LANGUAGE AND THE CREATIVE MIND

edited by

MIKE BORKENT, BARBARA DANCYGIER, and JENNIFER HINNELL

CSLI PUBLICATIONS
Center for the Study of Language and Information
Stanford, California

Printed in the United States
17 16 15 14 13 1 2 3 4 5

ISBN 978-1-57586-671-0 (cloth : alk. paper) –
ISBN 978-1-57586-670-3 (pbk. : alk. paper)
ISBN 978-1-57586-672-7 (electronic)

∞ The acid-free paper used in this book meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

CSLI was founded in 1983 by researchers from Stanford University, SRI International, and Xerox PARC to further the research and development of integrated theories of language, information, and computation. CSLI headquarters and CSLI Publications are located on the campus of Stanford University.

CSLI Publications reports new developments in the study of language, information, and computation. Please visit our web site at
http://cslipublications.stanford.edu/
for comments on this and other titles, as well as for changes and corrections by the author and publisher.

Contents

15

Creativity Across Modalities in Viewpoint Construction

EVE SWEETSER

1 Introduction

It is no news to embodied cognitive science researchers that perception and cognition—even imagined perception—are not just embodied but *viewpointed*. We have no God's-eye perception and cognition.[1] But so far, these researchers have dealt primarily with the question of single viewpoints, not with the ways in which multiple viewpoints live in single minds and bodies. I shall here discuss ways in which language and co-speech gesture are creatively used to represent viewpoints—even complex combinations of viewpoints which no single described person is thought of as experiencing, although the narrator of the description, and the reader or listener, get to experience them all. To give one very simple example, imagine that a speaker moves one hand in fast curves through the air, as she says *We were just careening down the mountain road*—and simultaneously manifests on her face the expression of terror that she felt as a passenger in the careening car. Her hand gesture gives more the visual viewpoint that a spectator might have had of the car's overall path, while her face gives the emotional viewpoint of her past frightened passenger self.

[1] See Dancygier and Sweetser (2012) and Bergen (2012) for relevant reviews of some of this research on viewpoint, and Barsalou (2010), Bergen and Chang (2005) and Feldman (2006) for general reviews of embodied cognition and its relation to language.

Language and the Creative Mind.
Mike Borkent, Barbara Dancygier, and Jennifer Hinnell (eds).

As cognitive science now knows well, viewpoint is not just a pervasive feature of language and cognition, we might even call it an *invasive* one: we are not only constantly but *inevitably* aware of other people's viewpoints as well as our own. We have no non-viewpointed perception of the world: our bodies are always asymmetrically constrained in visual and manual access to the world, as well as in motion—the space in front of us is accessible in a way that the one behind us is not. And there's more to it, which has been much less examined than the single viewpoint. As soon as there's another person sharing space with me, my body tracks what they can see or reach, as well as what I can see or reach. This awareness doesn't necessarily mean that I participate actively in the other viewpoints which I am tracking—though I could. For example, I *could* quite vividly imagine your view of the sunset, and move to look over your shoulder to share your vantage point because my previous view wasn't good enough. But without any vivid imagining, I still can't help knowing in what general direction you're able to see. I also can't help what your body posture and face tell me about your emotional states; and even though I may have rather different emotional states of my own, I can't help inhabiting yours too, to some extent. This is why it's exhausting to spend a lot of time with someone who's in severe pain, or in serious emotional depression.

In everyday life experience, what this means is that my body fully represents, naturally and automatically, its own perceptual and emotional and cognitive experiences—and it also both absorbs and impinges on aspects of the experiences of people around me. To do all this, it uses my own neural hardware for embodying experience. The old *out of sight, out of mind* dictum has some truth to it: I will be less strongly affected by others' experiences if they are not physically enacting their experience in my presence.

But what does all this mean for linguistic representation of viewpoint? Or for that matter, for linguistic and gestural representation of viewpoint in oral narrative? The narrator has to use her single text, or her single body and voice, to represent all the aspects of viewpoint which are imagined to be distributed among characters in the narrative. In the imagined situation, as in life, differing viewpoints are distributed across many bodies—each body has viewpointed experience, impinged on by others but distinct. Nobody can simultaneously *be* the car passenger and the spectator watching the car's path from an outside vantage point. But in the representational form, one single text—or one single body—may be required to represent all of those.

Consider an American Sign Language narrative, described by Paul Dudis (2004a,b), which tells the story of an encounter between a hearing librarian and a Deaf library patron who has inadvertently been making noise in a library. The signing narrator's body represents the seated patron, while

a V-hand point downwards towards his face represents the disapproving gaze of the standing librarian. But simultaneously the signer's face initially represents the librarian's disapproving expression, which is presumably located not on the patron's body but on the (unrepresented) body of the librarian at the other end of the gaze. The signer then switches facial expression to represent the discomfited look of the patron, as he becomes conscious of the librarian's disapproving attention. What might be surprising to some is the first combination: it might look to a naïve nonsigning observer as if the signer's body is representing someone with a disapproving facial expression (who is also being looked at), whereas in fact at that point it represents someone *being looked at by* a person with a disapproving facial expression. The signer has only one face—which has to represent the faces, in turn, of both participants in the described exchange, even while his hand continues to represent the visual view of the librarian, and his trunk continues to represent the seated library patron who is being looked at.

2 Embodied Viewpoint in Language

Language itself is as viewpointed as cognition; in this section of the paper, I will resume some of the basic observations about language and viewpoint (Sweetser [2012] gives more detail and references). Speakers cannot use basic linguistic forms such as pronouns and deictics without 'giving away' their choice of viewpoint. *This, that, here, there*, choice of tenses, and so on—all these tell you when and where objects and events are with respect to some Ego, some perceiver and observer. Your *here* is often my *there*, so speakers have to choose whose viewpoint to represent, as in English when someone accepts an invitation by saying *Sure, I'd love to* ***come*** *to your party*—the verb *come* is representing the addressee's (party-giver's) viewpoint. If the speaker were to subsequently turn to a housemate, however, she might say *Let's* ***go*** *to Chris' party on Saturday* and would be most unlikely to say *come* (*cf.* Fillmore 1997[1971]; Sweetser 2012).

Languages vary as to the ways in which some Ego's viewpoint is represented in standard spatial reference. Relative spatial reference systems refer to spatial relations in small, local space in terms of viewpointed body-based spatial relations such as back/front and left/right. So in English I might say *the paper is to the left of my computer on the desk*, even when I'm giving someone directions about how to find the paper, when neither of us is actually present in the office described. In an Absolute directional language (Levinson 1996, 2003) I would need to say instead something like *the paper is to the east of my computer on the desk*, which would describe the same situation, since I face south as I work in my campus office, so the east side of the computer is the left side from the computer-user's perspective. The

difference is that the east/west distinction does not require the bodily position of an imagined viewer or computer user to be comprehensible: a global directional viewpoint is enough. But to understand what would be *the left of the computer*, I have to imagine a computer user in canonical orientation with respect to the computer, and compute that user's left and right.

Languages also vary in how they use spatial deictic forms to represent social rather than physical spatial viewpoint. Hanks (1990) cites Mayan speakers using distal ('that-there') deictic forms to refer to objects which were physically right next to them, but which belonged to (were socially connected to) someone else, and not available for the speaker's use; conversely, the same speakers used proximal 'this-here' forms to refer to more physically distant objects which were the speaker's own belongings or normally accessible for their use. Rubba (1996) reports similar usages by American English speakers in San Diego, where speakers (all interviewed in the same location on the UCSD campus) used distal forms to refer to neighborhoods that were culturally alien to them, and proximal ones to refer to neighborhoods where they felt comfortable. In these cases, linguistic forms associated with physical viewpoint are extended to refer to social viewpoint.

But viewpoint doesn't stop with the set of forms standardly labeled as deictic. Any choice of a linguistic form expresses viewpoint, inasmuch as it expresses some conceptualizer's *construal*, in Langacker's sense (1987, 1991). The same person might correctly be described as *Professor Jones*, or *Joe*, or *my cousin*, or *a guy from Berkeley*, or *Sue's father*, or *Daddy*, depending on who is describing and who is addressed. In choosing any one form, I have made a decision as to how to prompt construal from my particular addressees in that particular situation.

Most crucially, such choices are already intersubjective, in the sense of Verhagen (2005). I am likely to refer to my across-the-hall office neighbor as *Lev* among colleagues and graduate students, as well as in addressing him, but I might call him *Professor Michael* to an undergraduate student. This is presumably because I am adopting the description form which the undergraduate would use. (The undergraduate is not equally free to do likewise—that is, s/he would not as plausibly use Lev Michael's first name in talking about him to me, on the grounds that I would be likely to use it.) Similarly, it would be inappropriate to say *the book* (as opposed to *a book)* unless I think my addressee shares my cognitive ability to readily identify a book being referred to. In using a definite article, I am crucially manifesting my understanding of my addressee's cognitive relationship to the referent. So even when a linguistic expression only manifests one particular viewpoint (such as more or less formality in describing a faculty member, or the choice of *come* or *go* to describe attendance at a party), the choice of that

expression often normally involves taking multiple viewpoints into account, not just the speaker's own most likely individual perspective.

3 Mixed Viewpoint in Language Use

Narratologists—in this respect well ahead of cognitive science research—have long been aware of multiple viewpoint representation in language, Genette (1980) being a major milestone for this aspect of narratology. What is termed Direct Speech and Thought is characterized by representation of what a single person would say or think—often, though not always, this corresponds to expression of one viewpoint. And in particular, we seem to expect a single utterance or sentence to express a coherent viewpoint; obviously the next utterance in the sequence could be a reply from another participant, but in general there seems to be an unspoken expectation that sentence boundaries, or speaker shifts, correspond to possible places for shifts in viewpoint. But of course that's not always so. Ann Banfield's (1982) 'unspeakable' sentences are literary examples of combinations of linguistic forms which can only be interpreted coherently in a context where they express embedded narratorial viewpoint. A sentence such as *Soon she would be telling Daddy all about it, the nasty tattle-tale!* expresses both the imagined narrator's description of the sentence's subject (third person form *she*, and the description *nasty tattle-tale*), as well as the described subject's own description of her parent (*Daddy*). The narrator would presumably not directly refer to this person as *Daddy*, but uses the form to represent the subject's usage and viewpoint. Without the support of a context involving multiple established viewpoints, these combinations of forms don't even make sense; in such a context, they powerfully represent the embedded structure.

I want to issue a caveat here; Direct Speech and Thought can express embedded viewpoints too. A character who is not the addressee's sibling could say *So I suppose you'll tell dear Daddy all about this*, intending *dear Daddy* to be a sarcastic quotation of the addressee's probable reference form for her parent. In that sense, these sentences are not unspeakable: one does not have to be a literary narrator to embed viewpoint, and thus produce an utterance with heterogeneous viewpoint markers. However, third-person narration in literary texts has been a major locus of exploration and creativity for authors trying to express both the narrator's and the characters' viewpoints, and hence a productive focus for research in narratology. Genette's work led the way in this area; Sanders and Redeker (1996) have offered a model of indirect narrative viewpoint in terms of Mental Spaces Theory, while Zunshine (2006) has examined historical development of strategies of viewpoint embedding in English literature.

A particularly thoroughly documented example is narratorial exploitation of the fact that imperfective verb forms evoke immersion in (internal experience of) an ongoing situation, while perfective forms evoke external viewpoint on a situation as a whole (Genette [1980]; Fleischman [1990]). Because this is so, narrators of third-person narratives (famously and innovatively, Flaubert's narratorial voice in *Madame Bovary*, but also of course Virginia Woolf and many more authors since) can combine a past-tense third-person narrative with imperfective verb forms, as ingredients in the Free Indirect Style, in what is now commonly called Free Indirect Speech and Thought (FIST). *Emma sat at the window; the snow fell* does not produce the same impression as an imperfective *Emma was sitting at the window; the snow was falling*. The impression given by the imperfective past is one where the narrator's viewpoint is still present (hence, past third-person references to the character and the character's present) but is 'immersed' in the local imagined experiential viewpoint of the character (hence the imperfective forms).

Similarly, Nikiforidou (2012) examines the 'impossible' and very ordinary combination of past tense with *now* in English literary texts, often used in FIST contexts and perhaps to be seen as a component of FIST.

(1) They now saw, tied to the fence, Ratliff's buckboard and team. (cited by Nikiforidou from Faulkner, *The Hamlet*)

The PAST + *now* construction is not restricted to literary texts: history books and journalistic settings also use it, with the same general effect of maintaining the narrator's distance from a past third-person setting, while still partially empathizing with the viewpoint of participants in that setting. Nikiforidou gives a formal analysis of this as a mental-space blend; it is also discussed in Dancygier (2012, Chapter 7) as part of a broader treatment of formal blends which create viewpoint blends. Similar blends of narratorial and character viewpoint have been analyzed as built up by connective uses in third-person text (Sanders, Sanders and Sweetser 2009, 2012).

We might note here the contrast between FIST usages and the English Narrative Present construction, which has been little studied because it seems to occur primarily in spoken language. Canonical usages in jokes are helpful data, however, as in (2).

(2) A Martian **walks** into Zabar's [a New York deli], and he **looks** around in **there** and **says**, 'I'll have one of those round things with holes in the center.' So they **give** him a bagel and he **takes** a bite of it. And **then** he **says**, 'Gee, this would go great with lox and cream cheese.'

This is apparently the inverse of PAST + *now*; it might be called the PRESENT + *then* construction. *Now* would be inappropriate as a replacement for *then* in this context—perhaps more obviously disruptive than *then* would be in (1). Notice also the use of distal *in there*; it could not readily be replaced by the proximal *in here*, which would express the viewpoint of 'present' participants in the scenario. The Narrative Present construction, like FIST, combines formal elements marking the narrator's viewpoint and the described participants' viewpoint—but chooses different elements to combine. And the resulting overall viewpoint is strikingly different from FIST. Though much further examination is needed to define the difference, the Narrative Present seems in part to negotiate a close, informal relationship between the narrator and the listeners, in a way that FIST does not.

Vandelanotte (2009, 2012) points to another interesting combination of linguistic forms, not previously recognized by narratologists, which he labels Distanced Indirect Speech and Thought. DIST combines pronoun and tense use appropriate to the actual speaker in the speech situation (in Example (3), notice the use of *you* and past tense) with content, descriptions and phrasing attributed to the quoted speaker (it is not the current speaker, but the past quoted speaker who is imagined to have used phrases such as *not going to be a bookkeeper long* or *going to do wonders*). Like Free Indirect Speech and Thought, DIST combines elements of two voices or viewpoints. But unlike the immersive effect of FIST, where the narrator (despite past tense and third person forms) aligns with the character's perspective, DIST produces a different, distancing effect in viewpoint.

(3) **You** was goin' to do wonders, **you** was! **You** wasn't going to be a bookkeeper long – oh no, not **you**. (cited by Vandelanotte [2012] from Elmer Rice, *The Adding Machine*)

Direct Speech equivalent: I'm going to do wonders, I am!...
FIST equivalent: He was going to do wonders, he was! ...

Dancygier (2012) offers a wide-ranging treatment of relationships between linguistic and narrative structures, including in particular the kinds of blending of viewpointed labels (pronouns, address forms), tense markers, and other forms to build multiple viewpoints in relation to each other. Her analysis of viewpoint embedding and compression (Chapter 7) offers blending analyses of several of the constructions discussed here. In other parts of the book she discusses topics such as the mixing of visual viewpoint in narrative—an issue which should have been brought into this paper if there had been space and time. One of her over-arching points, shared with Nikiforidou, Vandelanotte and Sanders, Sanders and Sweetser, is that to a significant extent we can pinpoint the individual contributions of forms to the

composite viewpoints constructed by narratives. In (3), for example, we know that the use of *you* to refer to the present interlocutor is a link to the speaker's real interaction space, while the past tense forms indicate that the quoted speech was in the past, and the 'quoted' expressions relate to past speech of the current addressee. We can even say that the choice of *you* marks a clear decision *not* to identify with the speaker of those past quoted speech acts—to keep those referential systems distinct from the current frame of reference.

Despite these observations about compositionality in viewpoint constructions, linguists are very far from being able to predict, based on the semantics of particular elements such as tense markers or pronouns, the kinds of different overall viewpoint effects produced by particular combinations of those forms. But it is becoming increasingly evident that this is a major locus of creativity in writing. Not only is embedded viewpoint representation done by combining forms expressing the narrator's and the character's (or the speaker's and the described subject's) different viewpoints, but different kinds of embedded viewpoint can be built by choosing different combinations of formal elements.

4 Multiple Viewpoints in One Body: 'Impossible Gestures'

The next question, however, is whether gesture can—like language—simultaneously represent more than one viewpoint. And the answer is clearly yes. Parrill (2009) documents 'dual-viewpoint' gestures, where for example both an observer-viewpoint large-scale path and a character-viewpoint handshape or facial expression are represented simultaneously (see also Stec [2013] for a review of literature on gestural viewpoint). In a way, this is predictable, or at least expectable, since gesture has so many more degrees of simultaneity than speech. Only one sound at once can come out of a vocal tract; but co-speech gesture is more like signed language in this regard. The independence of the various gestural articulators (two hands, trunk posture, head and trunk orientations, facial expression, eye gaze, eyebrow position, etc.) seems even at first glance to offer rich opportunities for the same kinds of simultaneous multiple viewpoint representations which were observed in Dudis' (2004a,b, 2007) ASL examples. In another example, Dudis (2007) cites an ASL story-teller who is describing a fight over a card-game; the signer's body remains a card-player, with his left hand shaped to hold cards, while his right arm reaches up to mime hitting his own face—enacting the arm of another participant in the story who is hitting the card-player. This kind of *partitioning* of the body—which can assign one part of the body to represent parts of one scene-participant, and

other parts to represent a different participant—has been described in detail in Dudis' work. The scene depicted in these cases is distributed over multiple imagined bodies—but enacted by a single body.

In analyses presented at the 2012 ISGS meeting in Lund, Stec and Sweetser discussed examples where videotaped narrators (English speakers, telling stories about autobiographical incidents) simultaneously manifest more than one physical viewpoint in their bodily gestures and postures. And by this we mean 'impossible gestures' on the order of Banfield's unspeakable sentences. As an example, let us take a story where the narrator is telling how when she was a college student, she helped a female friend with a class project by agreeing to go with the friend and apply for a marriage license at a town hall. The reason she was the one doing this was that her birth certificate says 'male' on it; it is accompanied by a stapled-on correction sheet saying 'female.' They wanted to find out how the town hall officials would respond to this. As she tells about the encounter, she enacts the town hall official and holds in her hands an imaginary birth certificate document, looking down at it skeptically (Figure 1).

FIGURE 1

So far, so good, she is enacting only one person's body. But she continues to hold the imagined document in her hands as she raises her head and voices her own past self's responses to the official—clearly at this point, her voice and face are enacting one participant's viewpoint while her hands are still enacting the other participant (Figure 2). This is an impossible gesture, if we were to imagine it as representing a single bodily viewpoint. The town

hall official could not have played the part of the narrator's past self or made her utterances (and was apparently looking down at a document rather than at her interlocutor); the narrator's past self did not have the document in her hands. We thus have dual-viewpoint enactment here, with two character viewpoints enacted, rather like Dudis' ASL library narrative. And the reason appears to be the same obvious one adduced by Dudis: the narrator has only one body, and needs to express more than one body's behaviors.

FIGURE 2

Further, this narrator continues to keep her hands in 'document-holding' position as (in Figure 3) she turns her head and eye-gaze to her real-world interlocutor, who has asked her why she didn't just detach the stapled-on correction sheet before giving the document to the official; she responds that she didn't think of it. She then returns her gaze to the story world, which is 20 degrees or so to the right of the interaction line with her real interlocutor. This is rather a different phenomenon from the two-character gestures. Here, she 'holds onto' a story character's manual pose while enacting verbally (and with facial expression) her own current part in the Real Space conversation with her real-world interlocutor. Once again, this is not a possible single-viewpoint gesture/posture combination: the narrator is not holding a document in her hands, and the official is not turning to talk to the real-world interlocutor. In this case, the dual gestural structure represents narrative embedding, as do the linguistic strategies mentioned above (FIST, DIST, Narrative Present). With gestural embeddings of this kind, the retention of the story-character's viewpoint seems frequently to be a way of visibly holding the floor for return to the narrative, even while granting the temporary floor to an interruption by the story-listener. We have other such examples from other narrators; sometimes the narrator is not retaining a character's pose but simply keeps her hands up, ready for action, in the space currently reserved for the narrative gestures, while she turns away from that space and interacts with the real-world interlocutor.

FIGURE 3

We did note in our English story-tellers that seven of the ten subjects at some point briefly 'enlisted' the real-world interlocutor as a stand-in for a depicted interlocutor in a past recounted conversation. That is, the narrators enacted their Past Selves, and shook fingers at, or expostulated with, their real-world interlocutors directly, turning towards them and making eye contact. This blend would be impossible for ASL story-tellers, who are required to do what English story-tellers seem to do *most* of the time, namely disengage gaze from the interlocutor and turn 20 degrees or so away from the interlocutor to build a new space for the fictional interaction. Since gaze is a (perhaps the) primary cue distinguishing real from fictive interaction for ASL narrators, ASL narrators cannot 'off-load' representation of a fictive interlocutor onto the real interlocutor, while English narrators can and sometimes do.

As a footnote here, visual art would provide us with yet more objects for analysis of viewpoint blends. It is commonplace in painting to use a depicted character's gaze as a cue to guide the viewer's gaze. De Mey's (2006) work on van Eyck has documented this, as well as other aspects of

van Eyck's uniquely sophisticated viewpoint depiction. The artist's Ghent Altarpiece, painted to hang in a particular chapel of Ghent Cathedral, replicates in the painted scene exactly the angle of light which would have been coming through the windows of that chapel. Every reflective surface depicted (every glass or metal object) and every shadow cast, is depicted as if the light came into the painted world from the windows of the chapel. This is most unusual, since paintings normally represent the light of the depicted scene completely without reference to any imagined real-world light setting in which the painting might be displayed. In this case, van Eyck could be seen as creating the impression for the viewer that s/he is in the same physical space as the depicted painted characters (the Virgin Mary and the Angel Gabriel, as well as others), since s/he shares their visual access to light.

This is only a start, and we have much more analysis to do. But it seems that oral storytellers use their bodies in space—in quite creative ways—to represent complex combinations of narrator and character viewpoints. And some of those combinations are quite analogous to the solutions found by ASL narrators, or by literary authors, to the same kinds of problems. Visual artists are finding creative solutions of different kinds.

5 Conclusions

Given how complex the representation of 'single' viewpoint can get, it is fascinating to go one step further and see the creativity involved in representation of multiple viewpoints in language and gesture, and even visual art. As artists and art critics have always known, creativity consists in finding new and elegant ways to use the resources of your chosen medium. And each of the media mentioned above has different resources.

Face-to-face interaction, in spoken or signed language, has all the resources of an inherently viewpointed body in space, to represent viewpoint phenomena. It also, as literary narratologists might not predict, has the viewpoint resources of language and can combine linguistic forms in ways that mark embedding or combining of viewpoints: in the ironic *Yeah, I'm just so annoying, insisting on transparent grading standards*, the speaker attributes the judgment *so annoying* to the imagined quoted speaker, but the *I* and the phrase *transparent grading standards* are clearly her own. We might initially think that the single body is a limitation on representing multiple bodies—or at least, we might think that it would be confusing to addressees to unpack the blends and realize that one fist represents part of an attacker, while the rest of the body represents the person being attacked. But clearly this is no problem for humans interpreting interaction—and why should it be, for people who can understand Free Indirect Style?

Written language lacks the face-to-face affordances—gestural and intonational—and has a long tradition of making up for this lack by various sophisticated narratorial techniques, including blending of linguistic forms from different viewpoints. Some blending 'recipes' are characteristic primarily of spoken language (the Narrative Present) while others (FIST, DIST) are more characteristic of particular genres of written language. Readers and listeners unpack these blends smoothly and comfortably, building the multiple points of view in relation to each other.

Current cognitive science has not fully assimilated the study of literature or gesture, let alone the study of visual art. So there are almost entirely different sets of scholars examining literary texts, oral narrative, co-speech gesture, signed language structure, speech processing in laboratory settings, and visual art. But since all of these phenomena are created and performed by humans, with deeply viewpointed perception and cognition, it seems important to consider them together. Some progress is happening in this area. Dancygier and Sweetser (2012) have brought together studies of written literary narratives, ASL narratives, and co-speech gesture. Turner (2006) brought studies of art together with cognitive science in a new way, while numerous scholars, notably Dancygier (2012), Hogan (2003a, 2003b) and Herman (2004, 2009), have been working to build more cognitively oriented theories of literary narrative. We can hope for more interdisciplinary connections in the future. But the full range of human creativity—and the ways in which creativity is similar or different across modalities—is waiting to be investigated.

References

Banfield, Ann. 1982. *Unspeakable Sentences: Narration and Representation in the Language of Fiction*. Boston: Routledge and Kegan Paul.

Barsalou, Lawrence. 2010. Grounded cognition: past, present and future. *Topics in Cognitive Science* 2, 716–24.

Bergen, Benjamin. 2012. *Louder Than Words: The New Science of How the Mind Makes Meaning*. New York: Basic Books.

Bergen, Benjamin and Nancy Chang. 2005. Embodied construction grammar in simulation-based language understanding. In Jan-Ola Östman and Mirjam Fried (eds). *Construction Grammars: Cognitive Grounding and Theoretical Extensions*. 147–90. Amsterdam: John Benjamins.

Casasanto, Daniel. 2009. Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*. 138(3), 351–367.

Casasanto, Daniel and K. Jasmin. 2010. Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. PLoS ONE 5(7), e11805.

Dancygier, Barbara. 2012. *The Language of Stories: A Cognitive Approach*. Cambridge: Cambridge University Press.

Dancygier, Barbara and Eve Sweetser (eds). 2012. *Viewpoint in Language: A Multimodal Perspective*. Cambridge: Cambridge University Press.

De Mey, Marc. 2006. Mastering ambiguity. In Mark Turner (ed. *The Artful Mind: Cognitive Science and the Riddle of Human Creativity*. 271–304. New York: Oxford University Press.

Dudis, Paul. 2004a. *Depiction of Events in ASL: Conceptual Integration of Temporal Components*. Ph.D. dissertation, University of California, Berkeley.

_____. 2004b. Body partitioning and real-space blends. *Cognitive Linguistics* 15:2, 223–38.

_____. 2007. Types of depiction in ASL. Ms., Gallaudet University.

Fauconnier, Gilles and Mark Turner. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Feldman, Jerome. 2006. *From Molecule to Metaphor: A Neural Theory of Language*. Cambridge, MA: MIT Press.

Fillmore, Charles J. 1997 [1971]. *Lectures on Deixis*. Stanford CA: Center for the Study of Language and Information. (Originally published 1971 by the Indiana University Linguistics Club as *The Santa Cruz Lectures on Deixis*.)

Fleischman, Suzanne. 1990. *Tense and Narrativity: From Medieval Performance to Modern Fiction*. Austin: University of Texas Press.

Genette, Gérard. 1980. *Narrative Discourse: An essay in Method*. Ithaca: Cornell University Press.

Hanks, William F. 1990. *Referential Practice: Language and Lived Space in a Maya Community*. Chicago: University of Chicago Press.

Haviland, John. 2000. Pointing, gesture spaces and mental maps. In David McNeill (ed.). *Language and Gesture*. 13–46. Cambridge: Cambridge University Press.

Herman, David. 2009. *Basic Elements of Narrative*. Malden MA and Oxford: Wiley-Blackwell.

_____. 2004. *Story Logic: Problems and Possibilities of Narrative*. Lincoln: University of Nebraska Press.

Hogan, Patrick Colm. 2003a. *Cognitive Science, Literature and the Arts: A Guide for Humanists*. New York: Routledge.

_____. 2003b. *The Mind and its Stories: Narrative Universals and Human Emotion*. Cambridge: Cambridge University Press.

Kita, Sotaro (ed.) *Pointing: Where Language, Culture and Cognition Meet*. Mahwah NJ: Lawrence Erlbaum.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*. 2 vols. Stanford: Stanford University Press, vol. I.

_____. 1991. *Foundations of Cognitive Grammar*. 2 vols. Stanford: Stanford University Press, vol. II.

Levinson, Stephen C. 1996. Relativity in spatial conception and description. In John J. Gumperz and Stephen C. Levinson (eds). *Rethinking Linguistic Relativity*. 177–202. Cambridge: Cambridge University Press.

_____. 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*. New York: Cambridge University Press.

McNeill, David. 1992. *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.

McNeill, David and Susan Duncan. 2000. Growth points in Thinking-for-Speaking. *Language and Gesture*. 141–161. Cambridge: Cambridge University Press.

Nikiforidou, Kiki. 2010. Viewpoint and construction grammar: the case of past + now. *Language and Literature* 19:3, 265–84.

_____. 2012. The constructional underpinnings of viewpoint blends: *The Past + now* in language and literature. In Dancygier and Sweetser (eds). *Viewpoint in Language: A Multimodal Perspective*. 177–197. Cambridge: Cambridge University Press.

Özyürek, Asli. 2002. Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language* 46:4, 688–704.

Parrill, Fey. 2009. Dual viewpoint gestures. *Gesture 9*(3), 271–289.

_____. 2010. Viewpoint in speech-gesture integration: Linguistics structure, discourse structure, and event structure. *Language and Cognitive Processes*, 25(5), 650–668.

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

Rubba, Johanna. 1996. Alternate grounds in the interpretation of deictic expressions. In Gilles Fauconnier and Eve Sweetser (eds). *Spaces, Worlds and Grammar*. 227–61. Chicago: University of Chicago Press.

Sanders, José and Gisela Redeker. 1996. Perspective and representation in speech and thought in narrative discourse. In Gilles Fauconnier and Eve Sweetser (eds). *Spaces, Worlds and Grammar*. 290–317. Chicago: University of Chicago Press.

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*. 19–59. Berlin: Mouton de Gruyter.

_____. 2012. Responsible subjects and discourse causality: How mental spaces and perspective help identifying subjectivity in Dutch backward causal connectives. *Journal of Pragmatics* 44, 191–213.

Stec, Kashmiri. 2013. Meaningful shifts: A review of viewpoint markers in gesture and sign language. *Gesture* 12(3), 327–360.

Stec, Kashmiri and Eve Sweetser. 2012. Significant breaks: Space and viewpoint transitions in gesture. Paper presented at the 2012 International Gesture Studies Conference, in Lund, Sweden.

Sweetser, Eve. 2012. Introduction: Viewpoint and perspective in language and gesture, from the Ground down. In Dancygier and Sweetser (eds). *Viewpoint in Language: A Multimodal Perspective*. 1–22. Cambridge: Cambridge University Press.

Sweetser, Eve and Marisa Sizemore. 2008. Personal and interpersonal gesture space: functional contrasts in language and gesture. In Andrea Tyler, Yiyoung Kim, and Mari Takada (eds). *Language in the Context of Use: Cognitive and Discourse Approaches to Language and Language Learning*. 25–51. Berlin: Mouton de Gruyter.

Turner, Mark (ed.) 2006. *The Artful Mind: Cognitive Science and the Riddle of Human Creativity*. New York: Oxford University Press.

Vandelanotte, Lieven. 2009. *Speech and Thought Representation in English*. Berlin/New York: Mouton de Gruyter.

_____. 2012. 'Wait till you got started': How to submerge another's discourse in your own. In Dancygier and Sweetser (eds). *Viewpoint in Language: A Multimodal Perspective*. 198–218. Cambridge: Cambridge University Press.

Verhagen, Arie. 2005. *Constructions of Intersubjectivity: Discourse, Syntax and Cognition*. Oxford: Oxford University Press.

Zunshine, Lisa. 2006. *Why We Read Fiction: Theory of Mind and the Novel*. Columbus, OH: Ohio State University Press.

16

Multimodal Construction Grammar

FRANCIS STEEN AND MARK TURNER

1 Introduction: Massive Multimodal Data

When people meet, they invariably communicate in multiple modalities: the eyes, gestures, and tones of voice merge with the perceived affordances of the surroundings into an integrated and partially shared experience. Multimodal communication predates and contextualizes language, and extends into a series of social, artistic, and technological innovations, from dance to cave paintings, from theater to cinema, from town criers to television news. Over the past hundred thousand years, our species has grown from a few roving bands to some seven billion individuals, linked by densely tangled electronic networks. That adds up to a lot of multimodal communication, a rich treasure trove of data comparable in complexity to the information available to astrophysicists, zoologists, and geneticists. Yet the datasets scholars rely on for deciphering the hidden orders of human communication remain overwhelmingly textual. The relatively few and small corpora we have of multimodal communication often come from specialized circumstances, such as interviews conducted by experimenters in whiteroom lab settings. In this article, we explore ways to broaden the foundations of our study of human communication.

Methodologically, we need to check our hypotheses against data. In the face of small, biased, and narrow archives of data, language scientists have often relied on personal introspection to choose between hypotheses. Yet, it is well-recognized by now in cognitive linguistics that although personal intuition can be a source of hypotheses, it has weaknesses as a test of hy-

Language and the Creative Mind.
Mike Borkent, Barbara Dancygier, and Jennifer Hinnell (eds).