74. Minnie reminded Dorothy that she knew Superman.
75. Minnie told Dorothy that she made Superman cry.

In 72, we assume that Minnie has access to knowledge about herself which she provides to Dorothy. In 73, on the other hand, we assume that Dorothy must be the source of the information, since Minnie would certainly have access to her own knowledge. Although both 74 and 75 could be read ambiguously, the most probable reading in each case is one that maintains the perspective of the starting point. Adults are able to maintain the viewpoint of the initial subject even in the complement clause. At the same time, they use facts about the verb *ask* to shift perspective in 73. However, between 5 and 8, children (Franks & Connell, 1996) are more likely to shift to the perspective of “Dorothy” in all of these sentences. This tendency to shift perspective arises from a general preference for local attachment evident at this age. It is possible that children are not able to coordinate the dual perspectives of the main and subordinate clause efficiently

for these structures during this age range (Huttenlocher & Presson, 1973).

Perspectival Overlays

In the preceding sections, we have focused on the ways in which grammatical devices provide cues that help listeners track shifts of perspective. These shifts have involved the actions and motions of agents, as they operate on other objects. This system of causal action marking is at the core of perspective tracking and it is the system that is marked most overtly through grammatical constructions and forms. However, there are at least five other systems of perspective shifting that function as linguistic overlays on this basic system of perspective tracking. These include the systems for marking perspective in space, time, empathy, evidentiality, and metaphor. Although these systems are of great importance for mental model construction, they have relatively little impact on grammatical structure, relying instead on marking through individual lexical items.

Space

Spatial localization is fundamental to perspective tracking. The linguistic marking of space relies on prepositions, motion verbs, and names for landmarks. Sometimes the marking of location can involve perspectival ambiguity. Consider this classic illustration from Cantrall (1974):

76. The adults in the picture are facing away from us, with the children behind them.

In this example, we see a competition between alternative reference points. If we take the perspective of the adults as our reference point, then the children are located between the adults and the viewer of the picture. If we take the perspective of the viewers of the picture as the reference point, the children are located on the other side of the adults, farther away from us. Ambiguities of this type are reminiscent of the shifts in reference point in sentences like 5 where we can imagine either imagine ourselves flying to New York or else

see the Grand Canyon flying to New York. However, on a structural level, 76 is not ambiguous, whereas 5 is. In other words, the ambiguities we find in spatial perspective taking are not reflected in the grammar. However, they are clearly reflected in our mental models. This is why we can consider perspective taking in space as an overlay on grammar.

Perspectival competitions can arise even from what seems to be a single reference point. 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.

Because spatial reference is so prone to ambiguity of this type, we have developed many linguistic devices for reducing such ambiguities. One way of reducing ambiguity is to use a third position as the reference point. For example, if we describe a position as being “50 yards behind the school,” it is not clear whether we are taking our own position as the reference point for “behind” or whether we are using the facing of the school as the reference point. To avoid this problem we can describe the position as “50 yards toward the mountain from the school.” In this case, 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. Languages such as Guugu Yimithirr (Haviland, 1993) and Mayan take this solution yet one step farther by setting up permanent map-like coordinates against which all locations can be pinpointed.

Time

Perspective taking in time is closely analogous to perspective taking in space. Like space, time has an extent through which we track events in terms of their relation to reference moments. Just as spatial objects have positions and extents, events have locations in time and durations. 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

77 with an iconic temporal order than ones like 78 with a reversed order. However, sentences like 79 which require no foreshadowing of an upcoming event are the most natural of all.

- 77. After we ate our dinner, we went to the movie.
- 78. Before we went to the movie, we ate our dinner.
- 79. 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.

Empathy

The third system of perspectival overlays involves marking for empathy. Here, Tenny and Speas (2002) have conducted a useful survey of devices used by various languages. Among the most marked of these devices are evaluative adjectives such as “beloved” or “damned.” Consider these examples:

- 80. John was looking for Sarah’s beloved cat.
- 81. John was looking for Sarah’s damned cat.

In 80, the cat is beloved from Sarah’s point of view. In 81, however, the cat is “damned” from either John’s point of view or the speaker’s point of view. Language is laden with evaluative perspectives of this type. However, much of the evaluation we convey in everyday interaction is encoded equally well through intonation and gesture—areas that lie outside the scope of our current exploration.

Evidentiality

A fourth area of perspectival overlay involves the marking of the evidence sources and types. For example, we may know some things because we saw them directly with our own eyes, whereas we know

others things because we have heard them from trusted sources. Often, we simply make assertions without providing information regarding evidence. In other cases, we may ask questions, indicate doubt, express belief, and so on. An example of the role of evidential perspective is the contrast between statements and questions. Consider the contrast between 82 and 83:

- 82. The bicyclist appears to have escaped injury.
- 83. Did the bicyclist appear to have escaped injury?
- 84. The reporter said that the bicyclist appeared to have escaped injury.
- 85. The reporter asked if the bicyclist appeared to have escaped injury.

In 82 the evidence is evaluated on the basis of evidence available to the speaker. In 83, on the other hand, the evidence is evaluated on the basis of evidence available to the listener. Examples 84 and 85 display a similar asymmetry.

In English, the marking of finer dimensions of evidentiality is conveyed by particles and adverbs such as *well*, *sure*, *still*, and *just*. In other languages, these same forms can appear as markings on the verb. Some languages pay close attention to fine distinctions in the source and nature of these evidences. Japanese displays a particularly interesting restriction on evidence reflected in 86 and 87:

- 86. You are sick.
- 87. You seem sick.

In Japanese, one cannot say 86, because it is presumptuous to imagine that one has access to inner states of another person. In fact, it might even be a bit inappropriate to produce 87. This constraint, known as “speaker’s territory of knowledge” (Kamio, 1995), involves aspects of both evidentiality and empathy.

The tracking of evidential perspectives occurs primarily on the level of the fuller discourse or narrative. On this level, we use particular constructions to help our listeners locate objects in their own mental models. In effect, we construct mental models of the mental models of our listeners and use these to determine the marking of evidentiality. As Givón (2005) puts it, speakers select grammatical constructions on the basis of their “mental models of the interlocutor’s current deontic and epistemic states.”

Metaphor

A fifth area of perspectival overlay on language involves the use of metaphor. The production and comprehension of metaphors, similes, and analogies has been a lively topic now for nearly three decades in both the linguistic and psychological literature. Examples 88 and 89 illustrate some of the typical patterns studied in this vast descriptive and experimental literature.

- 88. The road runs down to the river.
- 89. Headline: Congress stumbles in debate on tax reform.
- 90. Stocks took a plunge at the end of the day.
- 91. His marriage was like a glacier.

Everyday conversational speech makes very little use of metaphor, but other genres, ranging from financial reporting to pop psychology, rely heavily on extensions of the type illustrated in 89 and 90 to liven up otherwise boring prose. Lakoff and Johnson (1980) have shown how a small set of core metaphors, based on embodied cognition, dominates our construction of mental models. Fauconnier and Turner (1996) have further described the blending of multiple perspectives that occurs in mental models. Perhaps the most remarkable of these blends arise in drama and poetry when we track the perspectives of actors in plays within plays within stories, as in *Macbeth* and “The Story of the Tailor” in *The Arabian Nights*.

The processing of perspective in metaphors and blends involves conceptual overlays on language, much like the processing of perspective in space, time, empathy, or evidentiality. Like these other systems, the processing of metaphor makes no direct contact with grammar, functioning instead through semantic extension of the meanings of individual words such as *stumble* or *glacier*. However, it would be a mistake to ignore the ramifications of these types of processing for our theories of neuronal encoding of perspectives in mental models. The perspectival complexity of *Macbeth* or *The Arabian Nights* marks only the beginning of complex mental models that must be encoded in verbal form. For truly dazzling levels of complexity we can turn to Grigori Perelman’s solution of the Poincaré conjecture or the chains of reasoning in cases argued at the Supreme Court of the United States.

Production and Integration

Having examined the marking of perspective through grammar, we are now ready to consider how language integrates across these various levels or layers of perspective taking. The best way to explore this issue is to consider snippets of actual narratives and conversations. Consider this passage from an Associated Press release in 2001.

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

This passage begins from the perspective of the cyclone. It then uses a metaphorical image to allow the storm to act as an agent that wields a hammer. The past tense suffix on the verb *hammered* places this action into the past. The object of the hammering is the Bangladeshi coast. The coast is brought on stage, but there is no shift of perspective away from the cyclone and its hammering. Immediately after “coast” we have the introduction through “Monday” of a temporal overlay. Then, the metaphor of “hammering” is linked to the force of hundreds of demons. The perspective of the cyclone continues with “leveling,” “flooding,” and “killing.” In each case, our attention moves briefly to the objects without really shifting away from the cyclone.

In this sentence, there is a rich combination of images from many modalities. The cyclone is a visual image, the hammer is a visual image linked to motor movements of the arm and perceptions of noise and percussion. The image of the Bangladesh coast brings to mind the position of Bangladesh on a map of the Bay of Bengal, along with a trace of the delta formed by the Brahmaputra. The image of Monday forces us to refer to our recent calendrical memory to locate this event in a spatial analog to time. The images of demons bring to mind scenes from Indian art with black demon faces flying through the air and stories from the Vedas. We do not count out the demons in detail, but roughly imagine an array of many demons, perhaps extending over a wide physical space along the coast to accommodate their great number. When we link “entire” to “villages,” we have to engage in a mental activity of leveling that is completive and leaves no huts standing. Similarly, when we flood the crops, we have to imagine a whole scene with plants under water and when we envision

the killing, we actually envision six dead bodies, although we are not exactly sure how they died. Together, the construction of the image for just this sentence relies on diverse cognitive systems for motor action, space, time, enumeration, quantifier scope, visual imagery, semantic memory, metaphor, geography, geometry, and biology.

Some might like to think of these systems as cognitive modules (Pinker, 1997). It is certainly true that these diverse cognitions are supported by widely separated brain regions with highly differentiated functions, but it is misleading to think of language as popping together beads produced by encapsulated modules, as suggested by Carruthers (2002). Instead, language allows diverse areas to work in concert and to achieve communication by writing to the “blackboard” of the sentence currently under construction. In this sense, language does indeed promote integration across these nonmodular cognitive systems in a way that can promote cognitive growth (Spelke, 2002). In fact, it makes sense to think of language as providing a springboard for recent human cultural evolution, by allowing us to construct integrated mental models (MacWhinney, 2005; Mithen, 1996) that led to the further elaboration of consciousness (Dennett, 1991), social structure, and religious imagination (Campbell, 1949).

These reflections regarding information integration through perspective taking have been grounded so far on the single example sentence 92. This same passage continues with the material in 93.

93. Three men and two children were crushed under collapsed buildings or hit by flying pieces of tin roofs in the southern port of Chittagong. One man died in Teknaf, about 110 miles down the coast, when he was blown off his roof, while trying to secure it. The storm roared in from the Bay of Bengal with wind gusts of 125 mph, forcing a half-million people to flee their huts and huddle in concrete shelters. Many power and telephone lines were down, so a full account of casualties and damage was not available.

In this continuation, we shift from the initial perspective of the cyclone in 92 to the perspective of people being crushed. This involves a passive perspective, as marked by the -ed suffix. The actual causal agents follow later. The first agent is totally missing, since the fact that it was the cyclone that caused the buildings to collapse is not expressed. For the set passive verb “hit” the agent is “flying pieces of tin roofs.” Here, we begin with the notion of “flying” even before we

know what might be flying. We soon realize that what is flying are pieces of tin roofs, and we try to imagine how the cyclone pulled these pieces off of their roofs. As we read through a passage quickly, we may decide not to perform the extra mental work needed to fill out this further detail of the mental model. Even before we shift to the perspective of the flying pieces, we must see the victims crushed under buildings or being hit. We do not know which victims were crushed and which were hit by flying pieces of roof, so we just imagine some of each in both positions without trying to do any actual count.

Next we shift from Chittagong to Teknaf. The shift in space is accompanied by the introduction of the new perspective of “one man.” For this man, we first imagine him dead, then we see how he is passively blown off his roof by the cyclone as the unmentioned agent, and then we must put him back on his roof and imagine him trying to secure the roof. This order of events is the opposite of the actual order and building up a mental model in opposite order can be difficult. An alternative version of this sentence would read as in 94.

94. In Teknaf, 110 miles down the coast, a man was trying to secure his roof when the cyclone blew him off to his death.

Next we shift perspective back to the storm as in now acts not just on six people, but half a million. Finally, we move to viewpoint of the reporter, who explains that it was not possible to give a full report of the casualties because telephone lines were down.

This analysis of a news story selected at random illustrates the extent to which language weaves together diverse cognitions into the common grid of the sentence. Sentences achieve this integration by mixing together adjectives, lexical metaphors, quantifiers, descriptive verbs, numerals, temporals, prepositional phrases, passives, omitted subjects, participles, adverbs, conjunctions, and a wide variety of other grammatical devices. Some of these devices mark referents, some mark perspective shifts, and others add spatial, temporal, and evaluative overlays to the basic causal grid.

If we move beyond examples like 91 and 92 to examination of spontaneous conversation, the landscape changes markedly. Instead of weaving together places, times, and actors, conversations weave together diverse viewpoints, understandings, and goals. Conversations are heavily dedicated to the maintenance of interpersonal

relations and the establishment of mutual knowledge. Against this background, perspective tracking still plays an important role, but the perspective being tracked is one that is under continual negotiation between the conversational participants. A fuller examination of these issues is currently underway in the context of analyses of conversational interactions in classrooms (Greeno & MacWhinney, 2006).

Cultural Transmission

Vygotsky (1934) believed that children's conversational interactions played a fundamental role in their cognitive development. He viewed these interactions as setting a model for mental structures that children would then internalize as "inner speech." He characterized inner speech in terms of processes such as topic-comment structure and ellipsis, but provided no additional linguistic or cognitive detail. The perspective hypothesis can be viewed as an elaboration of the initial Vygotskian program. The idea here is that, by tracking perspective in conversations and narratives, children construct mental models that encode perspectival patterns in long term memory. Because these models are extracted from adult input, the perspective tracking and causal reasoning they contain will reflect the standards of the adult community. For example, if fairy tales begin with phrases such as "once upon a time, far, far away," then children will also learn to construct mental models for fairy tales in a cognitive space that is far away or imaginary in space and time. If the efforts of the valiant prince lead eventually to happiness and marriage, then children will build mental models in which these expectations are linked. If the efforts of a determined little locomotive allow it to pull a train up a hill, children will also imagine that they can achieve goals through determination. In effect, children will use perspectively constructed models extracted from conversation and narrative as a method for learning the systems and values of their culture. In this process, the links developed through perspective taking on all the levels we have discussed are crucial. Without language and the perspective tracking it allows, this level of cultural transmission would not be possible.

Developmentalists have extended the Vygotskian vision by linking the construction of mental models to play (Eckler & Weininger,

1989), games (Ratner & Bruner, 1978), narration (Bruner, 1992; Nelson, 1998), apprenticeship (Lave, 1991), learning contexts (Rogoff, 2003), and conversational sequencing (Bateson, 1975). What is common in all of these accounts is the idea that children are exposed to interactions in which they track the logical flow of ideas perspectively. From this process, they extract internalized mental models that are specific to their cultures and social groups (Spradley, 1972) that they can then transmit to others (Blackmore, 2000).

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. In this new formulation, the links in mental models are viewed as inherently perspectival and grounded on simulated, embodied cognition. This cognitive system relies on a wide range of neuronal structures for body image matching, spatial projection, empathy, and perspective tracking. Language uses this underlying system to achieve still further cognitive integration. When speakers produce sentences, they use grammatical devices to integrate diverse perspectives and shifts. When listeners process these sentences, they use grammatical markers, constructions, and lexical forms to decode these various shifted and overlaid perspectives. 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, causality, evidentiality, evaluation, empathy, and mental acts. Across each of these dimensions, we assume and shift between perspectives in order to construct a fully human, unified conscious awareness.

Acknowledgments

Thanks to William O'Grady, James Greeno, and Marnie Arkenberg for their comments on this paper. This work was supported by NSF Award SBE-0354420.

References

- Adolphs, R. (2003). Cognitive neuroscience of human social behavior. *Nature Reviews of Neuroscience*, 4, 165–178.
- Adolphs, R., & Spezio, M. (2006). Role of the amygdala in processing visual social stimuli. *Progress in Brain Research*, 156, 363–378.
- Avenanti, A., Buetti, D., Galati, G., & Aglioti, S. (2005). Transcranial magnetic stimulation highlights the sensorimotor side of empathy for pain. *Nature Neuroscience*, 8, 955–960.
- Bailey, D., Chang, N., Feldman, J., & Narayanan, S. (1998). Extending embodied lexical development. *Proceedings of the 20th Annual Meeting of the Cognitive Science Society*, 20, 64–69.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20, 723–767.
- Bateson, M. (1975). Mother–infant exchanges: The epigenesis of conversational interaction. In D. Aaronson & R. Rieber (Eds.), *Developmental psycholinguistics and communication disorders* (pp. 112–140). New York: New York Academy of Sciences.
- Blackmore, S. (2000). The power of memes. *Scientific American*, October, 64–73.
- Booth, J. R., MacWhinney, B., Thulborn, K. R., Sacco, K., Voyvodic, J. T., & Feldman, H. M. (2001). Developmental and lesion effects during brain activation for sentence comprehension and mental rotation. *Developmental Neuropsychology*, 18, 139–169.
- Bruner, J. (1992). *Acts of meaning*. Cambridge, MA: Harvard University Press.
- Budiu, R., & Anderson, J. (2004). Interpretation-based processing: A unified theory of semantic sentence comprehension. *Cognitive Science*, 28, 1–44.
- Campbell, J. (1949). *The hero with a thousand faces*. Princeton, NJ: Princeton University Press.
- Cantrall, W. (1974). *View point, reflexives and the nature of noun phrases*. The Hague: Mouton.
- Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, 33, 657–674.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1975). *Reflections on language*. New York: Random House.
- Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. Cambridge, MA: MIT 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.

- Dennett, D. (1991). *Consciousness explained*. New York: Penguin Press.
- Donald, M. (1991). *Origins of the modern mind*. Cambridge, MA: Harvard University Press.
- Eckler, J., & Weininger, O. (1989). Structural parallels between pretend play and narratives. *Developmental Psychology*, 25, 736–743.
- Fauconnier, G., & Turner, M. (1996). Blending as a central process of grammar. In A. Goldberg (Ed.), *Conceptual structure, discourse, and language* (pp. 113–130). Stanford, CA: CSLI.
- Franks, S. L., & 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* (Vol. 12, pp. 601–681). London: Erlbaum.
- Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.
- Givón, T. (2002). The visual information-processing system as an evolutionary precursor of human language. In T. Givón & B. Malle (Eds.), *The evolution of language out of pre-language* (pp. 3–51). Amsterdam: Benjamins.
- Givón, T. (2005). *Context as other minds: The pragmatics of sociality, cognition, and communication*. Philadelphia: Benjamins.
- Greeno, J., & MacWhinney, B. (2006). *Perspective shifting in classroom interactions*. Paper presented at the AERA Meeting.
- Grodzinsky, Y., & Amunts, K. (2006). *Broca's region*. Oxford: Oxford University Press.
- Grodzinsky, Y., & Reinhart, T. (1993). The innateness of binding and coreference. *Linguistic Inquiry*, 24, 187–222.
- Harris, C. L., & Bates, E. A. (2002). Clausal backgrounding and pronominal coreference: A functionalist alternative to c-command. *Language and Cognitive Processes*, 17, 237–269.
- Hausser, R. (1999). *Foundations of computational linguistics: Man-machine communication in natural language*. Berlin: Springer.
- Haviland, J. B. (1993). Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology*, 3, 3–45.
- Huttenlocher, J., & Presson, C. (1973). Mental rotation and the perspective problem. *Cognitive Psychology*, 4, 277–299.
- Jeannerod, M. (1997). *The cognitive neuroscience of action*. Cambridge, MA: Blackwell.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press.

- Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, 274, 114–116.
- Kakei, S., Hoffman, D. S., & Strick, P. L. (1999). Muscle and movement representations in the primary motor cortex. *Science*, 285, 2136–2139.
- Kamio, A. (1995). Territory of information in English and Japanese and psychological utterances. *Journal of Pragmatics*, 24, 235–264.
- Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11, 201–258.
- Kintsch, W., & Van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: Chicago University Press.
- Lave, J. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical nature of syntactic ambiguity resolution. *Psychological Review*, 101(4), 676–703.
- Macrae, C., Heatherton, T., & Kelley, W. (2004). A self less ordinary: The medial prefrontal cortex and your. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (Vol. 3, pp. 1067–1076). Cambridge: MIT Press.
- MacWhinney, B. (1977). Starting points. *Language*, 53, 152–168.
- MacWhinney, B. (1987a). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Hillsdale, NJ: Erlbaum.
- MacWhinney, B. (1987b). Toward a psycholinguistically plausible parser. In S. Thomason (Ed.), *Proceedings of the Eastern States Conference on Linguistics* (pp. . Columbus, Ohio: Ohio State University.
- MacWhinney, B. (2005). Language evolution and human development. In B. Ellis & D. Bjorklund (Eds.), *Origins of the social mind* (pp. 383–410). New York: Guilford.
- Marslen-Wilson, W. D., & Tyler, L. K. T. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1–71.
- Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature*, 379, 649–652.
- McClelland, J. L. (1979). On the time-relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, 86, 287–330.
- 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.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

- Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31, 838–850.
- Meltzoff, A. N., & Decety, J. (2003). What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London B*, 358, 491–500.
- Middleton, F. A., & Strick, P. L. (1998). Cerebellar output: Motor and cognitive channels. *Trends in Cognitive Sciences*, 2, 348–354.
- Miller, G., & Johnson-Laird, P. (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- Mitchell, D. C. (1994). Sentence parsing. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 375–405). San Diego, CA: Academic Press.
- Mithen, S. (1996). *The prehistory of the mind: The cognitive origins of art, religion, and science*. London: Thames & Hudson.
- Nelson, K. (1998). *Language in cognitive development: The emergence of the mediated mind*. New York: Cambridge University Press.
- O’Grady, W. (2005). *Syntactic carpentry*. Mahwah, NJ: Erlbaum.
- O’Grady, W. (2006). The syntax of quantification in SLA: An emergentist approach. In M. O’Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 98–113). Somerville, MA: Cascadilla Press.
- 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.
- Pecher, D., & Zwaan, R. (Eds.). (2005). *Grounding cognition*. Cambridge: Cambridge University Press.
- Pelphrey, K. A., Mitchell, T. V., McKeown, M. J., Goldstein, J., Allison, T., & McCarthy, G. (2003). Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. *Journal of Neuroscience*, 23, 6819–6825.
- Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2005). Neural basis of eye gaze processing deficits in autism. *Brain*, 128, 1038–1048.
- Pelphrey, K. A., Viola, R. J., & McCarthy, G. (2004). When strangers pass. *Psychological Science*, 15, 598–603.
- Pinker, S. (1997). *How the mind works*. New York: Norton.
- Ramachandran, V. S. (2000). Phantom limbs and neural plasticity. *Neurological Review*, 57, 317–320.
- Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia: A window into perception, thought and language. *Journal of Consciousness Studies*, 8, 3–34.

- Ratner, N., & Bruner, J. (1978). Games, social exchange and the acquisition of language. *Journal of Child Language*, 5, 391–401.
- 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.
- Rogoff, B. (2003). *The cultural nature of human development*. Oxford: Oxford University Press.
- Spelke, E. (2002). Developing knowledge of space: Core systems and new combinations. In S. Kosslyn & A. Galaburda (Eds.), *Languages of the brain* (pp. 239–258). Cambridge, MA: Harvard University Press.
- Spradley, J. (Ed.). (1972). *Culture and cognition: Rules, maps, and plans*. New York: Chandler.
- Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, 12, 153–156.
- Stein, D., & Wright, S. (Eds.). (1995). *Subjectivity and subjectivisation*. Cambridge: Cambridge University Press.
- Tenny, C., & Speas, P. (2002). Configurational properties of point of view roles. In A. DiSciullo (Ed.), *Asymmetry in grammar*. Amsterdam: Benjamins.
- Vygotsky, L. (1934). *Thought and language*. Cambridge: MIT Press.
- Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1196–1207.

