Introduction:
Viewpoint and perspective in language and gesture, from the Ground down.

Eve Sweetser
UC-Berkeley
sweetser@berkeley.edu

May 6, 2010

The anecdote of the jar

I placed a jar in Tennessee,
And round it was, upon a hill.
It made the slovenly wilderness
Surround that hill.

The wilderness rose up to it,
And sprawled around, no longer wild.
The jar was round upon the ground
And tall and of a port in air.

It took dominion every where.
The jar was gray and bare.
It did not give of bird or bush,
Like nothing else in Tennessee.   *- Wallace Stevens*

A landscape with a human in it is no longer just a landscape; it is also a human's egocentric conceptualization of a landscape. In Wallace Stevens' poem, even a human artifact like a jar proliferates new human-centered construal of the landscape. Viewpoint permeates human cognition and communication - predictably, since we never have experience of the world except as a viewpoint-equipped embodied self among other viewpointed embodied selves. Language reflects this fact of embodiment: linguistic structure shows no way entirely out of viewpoint to an objective pre-experiential description of the world. But it also shows in complicated and fascinating ways the possibility of a single mind accessing multiple different viewpoint affordances on the same scene. Without such cognitive flexibility, humans could not cooperate and communicate at the high level which is apparently unique to our species, and universal to neurally and developmentally typical members of the species (Tomasello 1999, 2008). For this reason, viewpoint is a phenomenon of special interest to almost anyone studying cognition or communication: linguists, cognitive scientists, literary analysts, philosophers and many more.

The title phrase in this introduction, *from the Ground down*, articulates something which the rest of this book confirms: cognitive perspective starts with bodily viewpoint within a real physical Ground of experience. In a mental space network, embedded spaces are thought of as being metaphorically "below" their "mother" spaces in the spatial network. By this metaphor, viewpoint flows downwards like water from the world as we directly experience it, to our embedded spaces of thought and speech – our invisible abstract conceptualizations and our construals for linguistic communication.

Viewpoint itself is universal. And alongside a great deal of crosslinguistic variation in how viewpoint structures are linguistically categorized and represented, there are also evident crosslinguistic patterns. Such patterns should come as no surprise: human neural architecture and experience put important constraints on the ways we are able to access perspectival construals. Both complex linguistic marking systems, and high-level literary and artistic manipulation of viewpoint, are built on shared early experience: for example, Primary Scenes (C. Johnson 1996, 1999a,b; Grady 1997a,b) link humans' experience of visual viewpoint with locational proprioception, and with spatiomotor strategies for access and reaching objects.

Even more interestingly, we are not just *capable* of multiple viewpoints; we are in fact incapable of keeping to one single viewpoint of space, or of cognitive structure, when other humans are present. A situation involving multiple humans is necessarily structured, for participants and for human observers, via complex multiple viewpoints. This is at least partly because of our mirror neurons (Rizzolati and Arbib 1998; Rizzolati, Fogasi and Gallese 2001) – some of the same neural activation involved in grasping, touching and acting on objects is also involved in viewing object manipulation by other primates. Thus, as Katherine Young (p.c.; cf. Young 2002) says, our bodies are naturally and constantly *occupied* by the neural patterns of surrounding human bodies – but not, apparently, in the same way by the spatial and force-dynamic relationships of inanimate objects. We are constantly aware of our bodily proximity to objects around us: but when another human is present, we are also unavoidably aware not only of our own human bodily affordances, but of (his or) hers as well – what can she reach, what can she see, and so on.

So a human neural system is from the start constrained to experience Ego's body as special and different - when our mirror motor neurons fire in response to watching someone pick up an apple, the non-mirror ones don't; so we neither use our own muscles to grasp a phantom apple, nor hallucinate that we are picking up an apple. On the other hand, the shared mirror neuron firing leads us naturally to categorize our own actions with those of others – and therefore to use the same verbs (*pick up, kick*) regardless of identity of the agent. This is in a sense a remarkable achievement: the visual and motor and tactile experience of picking up an apple is radically different from that of viewing such an action. Such "dual" experience of others' actions may well be part of the underlying basis for humans' universal ability to treat and understand other humans as conscious social agents like themselves, and thus to develop Theory of Mind. Without Theory of Mind – and the confidence that others also have Theory of Mind – complex viewpoint structures would be impossible. Other species do not, as far as we can tell, build complex counterfactual spaces, but they also don't (and can't) worry about specifically human concerns such as whether Joe has figured out that Chris and I are attracted to each other (*cf.* Zunshine [2006] for discussion of how these space embeddings play out in literary texts).

The Mental Spaces framework provides a way to represent these viewpoint phenomena. A mental space is a partial and local conceptual representation, which can be mapped onto or combined with other such spaces to build complex human conceptual structure. Mental spaces differ from other constructs, such as possible worlds, in being cognitive. A Mental Space analysis of linguistic meaning does not presuppose that there is some reality to which a speaker's understanding can be compared; all we humans have is our cognitive models of the world, based on

embodied experience. These are very powerful. For example, suppose we read a news story about a rock singer who is suing a magazine for libel in Italy. We have imagined mental spaces for law suits, journalism, libel, rock singers' lifestyle, and Italian culture, structured by frames with which we are familiar. Combining these, we build a richer cognitive representation than the news story could possibly evoke alone. Indeed, like most language, and as cogently remarked by Fauconnier (1994, intro. to 2nd edition of *Mental Spaces*), the news story's actual words and constructions are only prompts to readers to engage in space-building – they would be useless without the speaker's knowledge base and space-combining (or *blending*) abilities.

Since a given mental space is always attached to some perceiver or cognizer, mental spaces necessarily structure viewpoints. Of course, let's suppose that you and I are sitting next to a table, and both of us independently reach very much the same assessment of the table's physical size, hardness, and other physical characteristics. We are unlikely to notice that we have different construals of the table – so we feel as if we have access to Reality, when in fact we each have access to similar experiential data of the same object via similar embodied neural systems. But language is there not only to represent these unproblematic convergences in conceptual structure, but also to represent all the more complex situations which can arise as humans construe situations in varied ways.   Verbs such as *say* and *think*, which explicitly mark expressed or unexpressed cognitive states, necessarily build mental spaces, as Fauconnier has said.

Because humans can embed mental spaces, and hold contradictory spaces in mind at once, they can produce not only represented speech and thought, but also negation and counterfactual conditionals. They can also notice discrepancies between accessible spaces in complex networks, producing effects such as irony - discussed by Tobin and Israel in this volume – and humor.

Linguistic viewpoint.

Let us begin by surveying some of the range of linguistic forms which are markers of viewpoint. We shall label as *linguistic viewpoint* all the different ways that content is linguistically presented and construed differently depending on (at least) the following range of factors noted by linguists:

(1) Where the Speaker and Addressee are assumed to be, and what they are thought of as being able to see, be able to reach, etc. English uses such as **here, there, this, that, next door, …** depend for their reference on implicit information about the Speaker's and/or Addressee's presumed locations and their spatial relationships and access to objects designated. (Fillmore 1997[1971])

(2) When the Speaker and Addressee are assumed to be: just as with spatial terms, grammatical tenses and linguistic usages such as **now, then, tomorrow, last year…**depend for interpretation on the presumed time of utterance, writing, reading, or other communicative act. Deixis is neutral as to *scale* of construal. English *here* could mean the room we are in, the town or state we are in, or the planet where we are, each of which is more accessible than the contrasting *there*. Similarly, I can say *this pencil, this side of the room, this side of the Atlantic*, or *this side of the galaxy*. The less recognized deictic *home* works similarly; in a

science fiction story it can easily mean a character's "home" planet. (See Fillmore 1997[1971]; a mental space treatment of tense is laid out in Fauconnier 1997 and Cutrer 1994.)

(3) What the Speaker and Addressee are assumed to know, think, presuppose, and be able to calculate mentally about whatever mental space is involved. Examples of markers which give such clues are:

(a) Determiners. The choice of *a* as opposed to *the* says something about the Speaker's assumptions about the Hearer's ability to identify a referent. A mental space approach to determiners is laid out in Fauconnier 1985{1994}.

(b) Pronouns, address forms or honorific markers. The choice of a formal/distant rather than an informal/close second-person pronoun in languages with such a distinction (*tu* as opposed to *vous* in French) says something about the Speaker's construal of the social interaction, as do address choices such as *Professor Smith* versus *Mary* versus *Ma'am*.

(c) Connectives and evidential markers. Choosing *if* as opposed to *when* or *since* indicates the Speaker's lack of full positive epistemic stance commitment to the relevant mental space (Fillmore 1986, 1990; Dancygier and Sweetser 2000, 2005). A "hearsay" evidential marker (or non-grammatical marking such as *I hear that*) indicates the Speaker's lack of direct experience of the event referred to (Chafe and Nichols 1986).

(e) Presuppositional lexical items. The classic example is *stop*: saying either *Chris stopped smoking* or *Chris didn't stop smoking* indicates the Speaker's assumption that Chris smoked (Stalnaker 1974)). A mental space analysis of lexical presuppositional structures can be found in Ferrari and Sweetser (this volume).

(4) What the Speaker and Addressee feel about the contents of the relevant spaces - how they evaluate them affectively, culturally, etc. Such evaluation includes:

(a) Framing (Fillmore 1982, 1985). Calling a given behavior *thrifty* as opposed to *stingy* may not actually depict a contrasting set of behaviors, but certainly indicates that the speaker frames the (possibly identical) behavior of reluctance to spend money as in the one case prudent and reasonable resource conservation, and in the other case unreasonable and possibly selfish refusal to use resources as appropriate.

(b) Affective markers. Starting a sentence with *hopefully* marks the Speaker's positive emotional assessment of the eventuality mentioned, just as *maybe* marks her epistemic assessment. Dancygier and Sweetser (2005) have proposed the need to distinguish constructional semantics of *positive emotional stance* from Fillmore's (1990b) related concept of positive interest, in space-building constructions such as *if only*.

Many of these forms might better be analyzed as in some way negotiating Speaker and Hearer viewpoints, more in line with Verhagen's (2005) *intersubjectivity* than with the hypothesis that they simply mark Speaker viewpoint. Definite articles, for example, clearly mark some idea of shared cognitive accessibility, or Speaker's assessment of accessibility to the Hearer; negation and stance verbs (see Dancygier, this volume) negotiate stance between Speaker and Hearer. Intersubjectivity will be discussed in more detail in the next section.

But the list above is the tip of the iceberg. Since the whole point of mental spaces is precisely that humans can manage to separate (or blend) cognitive representations from different cognizers or experiences, naturally language can also express what *imagined* participants can reach, touch, perceive, know, think, presuppose, calculate, and feel about relevant spaces – not just what present speakers and hearers may be cognizing. And we may add that if we take embodied cognition seriously, all hearers and readers are imagined hearers/readers – we have no direct access to their cognitive states, so we are always speaking or writing to a reader or a hearer whose knowledge states, presuppositions, affect (and so on) we are estimating or imagining.

Like markers of thought and speech, linguistic markers of affective and perceptual states are themselves builders of mental space structure. A headache or fatigue can only be directly known by the experiencer, and so third-person sentences such as *His head ached* or *He felt tired* would require a non-direct evidential marker in some languages (a hearsay marker perhaps, indicating that he told me about his headache). As a result, such sentences can also indicate character viewpoint in a third-person narrative, since a separate narrator (not blended with the character) would not be able to speak directly about these aspects of character experience. Similarly, verbal aspect correlates with particular relationships to experiential states; imperfective aspect, which marks viewing an event from inside its temporal extent, is correlated with character-viewpoint in fiction for this reason. *It snowed that night* is not how the character experiencing the snow would put it; she would say or think *It is snowing*, or in Free Indirect Style accommodated to the narrator's past tense use, *It was snowing*. Vandelanotte's and Nikiforidou's papers in this volume examine some of the special literary stylistic exploitations of linguistic viewpoint markers in literary texts.

Metaphor adds yet more layers. Spatial closeness and distance, for example, have basic correlations with social intimacy and "distance" – the perfect basis for a Primary Metaphor in Grady's and Johnson's terms (C. Johnson 1996, 1999a,b; Grady 1997a,b). Unsurprisingly, therefore, most cultures use deictic terms to indicate social relationship and differentiation as well as spatial relation to Ego. Since temporal distance is correlated in turn with epistemic distance, a further Primary Metaphor (Fleischman 1989) motivates the use of temporal markers to indicate cognitive and social reticence.

In short, viewpoint is marked by just about anything which builds a particular individual's mental space construal in ways specific to that individual's cognitive and perceptual access. Authors of fiction exploit this constantly (Banfield 1982) – in a novel where narrative viewpoint shifts between characters, labeling a character *Mommy* as opposed to *Chris* can be quite sufficient to let us know which other character's viewpoint is currently on stage. Linguists have often preferred to focus on explicit grammaticized markers whose function is to express shifts in viewpoint,

such as pronouns or deictic markers. But in fact, any linguistic form choice is evidence concerning some mental space – and hence about the relevant cognizer's viewpoint.

Subjectivity and deixis.

Many analysts have noted that very often the presence and activity of the speaker or conceptualizer is left largely *implicit* in linguistic forms. Many utterances are not "about" the speaker – for example, *Joe walked into the café* seems to be primarily about the event of Joe's movement into the café. However, the past tense of the verb tells us that we should construe the reference event as taking place before the time of utterance. The use of *I, we*, or *you* can be seen as foregrounding, or bringing into explicit content, parts of the Ground (as Langacker [1987, 1990, 1991] has labeled the Speaker's or Conceptualizer's communicative setting) – these pronouns both refer to elements of the Ground, and require access to that Ground, to identify the referent.

Performative verb forms make certain aspects of the speech-act Ground into the explicit content of the utterance. An oath such as *I, Barack Hussein Obama, do solemnly swear that I will faithfully execute the office of President of the United States*...is intended to make it maximally clear who the speaker is, and exactly what speech act is being performed. Even the referent of *I* is explicitly specified, which is a highly unusual legal requirement for a spoken utterance. Most everyday interaction involves plenty of implicit shared Ground, which need not be mentioned. One friend could accept another's e-mailed dinner invitation by saying, *I accept your invitation*, but might more likely say something else such as, *Thanks, we can all come*, or *Great – what should I bring?* The prospective host would not have any trouble identifying the speech act of acceptance, or (given the e-mail address) the identity of *we* or *I*.

Langacker (1987, 1990, 1991) characterizes as **subjectivity** the "off-stage" *implicit* presence of the conceptualizer and the Ground in construal. The greater this implicit presence, the more subjective is a construal. Thus, for example, *Joe is sitting across the table from me* does require reference to the Ground; there's no mention of the speech act itself, but the speaker participant is explicitly referenced (brought "on stage") by using the first-person pronoun *me*. On the other hand, *Joe is sitting across the table* could be interpreted as meaning "across the table from me," even though there is no explicit first-person pronoun. It could also be interpreted as meaning "across the table from whoever is referentially accessible in the discourse context" – again, without explicit reference to Speaker, Hearer and discourse context. These construals are thus more subjective (or less objective) than the construal involved in *Joe is sitting across the table from me*.

Similarly, Traugott (1982, 1989) uses the term **subjectification** to refer to a unidirectional trend in semantic change towards meanings which are increasingly rooted in the discourse context of the Speaker-Hearer interaction.[1] The development of an epistemic modal meaning from a root modal meaning is a case of subjectification. The root modal sense of *should*, for example, refers to some

---

[1] By unidirectional, it is not meant that all meaning change is towards greater subjectivity. When a word for "feather" comes to mean "pen" (like English *pen*, from Latin *pinna*), there is no change in the referential relation to the Ground. The claim is that semantic change – particularly when associated with grammaticalization - often results in increased subjectivity, but *does not* result in decreased subjectivity.

obligation incumbent on the agent but not necessarily involving the conversational participants: *He should be home by now* (e.g., because he's ten years old and his parents have rules). Epistemic *should*, on the other hand, refers specifically to the speaker's epistemic evaluation of the rest of the utterance's content: *He should be home by now* (meaning "I judge it probable that he's home by now"). Aspect markers frequently develop historically into tense markers, for similar reasons: tense, unlike aspect, involves reference to a Speaker's Now. Although Traugott does not stress this, she clearly does mean *implicit* reference to the Speaker-Hearer interaction: modals and tense markers don't mention the action of speaking or the participants explicitly, but rather depend on them for interpretation. Stein and Wright (1995) expand subjectivity to the point where it might well cover all of viewpoint; it is an interesting question whether one can draw a line between the two, but here we maintain that it is a useful distinction.

Neither of these understandings of subjectivity and subjectification deal primarily with the complexity of relationships within the Ground: the Ground, or the Speaker-Hearer interaction, is treated as a whole, all of which has special implicit status unless parts of it are brought explicitly on stage by direct mention. Verhagen's (2005) concept of *intersubjectivity* (also brought up in Traugott and Dasher 2002) further elaborates the Ground for us, reminding us that it includes sharing and negotiation of viewpoint between Speaker and Hearer. Sanders, Sanders and Sweetser (2009) have recently proposed an understanding of the Communicative Ground itself as a mental space network (the Basic Communicative Space network), including representations of the Speaker's and Addressee's epistemic spaces as well as the space of speech act interaction and one or more content spaces.

Deixis (Fillmore1997[1971]; Levinson 1996a,b, 2003) refers to the conventional use of linguistic forms whose meaning depends on the (implicit) Ground. As Levinson has pointed out, the typology of deictic systems shows some structures to be strongly favored. One very common system for spatial deixis is a three-way one, where A ("this") means "near the Speaker, or nearer to S than to the Hearer", B ("that") means "near H and not near S", and C ("yon") means "not near either S or H." Factors that turn out to be relevant are whether the located entity is manually accessible to S/H, whether it is visible to them, and whether it is possessed by, or in the custody of, one of them.

The question of custody or ownership complicates the spatial understanding of deixis and forces a more social construal. Hanks (1985) points out the use of a proximal Mayan deictic to refer to the cooking hearth which is farther away but used by the speaker, and a distal one to refer to the hearth which is currently right next to the speaker but used/owned by someone else. English deictic verbs also make reference to locations conventionally or habitually associated with S or H, rather than only to S/H's physical locations at speech time. In English it would be usual to say *Can you come to my party?* as an invitation, even if I am not at home when I invite you, and even if the party is not being held at my home.

An under-studied aspect of deixis is the phenomenon of deictic *displacement* or deferral: a clear example is that in English, the correct response to the invitation *Can you come to my party?* is *Sure, I'd love to come*, not *Sure, I'd love to go*. The invitation accepter might later say to a third party *I'm going to Sandy's party on Friday*, and would be unlikely to say *come* in this context. The acceptance utterance thus participated in the inviter's deictic structure, displacing the accepter's deictic center

to the inviter's. Notice that *I* and *you* are not similarly displaceable – a speaker would not, and could not, accommodate an addressee by using personal pronouns relative to his deictic structure rather than hers. Nor are these particular viewpoint extensions of spatial deictic *come* paralleled in the COME and GO verbs of all languages. Emanatian (1992) details a remarkable example from Chagga, where there is displacement also in temporal uses of *come* as a future-marker, rather as if English speakers said *I'm coming to V* with the meaning of a GONNA future.

Diagram 1, below, shows a treatment of the English addressee-centered usage of *come* as a blend. Note that the speaker of *Can I come to your party?* has not completely recentered her deictic field on the addressee – *I* still refers to the speaker, *you* to the addressee, so the S/H referential system is not displaced. But the spatial deictic coordinate space, which is most canonically centered on Ego (hence on the speaker) in this blend is mapped onto the addressee as center.

Input 1: S/H space
with EGO at Speaker

H: "you"

S: "I"

Input 2: deictic
coordinate space

H: "you"
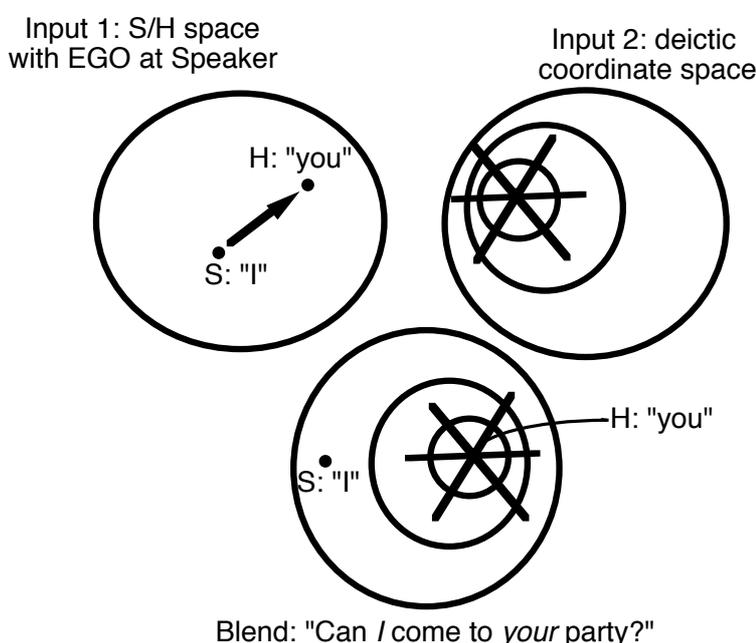
S: "I"

Blend: "Can *I* come to *your* party?"

FIGURE 1.

The crucial point here is that our everyday construal of personal viewpoint is a blend. It's a blend which is so common that it is hard to notice it. We normally experience our own bodies simultaneously as loci of our conscious Selves or Egos, agents of our speech and action, spatial sources of our fields of perceptual access and manual reach, interfaces of social interaction, and more. But as stated above, we also naturally create such models for other individuals around us – aided, very possibly, by our mirror neurons, which respond to other humans' grasping actions (for example) as well as to our own. Once that is accomplished, a speaker can naturally describe motion away from herself with *come*, if she is profiling the deictic field structure relative to another participant (and leaving out mention of her own deictic field). It seems much harder to displace *I* from the speaker's identity – one cannot say *Can you come to my party* meaning *Can I come to your party*? Embedded viewpoints are certainly possible, keeping separate mental spaces clearly separate (*You said, "Please come to my party"* – where *my* refers to the addressee within the addressee's quoted speech space) – but one cannot build this blend within a single space. As Nikiforidou points out in this volume, similar blends occur in temporal language, so

that a narrator's tense center may be combined with a character's temporal adverb to produce examples like *He was now living in retirement*; the opposite blend is in this case possible too (spoken English past narratives contain examples like, *So then the teacher leaves the room and the kids start talking*).

We also need to distinguish between a **viewer's perspective** and a full **deictic center** involving the speaker and the speech event. In English, *come* (and to a lesser degree, the less marked *go*) involve deictic centers, while verbs such as *arrive* involve a viewer's perspective which need not be that of the deictic center. *Arrive* clearly does take the perspective of the arrival point, in some sense. But *Sue has arrived in Los Angeles* could be spoken by either someone in Los Angeles or someone in San Francisco, and it is perfectly acceptable to use *arrive* with either a proximal or a distal deictic (*Sue has arrived there/here*). *Sue has come to Los Angeles*, however, is normally interpreted as spoken by someone in Los Angeles; and *come there* is distinctly strange. This is because not only *come*, but *here* and *there* invoke full deictic centers, unlike other perspectival adverbials such as *across the table*. As Langacker (1987, 1991) notes, *Sue jumped across the table* is normally interpreted as meaning "across the table from Sue's initial pre-jumping location". But *Sue sat across the table* means "across the table from someone whose perspective is taken" – the speaker and hearer are obvious options, but if we just mentioned Mary sitting at the table, then Mary is a good candidate.

Purely social viewpoint is built into many aspects of language, and indexes speaker-hearer interaction as surely as deictic physical viewpoint does. It is not normally referred to as deixis, and is generally studied by sociolinguists rather than semanticists. And yet consider the following attested scenario. A senior male university professor at the Sorbonne is about to introduce an honored speaker in a special lecture series; the speaker is his friend, his colleague and a senior professor from a university in South America. As they chat before the talk, they address each other by first names and use the informal French second-person pronoun *tu*. Moments later, as he addresses her in official welcome over the microphone, he is calling her *Professor X*, and using the formal address pronoun *vous*. An American introducer would not of course be able to change address pronouns, but would probably similarly switch from *Mike* to *Professor Jones* in moving from private conversation with the speaker to the public introduction: the audience is not part of the personal interaction which ratifies the first-name use. Similarly, an American academic would normally refer to faculty colleagues and office neighbors by first names, but in doing undergraduate advising, that same faculty member would probably suggest that the student consult "Professor X" to get references for her paper. These choices, then, are intersubjectively (in Verhagen's terms) indexical not only of the relationship between the speaker and the referent, but of the relationship between the hearer(s) and the referent, which may dominate in determining the choice.

Rubba (1996) has pointed out that supposedly spatial deictics such as *this* and *that*, *here* and *there*, are also used socially: the proximal ones may indicate "my cultural kind of neighborhood" and the distal ones "their (not my) cultural kind of neighborhood." Rubba's data is particularly compelling because, for other reasons, her interviews (concerning the English-only movement in California) all took place in the same room on the UCSD campus, relatively distant from all the (socially proximal or distal) neighborhoods referred to. But more such instances are readily accessible: those who listen at linguistics conferences will hear groups of same-

campus folks occasionally using *here* to mean "our campus", distant though that campus may be from the conference location.

Japanese honorifics are one much-studied area of social deixis. They are more complex than English or French address terms, since they apparently extend to cover identification of groups with individuals. Thus Dasher (1995) cites corporate telephone conversations where the secretary phones the President's wife to ask how he is recovering from the flu and says "How is the 'honored' President's health today," using the positive honorific for the President. This is an example of what is usually thought of as the central use: the secretary is lower in status than the President and so uses an honorific in addressing or referring to him. Moments later, however, the same secretary is canceling one of the President's external appointments for him, and says to the other executive's secretary "Our 'humble' President can't keep his appointment because he is ill today," using the (normally first-person-associated) humiliative marker. This is because in canceling the appointment she is acting on behalf of the President and his organization, and is thus identified with them; while in inquiring about his health, she is a family outsider and subordinate talking to an insider.

Collectively, the linguistic data suggest that we should in general be thinking of viewpoint as an intersubjective phenomenon (in Verhagen's (2005) sense) rather than a unitary first-person phenomenon – that is, addressees' and others' viewpoints are always relevant as well as the speaker's own viewpoint, in contributing to the speaker's choice of linguistically expressed viewpoint. We know that this is true with respect to how we negotiate spatial viewpoint and affordances; you would not ask someone to pass the salt unless it were *both* accessible to him and inaccessible to you. So we should not be surprised to find that similar joint determination applies to communicative and linguistic viewpoint structure.

Viewpoint across modalities.

Examination of some of the areas where perspective is known to be central – such as linguistic deixis (Hanks 1990), or gestural pointing (Kita 2003) – has helped motivate researchers to examine the much broader presence of perspectival phenomena in language and bimodal communication. Signed language forms necessarily embody portrayed viewpoints in ways which spoken language does not (Dudis 2004; Liddell 1998, 2003; Janzen 2008; and Schaffer's and Janzen's papers in this volume). Examining these phenomena has helped spoken language researchers to elaborate their understandings of the complexities of viewpoint. Data from all the relevant sources thus come together to push us towards a theory which includes mutual awareness of viewpoints between participants - and incorporation of addressees' and spectators' and readers' perceived viewpoints into the cognitive perspectives of speakers, narrators and writers. Intersubjective viewpoint construction is important at every level of communication, from co-speech gesture to literary narrative.

First, let us briefly examine gestural viewpoint. Gesture, like language, shows deictic centers, displacement phenomena and blended deictic structure. Our bodies are the most flexible and powerful **material anchors** (Hutchins 1995, Fauconnier and Turner 2002) for representing and expressing viewpoint. This is because, as is obvious but rarely stated, there is no more powerful icon for a bodily viewpoint than an actual body with an actual inherent viewpoint. Gesture therefore expresses spatial indexicality via embodied spatial indexicality – for humans, a pointing

gesture directs joint attention to a particular actual location relative to the actual pointing body. This in itself is a cognitive achievement; as Tomasello (1999, 2008) points out, great apes don't share human abilities to interpret points, though some dogs do.

But of course, just like linguistic deixis, gestural deixis is not that simple. Displacement or deferral of gestural deictic viewpoint is pervasive. Haviland (2000) notes what he refers to as *transposed* points in his Mayan data; these are are made relative to some imagined location rather than the gesturer's real current location. As an English example, I could say, *As you enter my office, the light switch is **here***, moving my left hand as if to turn a switch on a wall to my left. Since English is a relative spatial language, my fellow English speakers will understand that the light switch is located to the left of a person standing in the doorway and looking inwards - no matter which direction I'm currently facing, and no matter which direction my (non-present) office door faces. Crucially, they won't even think of looking for the light switch in the actual physical direction of my physical point in Real Space. Their understanding of this deferred point is a blend of the gesturer's current Real Space with the imagined space of my office door and its surroundings. If I were speaking an absolute spatial language such as Guugu Ymithirr (Levinson 1996a,b, 2003), interlocutors would instead probably create a different blended space, where my current absolute *direction* of stance (e.g. Northeast) was part of the input to the interpretation of the deferred point – that is, they might assume that someone entering my office would be facing Northeast.

In general, gesture allows Real Space blends of some or part of the gesturer's body with a character's body. For example, describing a character climbing a ladder, a gesturer may use her hands in "clambering" gestures as if going up an imaginary ladder, so that her hands represent the hands of the described climber. Alternatively, she may move one hand upwards in representation of the whole character's bodily upwards motion. This represents two rather different viewpoints on the same scene, the first a Character viewpoint (the gesturer is "being" the character in the blend) and the other a Narrator (or Global) viewpoint (the gesturer is more of an observer of the character, as represented by her hand), as described by Parrill's paper in this volume.

Gesture also uses Real Space more abstractly. It can exploit conventional Real-Space blends to give abstract meaning to spatial locations – for example, conventional mappings of Time onto Space allow forwards-directed pointing gestures to have future time reference for English speakers. This is complex and language-specific, but apparently cognitively pervasive for speaker-gesturers (Boroditsky and Ramscar 2002; Matlock, Ramscar and Boroditsky 2005; Núñez and Sweetser 2006). Gesturers can also establish *loci*, locations in Real Space, which correspond to referents in content being discussed, without necessary any mapping of spatial relations onto another imagined physical space – and even without the referents being physical entities. For example, loci can be entities being contrasted (my family and your family, logic and linguistics) or topics under discussion.

Signed languages famously not only represent the speaker's physical viewpoint, but use the visual-gestural modality to carry out linguistic deixis and reference. It is a question for research what similarities there may be between the use of referent *loci* in signed languages and the use of *loci* in gesture. But it is clear that in signed languages, pronouns consist of systematic pointing to (or moving signs towards or

away from) such *loci*, with grammatically determined hand shapes. Or, as some might say (Liddell 1998, 2003, Taub 2001), interaction with spatial *loci* does much of the same referential work for signed languages that spoken pronouns or person-agreement affixes do in spoken languages. And referring to objects and people who are actually present is normally done by exploiting (pointing to, signing towards, etc.) their actual locations.

Signed languages can also exploit blended viewpoint structures of the kind seen in gesture, but in a far more systematic and grammaticalized way, as Shaffer and Janzen discuss in their papers in this volume. American Sign Language grammars in particular lay out a system of *role shifts*, where a signer wishing to take on the role of a character rotates her body noticeably (perhaps twenty degrees or so) to one side or the other, and ends eye contact with the interlocutor (since the represented character is naturally not addressing the interlocutor). The signer's signed utterances, facial expressions and gestural enactments while in role shift are taken to be those of the character. The character's discourse may involve a different imagined physical space blended with the signer's actual Real Space (if he drives a car, the signer will be gripping an imagined steering wheel), and also different referential loci depending on what the character needs to refer to. That is, the signer's Real Space is simply re-mapped to the character's spaces, rather than to the signer's own mental spaces and physical surroundings. Return to facing the interlocutor, with eye contact, indicates an end to the shifted role and a return to the signer's own viewpoint. The resulting Role Shift blends are *grounded blends* (blends one of whose inputs is the Real Space) – as are many gestural blends.
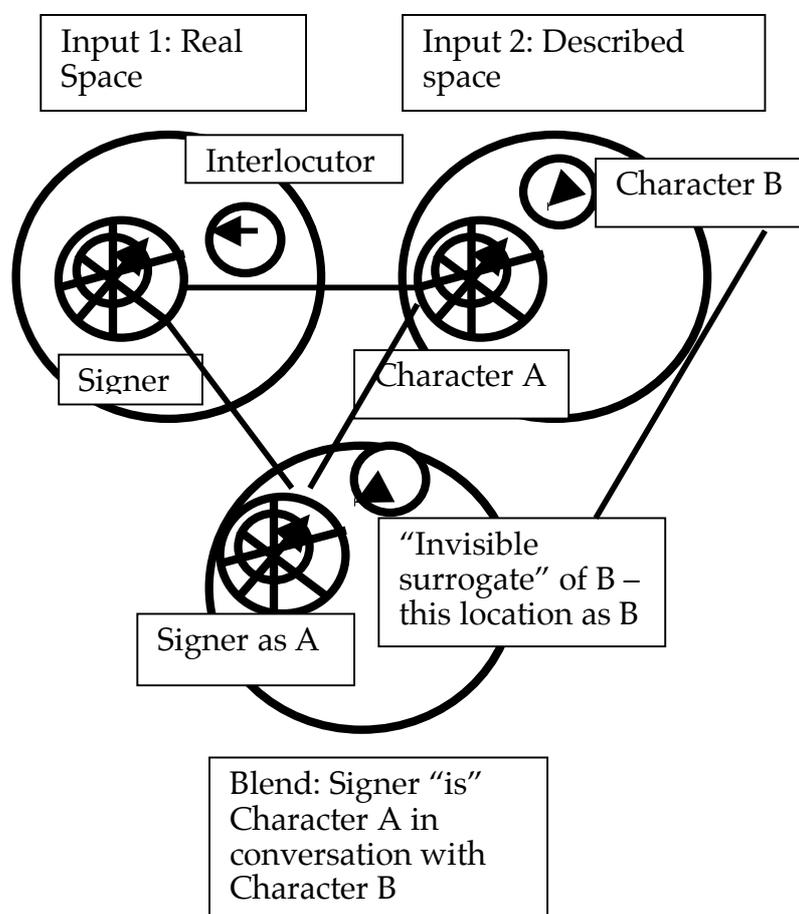

Figure 2.

Figure 2 shows the blend which results from role shifts. Here the signer is taking on the role of a described Character A; she turns slightly towards her right, thereby aligning her own Ego-centered spatial coordinates with those of the imagined Character A. In this role-shifted posture, the signer signs; we understand that this represents the character as signing to another character, who has probably been identified in the discourse. So in the described world, we construct Character B, facing towards Character A and conversing with her - but there is no Real Space representation of B, only a location towards which the Signer faces. In the Real Space, there is thus no Character B, only the Real-Space interlocutor(s) of the signer. In the described space and in the Blend, the Real-Space interlocutors do not exist, but Character B is present. A moment later, the signer may turn towards her left, creating a new blend in which she represents A's interlocutor (Character B), who is imagined to be facing leftwards towards (and having a conversation with) Character A. Her spatial coordinates will then be blended with those of the second character. Thus, although just a moment ago she was looking to her right at an imagined interlocutor, a moment later (as that interlocutor) she may be looking to her left at the same character she was enacting a moment earlier.

In this volume, Janzen further outlines another common ASL narrative viewpoint-changing strategy, not previously attended to by linguistic analysts, which involves *rotation* of the imagined space mapped onto the signing space, instead of rotation of the signer's body. If a signer is representing two characters who are facing each other and have opposite viewpoints, the signer may reassign the signing space alternately to each of the two characters. At least some of the same imagined physical space (the surroundings of the two characters) will be represented in the same signing space (the Real Space in front of the signer's body) – but the mappings will be rotated to fit each character's perspective. This is extremely different from role shift, but seems common and natural in ASL narratives.

In both role shifts and sign-space rotations, as in gestural blended "character viewpoints" (but unlike spoken-language viewpoint), a viewpointed body iconically represents another viewpointed body. This has immense consequences for representation of viewpoint, which often needs no active expression beyond offering a bodily posture. As Sweetser (2009) has pointed out, English speakers normally gesture about ongoing processes (solving a problem, getting a Ph.D.) with hand motions outwards from the body rather than inwards towards the body – that is, the start of the activity is mapped onto points closer to the body, and later stages are mapped onto more distant points. (The exception to this generalization is reference to physical motion towards the speaker, where actual physical deictic structure dominates and a speaker will gesture towards herself as she says *Come here*.) This is probably because we do experience our body as a source of actions – when we grasp something, we stretch out our hand to do so, and we stop stretching it out when our goal is attained. There is no such deictic structure built into spoken-language linguistic descriptions of getting a Ph.D. or solving a problem, confirming the instinctive impression that the visual/gestural modality is even more pervasively viewpointed than the spoken one.

Conclusion

The evidence is overwhelming that human thought is essentially embodied – literally in the sense that it is neural patterns, and also in the broader sense that its character necessarily reflects embodied perception and reasoning (see Lakoff and Johnson 1999, Gibbs 2006 for overviews of this issue). Embodied human cognition is inherently pervaded by viewpoint, and therefore communicative manifestations of cognition are too, in a remarkably diverse range of ways. This makes perfect sense if we assume that linguistic forms are (Fauconnier 1985[1997], 1997) prompts for mental space building, and that the process of space building involves mental *simulation* of the situations and events referred to. *Simulation-based* semantics (*cf.* Feldman 2006) was prompted by the discovery that mirror motor neurons fire not only when humans witness events but also when they hear or read about them (Rizzolati and Arbib 1998; Rizzolati, Fogasi and Gallese 2001). It is now a theoretical framework being developed by a large community of researchers. Those researchers are taking seriously the idea that although a sentence such as *Joe walked into the café* is in itself neutral as to whether it is describing the event as viewed from inside or outside the café, the simulations which it prompts in listeners or readers can't be equally neutral. Better understanding of all the varied manifestations of viewpoint in communication is therefore crucial to developing a more cognitively realistic understanding of processing – even of processing sentences which don't seem to prompt a particular viewpoint. And, as has been observed again and again (McNeill (1992, 2000, 2005; Cienki 1998; Goldin-Meadow 2003; Narayan's paper in this volume) multimodal data will give evidence about communicators' cognitive processing which is not accessible from just a single modality of output.

We also feel that, just as some theoretical frameworks are shown by these papers to be increasingly helpful in examining viewpoint, other conceptual baggage is being shed as the field moves forwards. Readers will note that, perhaps surprisingly in a book on viewpoint, this introduction has not discussed the role of context in determining viewpoint. This is not because we or our authors are inattentive to contextual factors in communicative situations – on the contrary. Rather, we feel that the cover term *context* is too broad to be meaningful, when the range of studies is as varied as this. For analysts of literary texts, *context* crucially means the surrounding text. For analysts of face-to-face recorded communication, although the immediate linguistic "record" is of course relevant to production and interpretation at any time, the Real Space is also relevant, and can become relevant in unforeseen ways. When someone describes a person pounding on a table, having a surface to gesturally pound during the description suddenly becomes relevant context – although the Real-Space surface may not have been very important to the communicative exchange till that point. In all communicative situations, context might also be said to include shared models of how communication works (spoken, written, or signed) – so Sanders, Sanders and Sweetser's (2009) Basic Communicative Space Network is context, as well as the actual circumstances of the exchange. Also ever-relevant to every communication are shared beliefs and frames which can be recruited to populate spaces – whether local or cultural frames. And the list could continue. For such a varied set of studies, we feel that it is more important to focus on the actually relevant aspects of context – different kinds of contextual affordances are present in different modalities and situations.

What this volume is doing, therefore, is bringing together researchers who work on widely differing kinds of data (experimental data, face-to-face recordings, corpora, literary texts) and in different modalities (signed and spoken languages as well as co-speech gesture) to bring a new overall picture of viewpoint. In some cases, our

authors are building on long traditions but significantly elaborating them (Vandelanotte) or reinterpreting them to offer new generalizations in the context of current grammatical and cognitive theories (Nikiforidou, Tobin and Israel). In other cases, the relevant phenomena are either underdescribed (Shaffer, Parrill) or previously undocumented (Janzen, Narayan). What fascinated the editors, when we initially gathered this group of contributors, was the degree to which the participants' presentations converged on a shared understanding – and also the degree to which that surprised them. We had initially thought that interaction between them was sure to be productive. But when we realized that this collaboration had helped the central participants to become more aware of the degree to which their work was addressing related issues, we were further convinced that the rest of the linguistic and cognitive science communities could also find this volume a useful contribution.

References

Banfield, Ann. 1982. *Unspeakable sentences: Narration and representation in the language of fiction.* Boston: Routledge and Kegan Paul.

Boroditsky, Lera and Michael Ramscar. 2002. The roles of body and mind in abstract thought. *Psychological Science* 13:2, 185-188.

Chafe, Wallace and Johanna Nichols (eds.) 1986. *Evidentiality: The linguistic encoding of epistemology.* Norwood NJ; Ablex.

Cienki, Alan. 1998. Metaphoric gestures and some of their relations to verbal metaphoric counterparts. In J.-P. Koenig (ed.), Discourse and cognition: Bridging the gap. Stanford CA: CSLI. 189-205.

Dancygier, Barbara. 1998. *Conditionals and prediction*. Cambridge: Cambridge University Press.

Dancygier, Barbara and Eve Sweetser. 2000. Constructions with *if, since* and *because*: Causality, epistemic stance and clause order. In E. Couper-Kuhlen and B. Kortmann (eds.) *Cause, condition, concession, contrast*. Berlin: Mouton de Gruyter.

Dancygier, Barbara and Eve Sweetser. 2005. *Mental spaces in grammar: Conditional constructions.* Cambridge: Cambridge University Press.

Dasher, Richard. 1995. *Grammaticalization in the system of Japanese predicate honorifics.* Doctoral dissertation, Stanford University.

Dudis, Paul. 2004. Body partitioning and real-space blends. *Cognitive Linguistics* 15:2, 223-238.

Emanatian, Michele. 1992. Chagga 'come' and 'go': Metaphor and the development of tense-aspect. *Studies in Language* 16, 1-33.

Fauconnier, Gilles. 1985 [1994]. *Mental Spaces*. 2nd ed. Cambridge: CUP.

_____. 1997. *Mappings in Thought and Language*. Cambridge: CUP.

Fauconnier, Gilles. and Eve Sweetser (eds). 1996. *Spaces, Worlds, and Grammar*. Chicago: Chicago University Press.

Fauconnier, Gilles and Mark Turner. 2002. *The way we think. Conceptual blending and the mind`s hidden complexities*. New York: Basic Books.

Feldman, Jerome A. 2006. *From molecule to metaphor: A neural theory of language.* Cambridge MA: MIT Press.

Fillmore, Charles J. 1997 [1971]. *Lectures on Deixis*. Stanford CA: CSLI. (originally published 1971 by the Indiana University Linguistics Club as *The Santa Cruz Lectures on Deixis*.)

      . 1990. Epistemic stance and grammatical form in conditional sentences. *CLS* 26, 137-62.

      . 1990(b). The contribution of linguistics to language understanding. In Aura Bocaz (ed.), *Proceedings of the 1st Symposium on Cognition, Language, and Culture.* Universidad de Chile. 109-128.

      . 1986. Varieties of conditional sentences. *ESCOL* 3, 179-95.

      . 1985. Frames and the semantics of understanding. *Quaderni di semantica* 6:2, 222-54.

      . 1982. Frame semantics. In Linguistic Society of Korea (ed.) *Linguistics in the morning calm.* Seoul: Hanshin. 111-38.

Fleischman, Suzanne. 1989. Temporal distance: A basic linguistic metaphor. *Studies in Language* 13: 1-50.

Gibbs, Raymond. W. 2006. *Embodiment and cognitive science*. Cambridge: Cambridge University Press.

Goldin-Meadow, Susan. 2003. *Hearing gesture: How our hands help us think.* Cambridge MA: Harvard University Press.

Grady, Joseph. 1997a. THEORIES ARE BUILDINGS revisited. *Cognitive Linguistics* 8, 269-290.

      . 1997b. *Foundations of meaning: Primary metaphors and primary scenes*. Doctoral dissertation, University of California, Berkeley.

Hanks, William F. 1990. *Referential practice: Language and lived space in a Maya Community*. Chicago: University of Chicago Press.

      . 1996. Language form and communicative practices. In J. J. Gumperz and S. C. Levinson (eds.), *Rethinking linguistic relativity*. Cambridge: Cambridge University Press, 232-70.

Haviland, John. 2000. Pointing, gesture spaces and mental maps. In D. McNeill (ed.) Language and gesture. Cambridge: Cambridge University Press. 13-46.

Hutchins, Edwin. 1995. *Cognition in the wild*. Cambridge MA; MIT Press.

Janzen, Terry. 2005. *Perspective Shift Reflected in the Signer's Use of Space*. CDS/CLCS Monograph Number 1, Dublin: Center for Deaf Studies, University of Dublin Trinity College.

Johnson, Christopher. 1996. Learnability in the acquisition of multiple senses: SOURCE reconsidered. *Proceedings of the 22nd Annual Meeting of the Berkeley Linguistics Society*, 469-81.

      . 1999a. *Constructional grounding: The role of interpretational overlap in lexical and constructional acquisition*. Doctoral dissertation, University of California, Berkeley.

      . 1999b. Metaphor vs. conflation in the acquisition of polysemy: The case of SEE. In. M. Hiraga et al. (eds.), *Cultural, psychological and typological issues in cognitive linguistics.* Amsterdam: John Benjamins. 155-171.

Kita, Sotaro (ed.) 2003. *Pointing: Where language, culture and cognition meet*. Mahwah NJ: Lawrence Erlbaum Associates.

Lakoff, George and Mark Johnson. 1999. *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar*. Vol.1. Stanford CA: Stanford University Press.

      . 1990. Subjectification. *Cognitive Linguistics I*: 5-38.

      . 1991. *Foundations of cognitive grammar*. Vol. 2. Stanford CA: Stanford University Press.

Levinson, Stephen. 1996a. Language and space. *Annual review of anthropology* 25, 353-382.

_____. 1996b. Relativity in spatial conception and description. In J. Gumperz and S. Levinson (eds.) *Rethinking linguistic relativity.* Cambridge: Cambridge University Press. 177-202.

_____. 2003. *Space in language and cognition: Explorations in cognitive diversity.* Cambridge: Cambridge University Press.

Liddell, Scott. 1998. Grounded blends, gestures and conceptual shifts. *Cognitive Linguistics* 9:3, 283-314.

_____. 2003. *Grammar, gesture and meaning in American Sign Language.* Cambridge: Cambridge University Press.

Matlock, Teenie, Michael Ramscar and Lera Boroditsky. 2005. The experiential link between spatial and temporal language. *Cognitive Science* 29, 655-664.

McNeill, David. 1992. *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

_____. (Ed.) 2000. *Language and gesture.* Cambridge: Cambridge University Press.

_____. 2005. *Gesture and thought.* Chicago: University of Chicago Press.

Núñez, Rafael and Eve Sweetser. 2006. With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science* 30, 401-450.

Parrill, Fey and Eve Sweetser. 2004. What we mean by meaning: Conceptual integration in gesture analysis and transcription. *Gesture* 4: 197-219.

Rizzolati, G., and M. Arbib. 1998. Language within our grasp. *Trends in neurosciences* 21:5, 188-194.

Rizzolati, G., L. Fogasi and V. Gallese. 2001. Neuropsychological mechanisms underlying the understanding and imitation of action. *Nature Neuroscience Reviews* 2:661-670.

Rubba, Johanna. 1996. Alternate grounds in the interpretation of deictic expressions. In Gilles Fauconnier and Eve Sweetser (eds), 227-261.

Sanders, José and Gisela Redeker. 1996. Perspective and representation in speech and thought in narrative discourse. In Gilles Fauconnier and Eve Sweetser (eds.), 290-317.

Sanders, José, Ted Sanders and Eve Sweetser. 2009. Causality, cognition and communication: A mental space analysis of subjectivity in causal connectives. In Ted Sanders and Eve Sweetser (eds.) *Causal categories in discourse and cognition*. Berlin: Mouton de Gruyter. 19-59.

Sizemore, Marisa and Eve Sweetser. 2008. Personal and interpersonal gesture space: Functional contrasts in language and gesture. In A. Tyler et al. (eds.) *Language in the context of use: Cognitive and discourse approaches to language and language learning*. Berlin: Mouton de Gruyter.

Stein, Dieter and Susan Wright (eds.). 1995. Subjectivity and Subjectivisation. Cambridge: Cambridge University Press.

Sweetser, Eve. 2009. What does it mean to compare language and gesture? Modalities and contrasts. In J. Guo et al. (eds.) *Crosslinguistic approaches to the psychology of language: Research in the traditions of Dan Isaac Slobin*. New York: Psychology Press. 357-366.

Stalnaker, Robert C. 1974. Pragmatic Presuppositions. In M. Munitz and P. Unger (eds.), *Semantics and Philosophy*. New York: New York University Press.

Taub, Sarah F. 2001. *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge: Cambridge University Press.

Tomasello, Michael. 1999. *The cultural origins of human cognition*. Cambridge MA: Harvard University Press.

_____. 2008. *Origins of human communication*. MIT Press: Bradford Books.

Traugott, Elizabeth Closs. 1982. From Propositional to textual and expressive meanings; Some semantic-pragmatic aspects of grammaticalization. In W.P. Lehmann and Y.Malkiel (eds.). *Perspectives on Historical Linguistics*, Amsterdam:John Benjamins, 245-71.

———. 1989. On the rise of epistemic meanings in English: an example of subjectification in semantic change. *Language 57*: 33-65.

Traugott, Elizabeth Closs and Richard B. Dasher. 2002. *Regularity in semantic change*. Cambridge: Cambridge University Press.

Verhagen, Arie. 2005. *Constructions of intersubjectivity: Discourse, syntax, and cognition.* Oxford: Oxford University Press.

Young, Katharine G. 2002. The dream body in somatic psychology: The kinaesthetics of gesture. *Gesture* 2:1, 47-50.

Zunshine, Lisa. 2006. *Why we read fiction: Theory of mind and the novel*. Ohio State University.