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

Eve Sweetser

In Wallace Stevens' poem "The anecdote of the jar", the narrator has put a jar on a Tennessee hilltop, with apparently dramatic results. The jar, we are told, *took dominion everywhere*, imposed order on *the slovenly wilderness*, and forced the wilderness to *surround* the hill on which it sat. Before the jar, it seems there were no human objects in the wilderness. And natural features like hills and valleys have no inherent preexisting relation between landmarks and "surroundings", since the relation between figure and ground is an aspect of a viewer's perceptual construal, rather than of the perceived objects.† The jar – a human artifact and a durable reminder of the human presence of the narrator – has proliferated new human-centered construal of space. It has made the land into a landscape.

A stretch of country with a human in it is no longer just a stretch of country – it is also a human's egocentric conceptualization of that physical area. 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 that 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 that the rest of this book confirms: cognitive perspective starts with bodily viewpoint within a real physical Ground of experience. In a mental space

† Readers are invited to enjoy the full text of the poem, as one of the best examinations of viewpoint and perspective we know. Copyright issues sadly prevented us from including it in the paper.

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 cross-linguistic variation in how viewpoint structures are linguistically categorized and represented, there are also evident cross-linguistic 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 (Johnson 1996, 1999a, 1999b; Grady 1997a, 1997b) 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 some of the ways in which our brains process other humans' activities. Research on monkeys has shown that some of the same neural activation patterns involved in grasping, touching, and acting on objects is also involved in viewing object manipulation by other primates (Rizzolati and Arbib 1998; Rizzolati *et al*. 2001) – these are so-called mirror neurons. Mirror neuron data come from monkeys, but have motivated work with human subjects, which has shown that there are in fact quite widespread patterns of such activation – motor cortex areas involved in actions are also partially involved in observing or hearing or reading about similar actions (Pulvermüller *et al*. 2001; Hauk *et al*. 2004; Buccino *et al*. 2005). Thus, as Katherine Young (personal communication; 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 she can reach, what she can see, and so on.

So, from the start, a human neural system is 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 do not fire; 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 the 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 do not (and cannot) 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).

More surprisingly, we build viewpointed, sensory simulations in response to linguistic stimuli as well – and not just simulations of actions described, but of the situations involved (Richardson *et al*. 2003), and multimodally (Pecher *et al*. 2003, 2004). Subjects are faster at recognizing a vertically oriented picture of a nail than a horizontal one, after being primed with the sentence *He hammered the nail into the floor*, which implies a vertical orientation of the nail being hammered; the reverse is true after being primed with a sentence like *He hammered the nail into the wall* (Stanfield and Zwaan 2001). Subjects also simulate the temporal structure and completion of described actions differently depending on the linguistic aspect (Bergen *et al*. 2010) expressed. These are only a few examples of the increasing evidence (for overviews, see Barsalou *et al*. 2005; Barsalou 2010) that language prompts us to "run" simulations in the brain – simulations that are necessarily viewpointed, because the experiences on which they are based are necessarily viewpointed. This volume focuses on a much wider range of ways in which language and bimodal communication represent viewpoint – many of which have not yet been tested for simulation correlates in the brain. But since the volume is a contribution to the study of language in the context of *embodied* (or *grounded*) cognition, it is important for these works to be set in the context of evidence that communication activates the embodied neural system's representation of viewpoint.

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 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 lawsuits, 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 (1985 [1994], introduction to 2nd edn. 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. Let us 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 that 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 that 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, and so on. 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 we are on, 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 that 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, 1990a; Dancygier 1998; 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).
   (d) 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, and so on. 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, 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, the use of such sentences without hearsay marking 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* (for a mental spaces analysis of free indirect style, see Sanders and Redeker 1996). Vandelanotte's and Nikiforidou's chapters 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 (Johnson 1996, 1999a, 1999b; Grady 1997a, 1997b). 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 that 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 emailed 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 email address) the identity of *we* or *I*.

Langacker (1987, 1990, 1991) characterizes as **subjectivity** the "offstage" *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 is no mention of the speech act itself, but the speaker participant is explicitly referenced (brought "onstage") 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 that 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 obligation incumbent on the agent, but not necessarily involving the conversational participants: *He should be home by now* (e.g. because he is 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 is 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 do not 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 onstage 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. And Sanders *et al*. (2009) have 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 acts and speech act interaction between participants, and one or more content spaces.

Deixis (Levinson 1996a, 1996b, 2003; Fillmore 1997 [1971]) 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 (1990, 1996) points out the use of a proximal Mayan deictic to refer to the cooking hearth that is further away but used by the Speaker, and a distal one to refer to the hearth that 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.

Figure 0.1 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, and *you* to the Addressee, so the S/H referential system is not displaced. But the spatial deictic coordinate space, which is most canonically

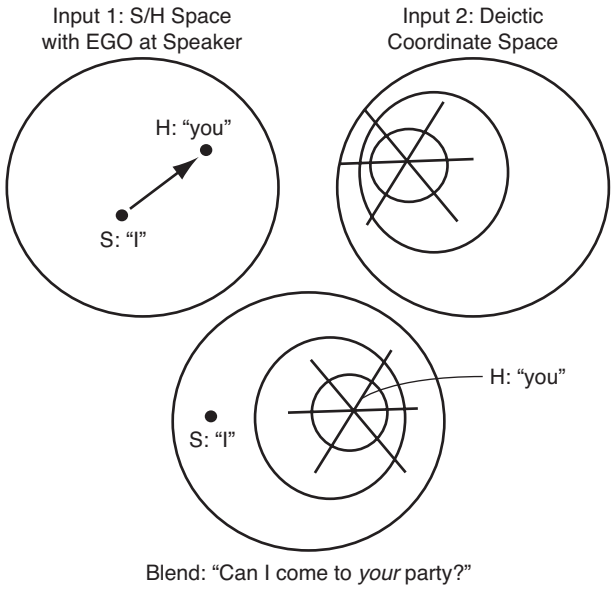

Figure 0.1

centered on Ego (hence on the Speaker), in this blend is mapped onto the Addressee as center.

The crucial point here is that our everyday construal of personal viewpoint is a blend. It is a blend that 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