Wilbur, R. B. (1987). *American Sign Language: Linguistic and applied dimensions.* Boston, MA: College-Hill Press.
Wilbur, R. B. (1999). Stress in ASL: Empirical Evidence and Linguistic Issues. *Language and Speech, 42*(2–3), 229–50.
Wilcox, P. P. (2000). *Metaphor in American Sign Language.* Washington, DC: Gallaudet University Press.
Wilcox, P. P. (2004). A cognitive key: Metonymic and metaphorical mappings in ASL. *Cognitive Linguistics 15*(2), 197–222.
Wilcox, P. & Wilcox, S. (1995). The gestural expression of modality in American Sign Language. In J. Bybee & S. Fleischman (Eds.), *Modality in grammar and discourse.* Amsterdam: John Benjamins Publishing Company.
Wilcox, S. (1992). *The phonetics of fingerspelling.* Amsterdam: John Benjamins.
Wilcox, S. (2002). The iconic mapping of space and time in signed languages. In L. Albertazzi (Ed.), *Unfolding perceptual continua* (pp. 255–81). Amsterdam: John Benjamins.
Wilcox, S., Wilcox, P., & Jarque, M. J. (2004). Mappings in conceptual space: Metonymy, metaphor, and iconicity in two signed languages. *Jezikoslovlje, 4*(1), 139–56.
Wilcox, S. (2004a). Conceptual spaces and embodied actions: Cognitive iconicity and signed languages. *Cognitive Linguistics, 15*(2), 119–147.
Wilcox, S. (2004b). Gesture and language: Cross-linguistic and historical data from signed languages. *Gesture, 4*(1), 43–75.
F. Wilkinson, E. (2006, January). Exploiting location and movement to construct mataphors: A cross-linguistic analysis on American and Italian Sign Languages. Poster presented at the workshop on Cross-linguistic Sign Language Research, Max Planck Institute, Nijmegen.
Woodward, J. (1979). Quelques aspects sociolinguistiques des langues des signes americaine et française [Sociolinguistic aspects of French and American Sign Language]. *Langages, 13,* 78–91.
Woodward, J. & De Santis, S. (1977). Negative incorporation in French and American Sign Language. *Language in Society, 6*(3), 379–388.
Wulf, A., Dudis, P., Bayley, R., & Lucas, C. (2002). Variable Subject Presence in ASL Narratives. *Sign Language Studies, 3*(1), 54–76.
Zeshan, U. (2003). 'Classificatory' constructions in Indo-Pakistani Sign Language: Grammaticalization and lexicalization processes. In K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 113–141). Mahwah, NJ: Lawrence Erlbaum.
Zeshan, U. (2004a). Negative constructions in sign languages. *Linguistic Typology, 8*(1), 1–58.
Zeshan, U. (2004b). Interrogative constructions in signed languages: Crosslinguistic perspectives. *Language, 80*(1), 7–39.

Looking at space to study mental spaces

Co-speech gesture as a crucial data source in cognitive linguistics

Eve Sweetser

1. Introduction

The vast majority of linguists, psychologists and cognitive scientists do not look at gestural data. Most such departments offer no training in the transcription or analysis of gesture – and this includes departments that pride themselves on being oriented towards the rich data of embodied cognition and language. Despite this lack of institutionalized curricular niches, research on gesture has made major strides in the last couple of decades. And as a result, widely dispersed individual researchers on language and cognition are finding that the study of normal bimodal language production *as* bimodal, rather than as written transcripts of the auditory modality, gives us far richer data for the study of cognition. The formation of the International Society for Gesture Studies in 2002 and the publication of the journal *Gesture* (starting in 2001) are both results of interaction among this diverse community, and ongoing reinforcements of such interaction.

Real spoken language production in context is always a multimodal process, performed by a gesturing body embedded in a physical setting. Many cognitive linguists now take seriously the embodiment of human thought[1] and language – the idea that our conceptual and linguistic structures could not be as they are if they were not based in human bodily experience of the world. Nonetheless, as a cognitive linguist, I have undergone successive processes of recognizing the degree to which linguistic methodologies fall short of investigating language as embodied. If we linguists are not still focused on written language, we are still focused on the output of the vocal track in spoken language communication, neglecting the visual-gestural track that is co-performed with it.

And yet the visual-gestural medium offers information almost impossible to squeeze into the oral-auditory medium. Vocal production has highly sequential structure (one

1. See the entire cognitive linguistic corpus of work, but notably Lakoff and Johnson 1999, Lakoff and Núñez 2000, Fauconnier and Turner 2002, and the works leading up to them (Fauconnier 1998; Fauconnier & Turner 1996; Turner & Fauconnier 1995). More recent work on mirror neurons supports this embodied viewpoint as well; cf. for example Gallese and Lakoff 2005.

we are well aware that we are producing language (except in cases such as sleep-talking) and we are often aware at intervals of particular form choices ("oh no, I said *men* again, and I should have said *people*"). Speakers notice and may correct the errors of first- and second-language learners: observers do say things like "don't use that word"; "that's *ran*, not *runned*"; "that should be *Do you go*? and not *Go you?*" Admittedly such correction is not common in caregivers of children acquiring language, and second-language learners outside the classroom may also not get much overt correction – and when it does occur it is often ineffectual. But it is easy for an observer to notice such linguistic errors – all speakers are aware of whether a past tense form is correct or not in their dialect. This is much less the case with gesture. Parents in some cultures encourage children to gesture less, or smaller – but they do not notice errors in the form or order of gestures, or suggest alternate better forms. For instance, no researcher has reported a parent saying to a child, "no no, that gesture should have been timed with the subject of the sentence rather than with the verb." And this is unsurprising, given that the parents are themselves unaware of the generalization about gesture timing.

So, as McNeill (1992) and others (see particularly Alibali 1994; Alibali & Goldin-Meadow 1993a, b; Goldin-Meadow 1999, 2003; McNeill 2000) have cogently argued, gesture has immense potential as a source of information about cognitive processes during linguistic production and comprehension. Goldin-Meadow (2003) gives strong evidence that when gesture and speech conflict, two cognitive models are represented. This can be used as a metric of when a child is in the process of catching onto a new concept. For example, Goldin-Meadow observed children responding to Piagetian conservation tasks; at a certain point, when a child has watched the water poured from a tall thin glass to a short fat glass, the child may still *say* that there is less water now – but may gesture a width-wise gesture which suggests awareness of the differences in circumference of the containers. Compared to other children who give the same verbal answer, such a child is more likely to be ready to go beyond a simple metric of height for quantity, and learn the more complex conservation relationship.

Cognitive linguists will naturally want to know whether everyday speakers and gesturers, as well as scientists, can access the information provided by gesture alongside that provided by speech. The answer is pretty clearly *yes*. It does not seem that the cognitive utility of gesture is limited to one side of the communicative exchange, although gesture is differently useful to the performer and the viewer. Speakers gesture even when not in the physical presence of a viewing interlocutor (Bavelas et al. 1992; de Ruiter 1995). Telephone speakers seem to make fewer interactive, discourse-regulating gestures than face-to-face speakers do, although they still make highly routinized interactional gestures such as head-nods. However, they still make iconic gestures about the content of their speech (including spatial gestures while giving directions by phone). All this would seem to indicate that gesture is not there only for the hearer. Krauss et al. (2000) and others have documented the impact of gesture on lexical access; speakers who cannot make manual gestures (e.g., are requested to keep their hands grasping the top of a chair while speaking) have more difficulty with lexical access, particularly when the content of their speech is spatial. This is what a neuroscientist would expect. When two neural routines are closely correlated in performance, activating one of them will help activation of the other (as with humming a tune to help you remember the words of the song). McNeill and Duncan (2000) have posited a more complex model of *growth points*, wherein gesture is a *material carrier* for the dynamic development of thinking for speaking.

Gesture also has communicative impact on hearers/viewers. McNeill (1992) records an experiment where subjects viewed a videotaped narrative that included a gesture-speech mismatch (the videotaped speaker gestured stair-climbing with his fingers while saying *he climbed up the drainpipe*). A significant number of viewers, in retelling the story, said that the character climbed up stairs rather than up a drainpipe. Kendon (1995) presents examples where information is clearly available from the gestural track, complementing the spoken information. Goldin-Meadow (2003) presents an impressive body of evidence that hearers make use of speakers' gestural information. Özyurek (2000) presents interesting evidence of "addressee design" in gesture, and similar patterns are suggested by Sweetser and Sizemore (in press). Perhaps the most impressive evidence of all for the communicative efficacy of gesture, however, are examples like McNeill's (1992) Snow White experiment, wherein subjects were required to tell the story of Snow White without language. The storytellers' communicative partners "took up" the gestural content with remarkable speed, and the speedy reduction and conventionalization of patterns representing ongoing characters give clear evidence of what happens to gesture when it bears the full communicative load unassisted by another track. Goldin-Meadow (1993, 2003) similarly chronicles the use of gesture by deaf children not exposed to a signed language.

Signed languages emerge when gesture is the central vehicle of communication among a community, over a long enough period of time.[3] We have recently been given glimpses of this emergence process in the work of Kegl and colleagues (Kegl & Iwata 1989; Kegl, Senghas, & Coppola 1999) on the development of Nicaraguan Sign Language within a community which began around a Deaf school. Signed languages are thus the endpoint of a communicative development which starts when the visual-gestural modality is regularly given the full communicative load, as spoken languages are the endpoints of a developmental sequence wherein vocalizations ended up bearing the primary semantic load but remained paired with visual-gestural expression. There is, however, good evidence that signed languages also have gesture, performed in the same visual medium as the language performance (Liddell 2003); the same kinds of evidence also suggest that we should be thinking of some components of spoken language prosody as gestural (in *he gave a looooong talk*, the iconic vowel lengthening is gestural). The similarities and differences between spoken and signed language gestural systems are only beginning to be investigated.

I shall also not primarily discuss fully conventional gestures here – what many researchers have called *emblems* and Kendon calls *quotable gestures*. The affirmative head-nod, and the thumb-index circle (ASL "F" handshape) which English speakers use to mean "just right, perfect" are examples; so are various obscene gestures. Basic differences between these gestures and flexible co-speech gesture include (1) that the head nod or the "just right" gesture does not need language at all, but can substitute for words if necessary,

3. See Armstrong, Stokoe and Wilcox (1995) for arguments that such signed languages may have been crucial to the evolution of modern spoken languages.

and (2) that the head nod and the "just right" hand gesture are fully conventional, always formed in the same way and subject to judgments of mis-performance if done wrong. Quotable gestures are complex and well worthy of discussion in their own right; readers are referred to Kendon 1990, 2004; Calbris 1990.

I also do not here deeply discuss all the ways in which bodily location, and in particular the relative bodily locations of interlocutors, are relevant and meaningful to the interaction, or the ways in which gesture indexes all kinds of meanings in discourse. Some of these (particularly Haviland 1993, 2000) will come up in my final discussion of viewpoint. Clark (1996) and Goodwin (2000) are excellent examples of work on these subjects; Engle (2000) and Smith (2003) are fascinating studies of the way different modalities of gesture are used, and of the interaction of gesture with surrounding objects.

And I do not here follow the ways in which gesture has been used as a window on speech processing; this is studied in the work of researchers such as Krauss et al. (2000) and Butterworth and Beattie 1978, 1989. Nor do I adequately cover the ways in which gesture is involved in social interaction; the work of Bavelas (Bavelas et al. 1992 and elsewhere) is of particular interest here. Clearly cognitive linguistics should be interested in all these issues.

This chapter will present some ways that looking at gesture can be useful, in my view, specifically to cognitive linguists. It cannot possibly discuss all the ways. Gesture is an immensely varied phenomenon, and far less studied to date than language. Susan Duncan once said to me that when researchers make conflicting claims about gesture – e.g., that it expresses content, and that it negotiates interaction; or that it facilitates lexical access, and that it is used independently of words – she felt this was usually only evidence that they were looking at different aspects of this pervasive and complex range of human behaviors.

3. Transcription

Every analyst will have different needs, in analyzing and transcribing gesture, just as in analyzing and transcribing language. One linguist may need spectrograms and close phonetic transcription, another may find that standard English orthography is fine for her purposes; another (a discourse analyst perhaps) may not want phonetic transcription but need to measure pauses. Similarly, in studying gesture, there are cases where it is crucial to measure time in milliseconds, as in the research (cited above) leading to our current understanding of the co-timing patterns of gesture and speech, and the evidence for neural co-performance of the two routines. For my own work, it has been sufficient to identify co-timing of a gesture stroke with language at the level of the syllable or word, since my goal has been to examine the relationship between the meanings of the two tracks.

I will therefore not be advocating and presenting in detail any particular gesture transcription system. My own system has been strongly influenced by the McNeill group's work, which uses a grid around the speaker's body for spatial location description, and a modified version of the American Sign Language finger-spelling system for description of handshapes. But most of what you will see in this paper will be relatively abbreviated transcription forms, focusing on the correlations between form and meaning rather than on highly detailed formal description of gesture. Reading McNeill (1992), Kendon (2004) and other works on gesture will provide better resources for someone starting gesture transcription than I can provide here.

It would be yet another paper (and not one for which I am particularly competent), to assess the technical needs of a wide range of possible gesture research projects, including the variation between the needs of researchers with different computer systems. For all gesture analysis and transcription, researchers will need some system of video playback that permits simultaneous slowing down of the visual and sound tracks. Presuming that your data is digital, therefore, Final Cut Pro and other more sophisticated programs will work; QuickTime is not sufficient, although a convenient medium for exchanging video files and for making freezes for illustrations. A transcription program allowing co-indexing of the transcript with the video is also important as the data base grows; needs will vary depending on the project.

I close this section with a final word about the difficulties of data collection and presentation. It should be noted that gesture transcription is much more time-consuming than speech transcription. Like speech, gesture can be transcribed more or less finely; but the standard approximation is that transcribers spend at least ten times as much time on the gesture track as on the corresponding speech track. If you have done speech transcription, you know that this is a daunting figure; and indeed it is correspondingly difficult to build up large corpora of transcribed gesture data. Further, although one can (and linguists often do) present only a transcription of spoken data, it is essential to present photo freezes, if not whole video clips, in presenting analysis of gesture; no transcription system is simultaneously detailed enough and transparent enough to substitute for visual images in a scholarly article. However, since speakers are far more identifiable from visual images than from written transcripts of their words – and yet more from video clips – subjects' privacy is correspondingly more protected, and much analyzed gesture data would be barred from inclusion in any large publicly accessible database.

So linguists thought it was hard to analyze real language data, and they were right; but the complexity and difficulty of the endeavor – and therefore the time commitment – are even greater when you analyze multimodal communication. And of course, you will still be missing a lot, no matter what you choose to focus on – you won't be able to work on every aspect of multimodal communication simultaneously, and researchers with different emphases will see things you do not (just as linguistic researchers notice different things about the same language phenomena). But gesture analysts agree that we can't now go back into the box; communication is multimodal, and we need to work on it as it is.

4. Gesture, iconicity and levels of abstraction

Gesture can, as mentioned above, be literally iconic for a flexible array of physical and spatial entities. Most notably, a body can be iconic for a body: a gesturer who is presenting a represented character's words and simultaneously pounds on the table, for example, can be interpreted as meaning that the quoted character pounded on something. As noticed by Mandel (1977) and Taub (2001) for ASL, iconic mappings between the body and represented content also include:

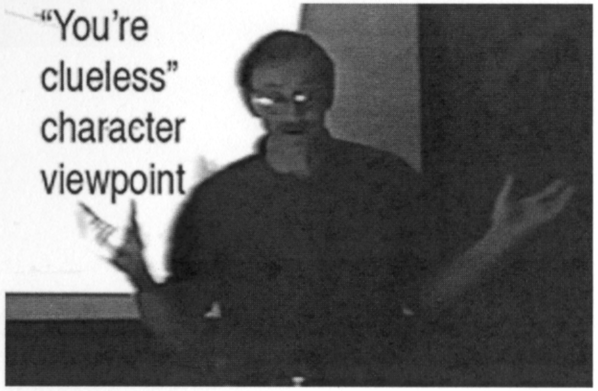

Figure 8.

will not refer to the Real Space of the surrounding city of Berkeley, or to the Real Space north-south and east-west parameters, but rather to imagined surroundings of Palo Alto, assuming an imagined imputed orientation of the speaker with respect to those surroundings. This means that the indexical function of gesture is much richer and more flexible than it appears to be on the surface – although Goodwin (2000) gives excellent examples of how flexible the connections can be between actual objects pointed to, and the intended meaning.

Similarly, deictic temporal gesture necessarily involves use of the body as establishing a spatial origo which is metaphorically interpreted as meaning NOW. Núñez and Sweetser (2006) have shown that in cultures whose languages follow the semi-universal metaphoric models PAST IS BEHIND EGO, FUTURE IS IN FRONT OF EGO, gesture does so too – English speakers clearly gesture forwards about the future and backwards about the past. Aymara, a native American language spoken in the Andean highlands of Chile, Bolivia and Peru, is a genuine exception to this quasi-universal of linguistic metaphor: Aymara linguistically treats PAST as IN FRONT OF EGO, while FUTURE IS BEHIND EGO. As Núñez and Sweetser show, Aymara speakers also gesture forwards when talking about *last year* or *last week*, and backwards when talking about *next year* or *next week*. The absolute universal here seems to be that the front of the body is the NOW origo. (In signed languages also, the NOW location seems to be not the body itself, but the front surface or the space immediately in front of the body). Lateral motion representing time (a speaker can say *it went on for years* and gesture from one side to the other) lacks a specific built-in NOW, and is not deictic in its function.

In Figure 9 below, the speaker from the Cognitive Studies colloquium is finishing his lecture. He says: *[it's time] to..uh for me to stop here and interact with you at this point and and and – take some – take questions and… so I think I'll do that here. Thank you.* As he says *I think I'll do that here* and *thank you*, he twice points downwards to the ground immediately in front of him. Both the word *here* and the point to a location "here" (right in front of Ego) refer metaphorically to NOW, rather than literally to a spatial location.

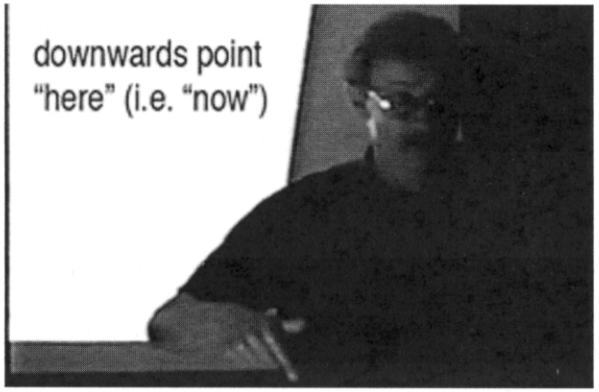

Figure 9.

These points are extremely common in English speakers' gestures – many a speaker saying *right now* can be seen pointing downwards for emphasis, just as she would if she were saying *right here*.

It would be particularly interesting to have data on the temporal representation gestures used by speakers of languages with *absolute* spatial systems (Levinson 1997, 2003). These speakers do not use their bodies as origos for everyday spatial representation as pervasively as users of relative spatial language systems. An English speaker might say *the ball is in front of the tree* (Hill 1982; Levinson 2003); the hearer would need to know that in English, trees are construed as facing the speaker (Ego-Opposed) and that therefore the ball is between the tree and the speaker. An absolute spatial system would dictate something more like *the ball is west of the tree*, needing no calculation from the speaker's location. It might therefore generate less egocentric spatialization of time, as well, at the metaphoric level.

7. Conclusions

My work on iconic and metaphoric uses of gesture (Sweetser 1998; Parrill & Sweetser 2004; Núñez & Sweetser 2006), conversation-regulation uses of gesture (Sizemore & Sweetser in press), and general mental space mappings in gesture (Parrill & Sweetser 2004) have convinced me that mental spaces theory is how to approach gesture analysis. Liddell (1990, 1995, 1998, 2003) and Taub (2001) have shown that the same is true for analysis of signed languages.

On the other hand (to use an appropriate metaphor), gesture analysis is crucial input to mental spaces theory and to cognitive linguistics in general. It is hard to exaggerate the degree to which gesture pervasively embodies the source domain spaces, in representing abstract concepts metaphorically. It is also hard to imagine that the performance of these physical routines in space is as unrelated to the source domain concepts as some sound at once) and its direct iconic capacity is limited (of course) to sounds and temporal or sequential structure: it can represent a cat's vocalization with the word *meow*, or ongoing length of an activity with repetition such *as she talked on and on and on*. But it cannot directly iconically represent spatial relations, motion, paths, shape, and size,[2] all of which are easily represented in manual and bodily gesture (Taub 2001). Gesture is also more temporally flexible than speech, since the hands (for example) are relatively independent articulators which can perform separate routines simultaneously; add motion of trunk, head, feet, mouth, and eyebrows, and you begin to get an idea of this flexibility. So we would expect that examining gesture might reveal much about embodied conceptual structure which is less obvious in spoken language.

And indeed, looking at spatial cognition and language, McNeill and co-workers have shown that gestural depictions of motion in space, as well as linguistic structures, are systematically different for typologically different languages; researchers on language and space are now making use of gestural data to reach new conclusions about crosscultural aspects of cognitive structure. Cienki (1998) has shown that systematic metaphoric gestural structures may occur separately from linguistic metaphoric usage, as well as accompanying such usage: in such a case, a cognitive semanticist or mental space theorist might want to re-assess the "meaning" of the accompanying language, in the light of the co-produced gestures. Sizemore and Sweetser (in press) show that gesture beyond the speaker's personal gesture space, into the interlocutor's space, happens specifically when the speaker is regulating the conversational exchange; discourse analysts might therefore wish to examine gestural structure alongside linguistic discourse markers. Gesture is as varied in its functions as language, perhaps even more so. But all of these directions, and more, will be of use to cognitive linguists and cognitive scientists.

This chapter has several goals. First, I shall briefly lay out some of the well-established findings which underlie modern gesture research: for example, the evidence that gesture and speech are co-produced as a single neural package, and that a specific pattern of neural packaging is shared by a given language community. This will bring up some basic work in the field – such as that of McNeill (1992, 2000) and Kendon (1990, 2000, 2004). I shall evaluate ways in which such analyses and such data are useful to cognitive linguists.

In the later sections of the chapter, I shall continue to examine the ways in which these tools for gestural analysis can help cognitive linguists to further understand the language and cognition which are going on as the observed gestures are performed – and vice versa. I shall argue in particular that the tools of Mental Spaces theory are extremely productive in this endeavor: gestures should be understood as grounded blends of Real Space (see Liddell 1990, 1995, 1998, 2000, 2003) with other mental spaces. Such an analysis allows us to see new generalizations at multiple levels of gesture analysis – and also gives us a single framework within which to model multi-modal communication.

Specific examples (presented as photographic frame-freezes from the relevant videos) will be analyzed, showing how gesture operates at abstract levels, expressing metaphoric meanings, discourse-structural meanings, and other abstract conceptual content (ongoing work by the Berkeley Gesture Project will be cited here), as well as iconically representing actual physical spaces. The *systematic* nature of these gestural structures will be laid out, and the ways in which they bear on claims about embodied cognition and mental spaces theory. The goal will be to help readers with differing research goals see the ways in which they can make use of gestural data – and to give them realistic expectations as to the kinds of work involved and the kinds of results which may emerge.

2. Indirect iconicity is possible as well; for example, sound symbolism systems very frequently involve high front vowels (performed with a felt routine of *small* mouth aperture) representing diminutivity or small size, in opposition to low vowels (with larger aperture) which represent large size (Sapir 1929; Ultan 1978; and see also the papers in Ohala et al. 1994). The vowel sound may be seen as metonymically connected with the size of the aperture involved, which in turn is iconic for the size of the object referred to in the linguistic form.

2. Gesture as a back door to cognition

Psycholinguists have become adept at getting below the surface of our conscious language production. Aspects of processing inaccessible to speakers are revealed in minute but readily measurable differences in linguistic processing time, for example, as well as by various scanning techniques revealing neural activity during linguistic processing. But speech-accompanying gesture offers another way of getting at the less conscious aspects of the cognitive processes involved in language. By gesture, I here primarily mean what Kendon calls *gesticulation* – movement of hands and body in flexible visual-gestural patterns which accompany the auditory track of spoken language.

Speakers of all languages gesture as they produce linguistic discourse, constantly. An interesting fact for cognitive science is that co-linguistic gesture is neurally co-processed with language, in tight language-specific neural routines. Stutterers "stutter" in gesture and resume gestural fluency when they resume speech fluency (Mayberry, Jacques, & DeDe 1998). Chinese gestural *strokes* (the main motion phase of a gesture) are systematically co-timed with the object of a sentence (McNeill & Duncan 2000), while English gestural strokes are co-timed with the verb (McNeill 1992). Gestural systems, and the coordination of the gesture-speech package, are acquired over the relatively long time span of language acquisition. Spanish has a different relationship between gestural and linguistic expression of manner and path of motion than English does, and Spanish-speaking children do not seem to have a fully adult system for gesturing about manner of motion until about age 12 (McNeill & Duncan 2000). The gesture-language complex that children learn is learned *as* a complex; coordination between the two tracks is language-specific, and the gestural track is complex and requires lengthy learning alongside the learning of the language system.

Another interesting fact for cognitive scientists and linguists is that the gesture track of this co-production routine is less monitored than the linguistic track. Of course, we cannot consciously choose every linguistic form we produce; we would never express things effectively if much of linguistic production were not automated and processed below the consciously accessible level. (This becomes painfully obvious to an adult learning a second language, where she lacks prefabricated unconscious production routines.) But

1. use of some part of the body to represent some object – a finger representing a leg, for example, as a hand gestures "walking" with fingers.
2. representation of an object by representing grasping or other interaction with it – e.g., a hand shaped to grasp a glass, to represent a glass.
3. moving the articulator to trace the shape of an object ("drawing" in the air or on some surface).
4. using motion of an articulator along a path to represent motion of an object along a path.
5. partial representation of some physical routine or activity – for example two hands grasping an invisible steering wheel to represent driving a car (LeBaron & Streeck 2000; this is also close to American Sign Language forms for CAR and DRIVE).

A major problem for iconic interpretation of a representation is always knowing how much to map from one domain to another. For example, in representing a person walking with two fingers pointing downwards, the gesturer and any interpreters have to know that only part of the gesturer's hand is to be mapped –the thumb and two smaller fingers need to be left out of the iconic representation. Dudis (2004a, b) has addressed this problem for ASL iconicity, arguing that signed language involves partitioned *zones* of representation. A parallel solution for gesture seems necessary.

A great deal of work has been done on gesture representing spatial concepts and accompanying spatial linguistic content. There is already solid evidence that cognitive differences correlate with differences in linguistic representation of spatial relations and motion. From Slobin's (1987, 1996, 2000) work on Thinking for Speaking and Levinson's (1997, 2003) work on spatial language and thought, we know that mental images, memory of spatial arrays and other cognitive factors seem to differ depending on whether the speaker's language has an absolute or relative spatial system, and on how the language's grammar combines expression of motion, path and manner. McNeill (1992), Kita (2000, 2003), Özyurek (2000), Nobe (2000), Haviland (1993, 2000) and many others have specifically studied spatial gesture, including the crosslinguistic differences between such gestures.

Equally interesting, however, are gestural representations of abstract concepts. As detailed in Taub (2001) for ASL signs with abstract meanings, and by Parrill and Sweetser (2004) for gesture, such representations are necessarily both iconic and metaphoric. For example, a speaker who shows a clenched fist as she says *rock-solid argument* is iconically representing a solid physical object; but (like the word *rock-solid*) this metaphorically represents an abstract argument which is perhaps likely to withstand counterargument.

In the following examples, transcribed by Fey Parrill and Eve Sweetser and analyzed by the Berkeley Gesture Project, you can see speakers representing abstract structure in terms of location and motion in physical gesture space. The first one is from a public lecture given by Mark Johnson at a bookstore in Los Angeles. Johnson is saying that a traditional understanding of meaning involves semantic concepts with fixed definitions which map neatly onto states of affairs in the world.

...have fixed definitions
and they map onto the world...um...
and that knowledge consists in...framing a set of concepts
that neatly map onto states of affairs in the world
whether those states of affairs have to do with morality or politics or um...or um...
quantum physics or whatever.

As he speaks, he gestures. On *fixed definitions* and *concepts* his hands (palms facing each other, on his upper right) delimit a space for concepts (Figure 1). As he says *map onto the world*, and once again as he says *states of affairs*, his hands move to delimit a space for real-world states, to the lower left of his gesture space (Figure 2). What he is doing is setting up a gestural space wherein the space to his upper right refers to concepts, and the space on his lower left refers to world-states. It is normal for a speaker contrasting two entities to use the two sides of the gesture space in this way; and it is also normal for gesturers to use the vertical dimension as he does (ABSTRACT IS UP, CONCRETE IS DOWN) in contrasting abstract concepts with concrete states of affairs.

Language and gesture show systematic use of the same metaphoric systems (Cienki 1998; Sweetser 1998; Núñez & Sweetser [2006]). Sometimes they are used simultaneously and in parallel – as in the case of the speaker who said *rock-solid evidence* and also made a solid fist gesturally – but gesture can also be used independently of language, as of course language can use metaphor independent of gesture (written language is full of metaphor, after all). Cienki (1998) documents a speaker who consistently gestures downwards for bad grades, up for good grades, and downwards for immoral behavior, upwards for moral behavior. Although the speaker (an American college student) is surely familiar with English linguistic metaphoric uses such as *high moral values*, *low-down dirty behavior*, and *high* or low *grades* (cf. Lakoff & Johnson 1980:200 he does not use such wording in this particular section of his discourse; only his gestures show clearly that he is conceptualizing morality and grades in terms of vertical scales.

Figure 1.

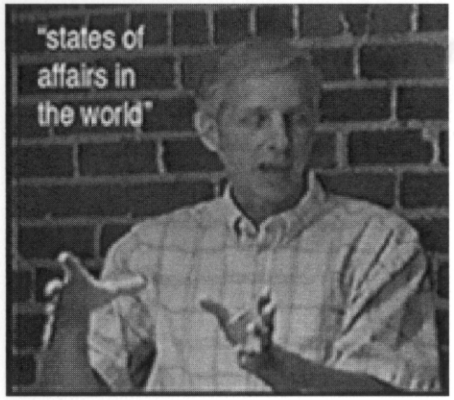

Figure 2.

Divisions of space in gesture similarly metaphorically represent *mental spaces* or areas of content; spatial loci become associated with subjects (McNeill 1992), as in signed languages (Liddell 2003). Relative position of loci in gestural space can systematically represent relationships between the content connected with those loci. Just as in linguistic usages such as *far apart* to mean "different" or "socially unconnected" and *really close* to mean "similar" or "socially connected", gesturers bring their hands or fingers into proximity to show similarity and relationship, and make a gesture of moving their hands apart when discussing difference.

The temporal structure of gesture is also exploited iconically to represent temporal structure of represented activities and situations. Dudis (2004a) has given an analysis of the multiple ways in which time is mapped between sign language performance and meaning. Similar structure regularly occurs in gesture. Our next example comes from a public lecture given to the UC-Berkeley Institute of Cognitive Studies colloquium. The speaker's topic is short-term planning memory. At one point he focuses on the "mental book-keeping" involved in spatiomotor tasks: a person dialing a phone number needs to know at what stage in the sequence she is, at any given point, even if she is repeating the same action as an earlier stage.

> ...and a key feature of spatiomotor tasks
> is the book-keeping you need to keep track of where you are
> You dial a phone number, 38537,
> you have to know...
> when you're doing the 3, you have to know where you are,
> is it the first three or the second three?

In this sequence, the speaker consistently gestures outwards and to the right, to show the development of a spatiomotor process; earlier stages are locations closer to his body (and reached earlier, since he's gesturing outwards) and later stages are locations farther from his body, as in the illustrations below. Figure 3 is the starting position (walking around, he

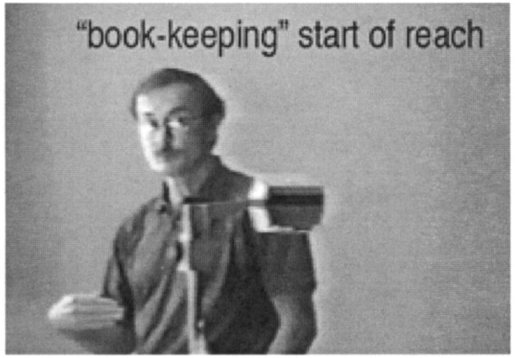

Figure 3.

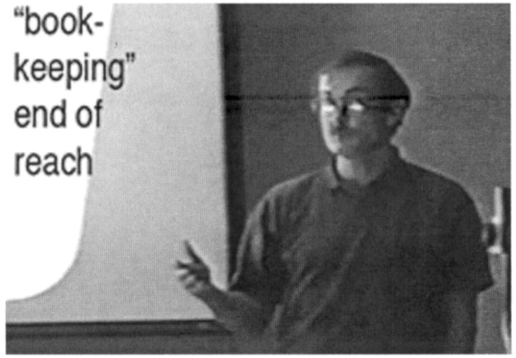

Figure 4.

performs beats in about this position from *book-keeping* to the start of the phone number dialing), and Figure 4 is the final position (on *seven*, the last digit of the phone number). He subsequently retravels this motion as he goes back and forth along the same path from *first three* to *second three*, showing that he has built up a spatial image which is still present for exploitation. Note that although dialing a telephone is a spatiomotor routine, and of course this outwards gesture is a spatiomotor routine too, the gesture does not represent the spatiomotor work of telephone dialing itself; it represents only the temporal sequential structure of the telephone dialing task. It is thus an abstract metaphoric gesture, even though it is about a physical routine.

In general, my observation is that gestural representation of temporally evolving processes, abstract or concrete, necessarily involves motion either *across* (from one side to the other) or *outwards from* the speaker's body (in this case, both). You cannot represent dialing a phone number (or getting a Ph.D., or solving a problem) by motion *towards* the body from a point farther away. There is only one exception, which is that if the process mentioned consists of physical motion towards a deictic center (e.g., if the speaker is saying *he came closer and closer to me*), then it is entirely appropriate to represent this process of approach iconically by moving the hand towards the body.[4] The body seems to constitute a spatial deictic center, which is an origo for the representation of spatial structure – and hence, metaphorically, for the representation of temporal structure as well.

For those interested in embodied cognition, it is important to understand restrictions on the mapping between Real Space (the speaker's perceived gesture space) and mental spaces such as the temporal structure of a process. What is it about our construal of embodied spatial motion – and its correlation with time – which allows our bodies to be the default temporal origo representation? Presumably it is related to the fact that when our muscles are relaxed, our hands fall back to the body, so that extension from the body is in some sense the primary gesture, necessarily preceding retraction.

5. Gesture and levels of content

A single linguistic form is often meaningful not merely in multiple senses, but at multiple levels of the communicative interaction. The labels I use for such levels are *content, epistemic, speech act*, and *metalinguistic*.[5] Gestures have the same multi-level property. Kendon (1995, 2004) documents the use of a single Neapolitan gesture, two hands sweeping in a barrier motion, with discourse meanings as well as content meanings. In one case it accompanies a speaker talking about *stopping* doing something, in another case it seems to indicate that the accompanying linguistic contradiction should cut off (*stop*) the interlocutor's train of reasoning, and in a third case it wordlessly tells someone not to interrupt while the gesturer is engaged in talking to someone else (*stop* the attempt at speech interaction). McNeill (1992) has a discussion of American discourse interactive gestures such as the hand held out, palm upwards, to mean that the speaker feels she has made her point; Sweetser (1998) discusses the American palm-outwards hand used to fend off interruption in conversation.

A crucial point about these gestures is that – being about abstract meanings – they are necessarily metaphoric. They could be used "literally" too; that is, they could in principle iconically represent a physical barrier, or showing a physical object to an interlocutor, or putting out a hand to protect oneself from a blow or a projectile. But from this iconic structure, they are metaphorically mapped onto thought and discourse interaction. Sweetser (1998) points out that the metaphors forming the basis for these abstract interactional

4. Taub (2001) brings up similar instances in ASL grammar, where actual physical motion direction is salient enough to trump other factors in motivating sign structure.

5. The first three of these four concepts were developed in a tradition of scholarship initiated by Halliday and Hasan and are now best known in the work of Traugott (1982, 1989) and Hopper and Traugott (1993) on grammaticalization; my specific terms and definitions were developed in Sweetser (1990). My use of a metalinguistic level stems from the work of Horn (1985, 1989) on negation and Dancygier (1998) on conditionals; it is further developed in Dancygier and Sweetser (2005).

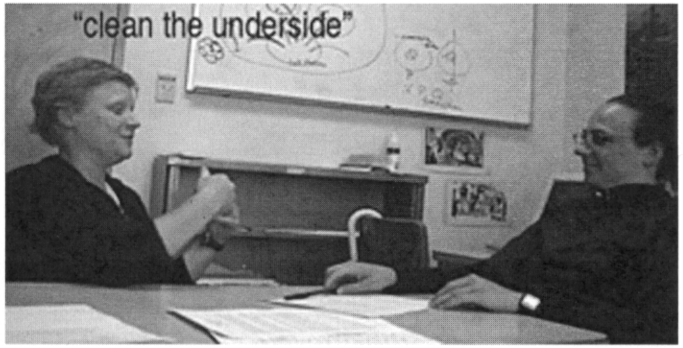

Figure 5.

gestures are also basic linguistic metaphors in English: for example, as detailed in Sweetser (1987, 1992), English regularly treats abstract communication as OBJECT EXCHANGE. Some metaphoric informational "objects", such as hostile questions or unwelcome interruptions, can be seen as projectiles, which in turn can be *rebutted* or *fended off* – iconically represented by the palm-outwards hand in gesture. Objects can be "shown" as well as exchanged; and indeed, a very common linguistic accompaniment to the palm-up gesture is *You see*?, or *See my point*?. The palm-up gesture thus appears to be iconic for showing an object – an *invisible surrogate* object, using Liddell's (1995, 1998, 2000, 2003) and Dudis' (2004b) terminology.

In the following examples, recorded by Marisa Sizemore for the Berkeley Gesture Project, two friends are talking together. S1, the blond woman, is describing a past unsatisfactory roommate to S2, the dark-haired woman. In Figure 5, as she says *clean the underside*, S1 makes a gesture of dishwashing by moving one hand (the "washing" hand) in front of the other (representing the plate being washed). This is a clear example of content gesture, iconic representation of one physical action by another.

S1:	She would never clean the undersides of the dis[hes]
S2:	[OK],
	so she'd wash the dishes
	but you had a problem with [how she did it.]
S1:	[And I…and I…]

S2 then interrupts to tease S1 about her fussiness; after all, this roommate did *wash* the dishes. As she interrupts, she reaches far forward into S1's normal *personal gesture space* (Figure 6b) and pats the table on *wash* (Figure 6c); then she leans yet further forward (Figure 6d) as she says *but you had a problem*; she finally leans back again (Figure 6e), satisfied that she has made her point and ready to listen to more of S1's story.

This gesture does not represent anything about the content of the utterance; it shows nothing about cleaning dishes, having problems, or manner of dishwashing. It is, however, a typical interactional regulation gesture. Sweetser and Sizemore (in press) show repeated forwards reaches into an interlocutor's gesture space at moments when the reacher wants

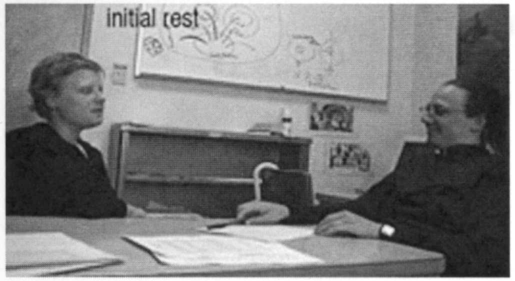

Figure 6a.

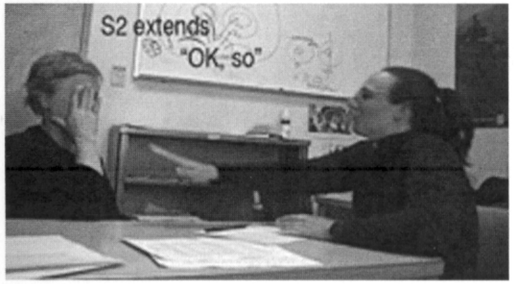

Figure 6b.

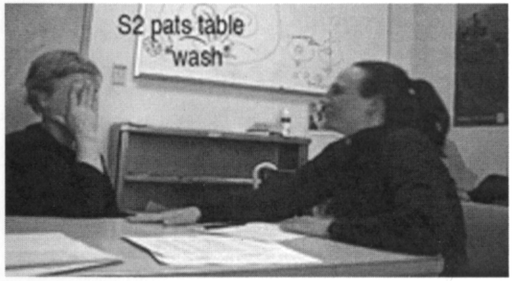

Figure 6c.

to (re)claim the floor or start up a new topic. Such interactional gestures are typically performed on the line between the speaker and the addressee, between their personal gesture spaces (in what Sweetser and Sizemore call *interpersonal space*). Other kinds of gestures can be performed in *extrapersonal* space, the space outside both personal and interpersonal spaces; for example, at another point in S1 and S2's discourse, S1 reaches far up to her left in an iconic gesture of reaching up for a plate in a cupboard (and finding it to be

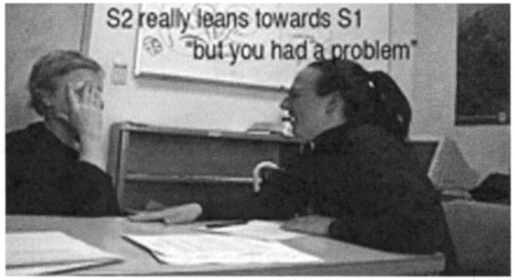

Figure 6d.

Figure 6e.

greasy). It would be unusual, at least, to build the cupboard in the interpersonal space; and it would be impossible to perform floor-claiming in the extrapersonal space.

Gesture thus not only functions at the same multiple levels of communicative meaning as speech, but spatial divisions in the gesturer's Real Space are relevant to functional levels in communication. Physical spaces once more stand for different Mental Spaces. This is of course important for cognitive linguists; and the finding points again to the need to look at all functions of language, not only at content communication.

6. Gesture, viewpoint and deixis

Gesture is performed by a body located and oriented in space, and can exploit the body's location and orientation in multiple ways. Spoken language also does this, of course; in normal face-to-face interaction, references such as *here* can depend on the hearer's knowledge of the speaker's location to index it successfully (a more problematic assumption when *here* is said instead over a cell phone connection). But speech is evidently lacking in indexing precision compared to gesture; *this one* or *that one* may need a supplementary gesture of pointing, but the point does not need a supplementary linguistic *this* or *that*.

Pointing is apparently a human universal, although with many cultural variations in form and use: Kita (2003) gathers recent work on such indexical gestures.

A body necessarily has physical viewpoint (Sweetser 2001, 2002, 2003); human bodies share structure which ensures that they can see forwards but not backwards, can access objects in front of them better than ones in back of them, can move forwards better than backwards, and of course are experiencing a gravitic environment in which we are normally able to stand on our feet rather than our heads (Clark 1973; Fillmore 1997[1971]). As Sweetser (2001, 2002, 2003) has pointed out, other subjective viewpoint phenomena correlate with physical perspective and access to a scene. There are *primary scenes* (Grady 1997, 1998; Johnson 1996, 1999) wherein we co-experience attentional focus, interactional affordances, and all the social and emotional correlates (satisfaction at successful physical interaction).

So one rather simple mapping is to have a viewpointed body represent another viewpointed body; and this is common in both gesture and signed language structure. It is predictably more conventionalized in ASL; grammarians use the term *role shift* to describe a grammatically distinguished stretch of signed discourse during which the signer ceases eye-contact with the addressee and turns her body slightly to one side, while enacting the linguistic performance and other actions of a "quoted" character. Similar shifts in gesture have been referred to as *character viewpoint* (McNeill 1992).

Global viewpoint has also been observed in gesture, as in signed languages; for example, when a hand represents a moving person, the other hand can represent another person moving to meet the first person. In this case, the gesturer's body trunk, and even the arms, are not representing anything, but are outside the mappings; only the moving hands represent the moving entities.

Rapid switches between these two kinds of viewpoint are apparently quite easy for speakers. The same Institute of Cognitive Studies lecture, cited in Section 4 above, contains such a case. The speaker is talking about reinforcement learning, and is using the metaphor of someone wandering around a maze, "running into" rewards and/or previously encountered locations. Like a rat running through a new maze, the wanderer has no initial direction, but is "wandering around idly". On the first utterance of *wander around idly* (Figures 7a–b) he rotates his right hand in front of his place-holder left hand, showing undirected motion (see Parrill & Sweetser 2004). In this case, only the hand and its motion represent the wanderer in the maze – and indeed the representation is rather abstract, since the general aimless motion is not the kind of motion involved in running around a maze.

Transcript:
you can think of reinforcement learning as dynamic programming done badly…
m'kay… he way reinforcement <Xworking it doesX>
you just wander around idly … m'kay
…and then when you bump into the reward you know what to do from the penultimate state.
You're not gonna have a policy,
you're clueless,
but the penultimate state, you know what to do…

and you know how to <Xdiscount itX> right?
Then you're wandering around idly a-again
you might bump into that penultimate state
the state before that knows how to do…
and you know how to <Xdiscount itX>, right?

However, when he says *you're clueless* and later *you're wandering around idly* again, he waves his arms in mime of a confused, lost person; his whole body represents this person. (The *clueless* image is below as Figure 8, because it is of more reproducible quality than the similar *wandering around idly* image.)

Even when the body is not involved in character viewpoint, it stands as an origo for spatial and temporal representation, as also mentioned above. However, to navigate from an origo one must know what space is being represented. As Haviland (2000) has cogently pointed out, a speaker in (for example) Berkeley may be giving directions to someone about a location in Palo Alto – and in such a case, the pointing and directional gestures

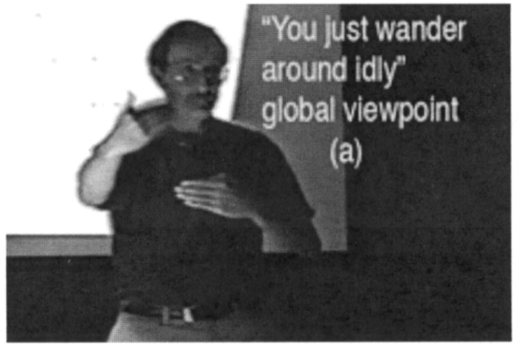

Figure 7a.

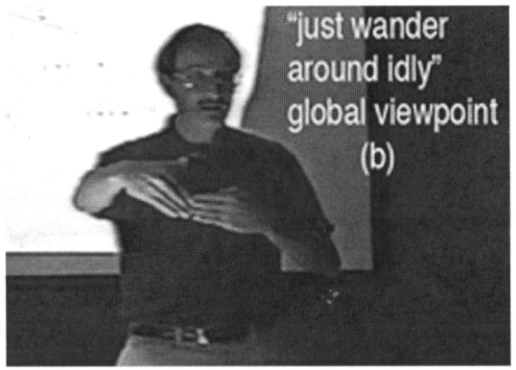

Figure 7b.

have claimed is the case with metaphoric language. Making a fist *is* an instance of solidity, and gesturing forwards *is* an instance of forward motion, unlike the linguistic forms *solid evidence* and *years ahead of us*. Discourse space structure is equally striking: the claim that two compared domains (such as short-term planning memory and chess) are two different mental spaces is enacted in gestural production by assigning them two different physical spaces as loci.

Gesture varies crosslinguistically and crossculturally, as has been shown by McNeill 1992, 2005, McNeill and Duncan 2000, Duncan (2001), Kendon (1990, 1995, 2000, 2004), Haviland (1993, 2000), Núñez and Sweetser (2006) and many others. It is a potentially crucial source of evidence for differences in linguistic cognition between communities. Fascinatingly, the limitations and constraints on gesture may show us cognitive universals – potentially even ones which we have not yet noticed in language. It was already well established (Moore 2000) that in language, deictic metaphors for time are universally back-front ego-referent metaphors; this claim is so far well supported by gestural data (Núñez & Sweetser [2006]). But would we have known that aspectual gestures about abstract processes therefore necessarily anchor themselves in an origo at the body – and how universal is this generalization, so clearly manifested in English? What exactly does it mean about cognition?

Gesture, it seems, offers potential evidence relating to linguistically expressed conceptual structure in a huge range of areas of interest to a linguist – concrete linguistic meaning, abstract linguistic meanings, discourse regulation and interaction, indexicality and viewpoint, and doubtless much more. All we have to do is reach out and observe the evidence at hand.

References

Alibali, M. W. (1994). *Processes of cognitive change revealed in gesture and speech.* Ph.D. dissertation, University of Chicago.

Alibali, M. W. & Goldin-Meadow, S. (1993a). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. *Cognitive Psychology, 25*, 468–523.

Alibali, M. W. & Goldin-Meadow, S. (1993b). Modeling learning using evidence from speech and gesture. *Proceedings of the 15th annual conference of the Cognitive Science Society*, 203–208.

Armstrong, D. F., Stokoe, W. C., & Wilcox, S. E. (1995). *Gesture and the nature of language.* Cambridge, England: Cambridge University Press.

Bavelas, J., Chovil, N., Lowrie D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*, 469–489.

Butterworth, B. & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In R. N. Campbell & P. T. Smith (Eds.), *Recent advances in the psychology of language: Formal and experimental approaches* (pp. 347–360). New York: Plenum.

Butterworth, B. & Hadar, U. (1989). Gesture, speech and computational stages: A reply to McNeill. *Psychological Review, 96*, 168–174.

Calbris, G. (1990). *The semiotics of French gesture.* Bloomington: Indiana University Press.

Cienki, A. (1998a). Metaphoric gestures and some of their relations to verbal metaphoric expressions. In J.-P. Koenig (Ed.), *Discourse and Cognition* (pp. 189–204). Stanford CA: CSLI Publications.

Clark, H. H. (1996). *Using language.* Cambridge: Cambridge University Press.

Clark, H. H. (1973). Space, Time, Semantics and the Child. In T. Moore (Ed.), *Cognitive Development and the Acquisition of Language* (pp. 27–63). New York: Academic Press.

Dancygier, B. (1998). *Conditionals and prediction.* Cambridge: Cambridge University Press.

Dancygier, B. & Sweetser, E. (2005). *Mental spaces in grammar: Conditional constructions.* Cambridge: Cambridge University Press.

Duncan, S. (2001). Perspectives on the co-expressivity of speech and co-speech gestures in three languages. In *Proceedings of the 27th Annual Meeting of the Berkeley Linguistics Society.* Berkeley CA: The Berkeley Linguistics Society.

Dudis, P. G. (2004a). *Depiction of Events in ASL: Conceptual Integration of Temporal Components.* Ph.D. dissertation, Department of Linguistics, University of California, Berkeley.

Dudis, P. G. (2004b). Body partitioning and real-space blends. *Cognitive Linguistics, 15*(2), 223–238.

Engle, R. A. (2000). *Towards a theory of multimodal communication: Combining speech, gestures, diagrams and demonstrations in instructional explanations.* Ph.D. dissertation, School of Education, Stanford University.

Fauconnier, G. & Turner, M. (1996). Blending as a central process of grammar. In A. Goldberg (Ed.), *Conceptual Structure, Discourse and Language* (pp. 113–130). Stanford: CSLI.

Fauconnier, G. & Turner, M. (2002). *The Way We Think.* Basic Books.

Fillmore, C. J. (1997 [1971]). *Lectures on Deixis.* Stanford CA: CSLI Publications. (originally published by Indiana University Linguistics Club).

Furuyama, N. (2000). Gestural interaction between the instructor and the learner in origami instruction. In D. McNeill (Ed.), Language and gesture (pp. 99–117). Cambridge: Cambridge University Press.

Gallese, V. & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology, 22*(3/4), 455–479.

Goldin-Meadow, S. (1993). When does gesture become language? A study of gesture used as a primary communication system by deaf children of hearing parents. In K. R. Gibson & T. Ingold (Eds.), *Tools, Language and Cognition in Human Evolution* (pp. 63–85). New York, NY: Cambridge University Press.

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Science, 3*, 419–429.

Goldin-Meadow, S. (2003). *Hearing gesture.* Cambridge MA: Harvard University Press.

Goodwin, C. (2000). Gesture, aphasia and interaction. In McNeill (Ed.) 2000, 84–98.

Grady, J. (1997). *Foundations of meaning: Primary metaphors and primary scenes.* Ph.D. dissertation, Department of Linguistics, University of California, Berkeley.

Grady, J. (1998). The "Conduit Metaphor" revisited: a reassessment of metaphors for communication. In J.-P. Koenig (Ed.), *Discourse and Cognition: Bridging the Gap* (pp. 205–217). Stanford, CA: CSLI.

Haviland, J. B. (1993). Anchoring, iconicity and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology, 3*, 3–45.

Haviland, J. B. (2000). Pointing, gesture spaces and mental maps. In D. McNeill (Ed.), *Language and gesture* (pp. 13–46). Cambridge: Cambridge University Press.

Hill, C. (1982). Up/down, front/back, left/right: A contrastive study of Hausa and English. In J. Weissenborn & W. Klein (Eds.), Here *and* there: *Crosslinguistic studies on deixis and demonstration* (pp. 11–42). Amsterdam: John Benjamins.

Hopper, P. & Traugott, E. Closs (1993). *Grammaticalization.* Cambridge: Cambridge University Press.

Horn, L. (1985). Metalinguistic negation and pragmatic ambiguity. *Language 61*, 121–174.

Horn, L. (1989). *A natural history of negation.* Chicago: University of Chicago Press.

Hutchins, E. (1995). *Cognition in the wild.* Cambridge MA: MIT Press.

Johnson, Ch. (1996). Learnability in the acquisition of multiple senses: SOURCE reconsidered. *Proceedings of the 22nd Annual Meeting of the Berkeley Linguistics Society*, 469–480.

Johnson, Ch. (1999). *Constructional grounding: The role of interpretational overlap in lexical and constructional acquisition.* Ph.D. dissertation, Department of Linguistics, University of California, Berkeley.

Kegl, J. & Iwata, G. (1989). Lenguaje de Signos Nicaragüense: A pidgin sheds light on the 'creole' ASL. In R. Carlson et al. (Eds.), *Proceedings of the Fourth Meeting of the Pacific Linguistics Conference* (pp. 266–294). Eugene OR: Dept. of Linguistics, University of Oregon.
Kegl, J., Senghas, A., & Coppola, M. (1999). Creation through contact: Sign language emergence and sign language change in Nicaragua. In M. De Graff (Ed.), *Language creation and language change: Creolization, diachrony and development* (pp. 179–237). Cambridge, England: Cambridge University Press.
Kendon, A. (1990). Gesticulation, quotable gestures and signs. In M. Moerman & M. Nomura (Eds.), *Culture embodied* (pp. 53–77). Osaka: National Museum of Ethnology.
Kendon, A. (1995). Gestures as illocutionary and discourse markers in Southern Italian conversation. *Journal of Pragmatics, 23*, 247–279.
Kendon, A. (2000). Language and gesture: Unity or duality? In D. McNeill (Ed.), *Language and gesture* (pp. 47–63). Cambridge: Cambridge University Press.
Kendon, A. (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.
Kita, S. (Ed.). (2003). *Pointing: Where language, culture and cognition meet.* Mahwah NJ: Lawrence Erlbaum Associates.
Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162–185). Cambridge: Cambridge University Press.
Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge: Cambridge University Press.
Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By.* Chicago: University of Chicago Press.
Lakoff, G. & Johnson, M (1999). *Philosophy in the Flesh: The embodied mind and its challenge to Western thought.* New York: Basic Books.
Lakoff, G. & Núñez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being.* New York: Basic Books.
LeBaron, C. & Streeck, J. (2000). Gestures, knowledge and the world. In D. McNeill (Ed.), *Language and gesture* (pp. 118–138). Cambridge: Cambridge University Press.
Levinson, S. C. (1997). Language and cognition: The cognitive consequences of spatial description in Guugu Yimithirr. *Journal of Linguistic Anthropology, 7*(1), 98–131.
Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity.* Cambridge: Cambridge University Press.
Liddell, S. (1990). Four functions of a locus: Reexamining the structure of space in ASL. In C. Lucas (Ed.), *Sign Language Research: Theoretical Issues* (pp. 176–198). Washington, D.C.: Gallaudet University Press.
Liddell, S. (1995). Real, surrogate and token space: grammatical consequences in ASL. In K. Emmorey & J. Reilly (Eds.), *Language, Gesture and Space* (pp. 19–41). Hillsdale NJ: Lawrence Erlbaum Associates.
Liddell, S. (1998). Grounded blends, gestures, and conceptual shifts. *Cognitive Linguistics, 9*(3), 283–314.
Liddell, S. (2000). Blended spaces and deixis in sign language discourse. In D. McNeill (Ed.), *Language and gesture* (pp. 331–357). Cambridge: Cambridge University Press.
Liddell, S. (2003). *Grammar, gesture and meaning in American Sign Language.* Cambridge: Cambridge University Press.
Mandel, M. (1977). Iconic devices in American Sign Language. In L. Friedman (Ed.), *On the other hand* (pp. 55–107). London: Academic Press.
McNeill, D. (1992). *Hand and mind.* Chicago: University of Chicago Press.
McNeill, D. (2005). *Gesture and Though.* Chicago: University of Chicago Press.
McNeill, D. (Ed.). (2000). *Language and Gesture.* Cambridge, England: Cambridge University Press.
McNeill, D. & Duncan, S. (2000). Growth points in thinking for speaking. In McNeill (Ed.), *Language and Gesture* (pp. 141–161). Cambridge, England: Cambridge University Press.

Mayberry, R. I., Jaques, J., & DeDe, G. (1998). What stuttering reveals about the development of the gesture-speech relationship. In J. Iverson & S. Goldin-Meadow (Eds.), *Nature and Function of Gesture in Children's Communication* (pp. 89–100). San Francisco: Jossey-Bass.
Moore, K. E. (2000). *Spatial experience and temporal metaphors in Wolof: Point of view, conceptual mapping, and linguistic practice.* Ph.D. dissertation, Department of Linguistics, University of California, Berkeley.
Nobe, S. (2000). Where do most spontaneous representational gestures actually occur with respect to speech? In D. McNeill (Ed.), *Language and gesture* (pp. 186–198). Cambridge: Cambridge University Press.
Núñez, R. & Sweetser, E. (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*, 1–49.
Ohala, J. J., Hinton, L., & Nichols, J. (Eds.). (1994). *Sound Symbolism.* Cambridge: Cambridge University Press.
Özyurek, A. (2000). The influence of addressee location on spatial language and representational gestures of direction. In D. McNeill (Ed.), *Language and gesture* (pp. 64–83). Cambridge: Cambridge University Press.
Parrill, F. & Sweetser, E. (2004). What we mean by meaning: Conceptual integration in gesture analysis and transcription. *Gesture, 4*(2), 197–219.
de Ruiter, J. P. (1995). Why do people gesture at the telephone? In M. Bieman & M. Woutersen (Eds.), *Proceedings of the Center for Language Studies Opening Academic Year 1995–96* (pp. 49–56). Nijmegen: Center for Language Studies.
Sapir, E. (1929). A study in experimental symbolism. *Journal of Experimental Psychology, 12*, 225–239.
Sizemore, M. & Sweetser E. (in press). Personal and interpersonal gesture speces: Functional contrasts in language. In A. Tyler, Y. Kim, & M. Takada (Eds.), *Language in the Context of Use: Cognitive and Discourse Approaches to Language and Language Learning.* Berlin.
Slobin, D. I. (2000). Verbalized events: A dynamic approach to linguistic relativity and determinism. In S. Niemeier & R. Dirven (Eds.), *Evidence for Linguistic Relativity* (pp. 107–138). Amsterdam: John Benjamins.
Slobin, D. I. (1996). From 'thought and language' to 'thinking for speaking.' In J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, England: Cambridge University Press.
Slobin, D. I. (1987). Thinking for speaking. In *Proceedings of the 13th Annual Meeting of the Berkeley Linguistics Society* (pp. 435–445). Berkeley, CA: The Berkeley Linguistics Society.
Smith, N. J. (2003). Gesture and beyond. Honors thesis, Program in Cognitive Science, UC-Berkeley.
Sweetser, E. (1987). Metaphorical Models of Thought and Speech: a comparison of historical directions and metaphorical mappings in the two domains. In *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society* (pp. 446–459). Berkeley, CA: The Berkeley Linguistics Society.
Sweetser, E. (1992). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure.* Cambridge: Cambridge University Press.
Sweetser, E. (1992). English Metaphors for Language: Motivations, conventions, and creativity. *Poetics Today, 13*(4), 705–724.
Sweetser, E. (1998). Regular metaphoricity in gesture: Bodily-based models of speech interaction. In *Actes du 16e Congres International des Linguistes* (CD ROM). Oxford: Elsevier
Sweetser, E. (1998). Regular metaphoricity in gesture: Bodily-based models of speech interaction. In CD-ROM *Proceedings of the 16th International Congress of Linguists*, held in Paris in July 1997.
Sweetser, E. (2001). Deixis, viewpoint and modality. Plenary talk at the International Cognitive Linguistics Conference, July 16–20, 2001 at the University of California at Santa Barbara.

2nd proofs

Sweetser, E. (2002). Conceptual blending and viewpoint. Plenary talk at the international workshop "The Way We Think," at Odense University, Demark, August 19–23, 2002.
Sweetser, E. (2003). Literal and metaphorical viewpoint in gesture. Paper presented at the 8th International Cognitive Linguistics Conference, at the University of La Rioja, Spain, July 2003.
Taub, S. F. (2001). *Language from the Body: Iconicity and metaphor in American Sign Language.* Cambridge, England: Cambridge University Press.
Traugott, E. Closs (1982). From propositional to textual and expressive meanings: Some semantic-pragmatic aspects of grammaticalization. In W. P. Lehmann & Y. Malkiel (Eds.), *Perspectives on historical linguistics* (pp. 245–71). Amsterdam: John Benjamins.
Traugott, E. Closs (1989). On the rise of espistemic meanings in English: An example of subjectification in semantic change. *Language, 65*, 31–55.
Turner, M. & Fauconnier, G. (1995). Conceptual integration and formal expression. *Journal of Metaphor and Symbolic Activity, 10*(3), 183–204.
Ultan, R. (1978). Size-sound symbolism. In J. H. Greenberg, C. R. Ferguson, & E. A. Moravcsik (Eds.), *Universals of Human Language, Vol. 2.* (pp. 527–568). Stanford: Stanford University Press.

Methodology for multimodality

One way of working with speech and gesture data

Irene Mittelberg

1. Introduction

Multimodality, besides entailing at least two modes of expression, also engages at least two modes of interpretation. For instance, when language and gesture coincide, the body of the speaker is as much involved in expressing thoughts, attitudes, or emotions, as are the linguistic articulators. Simultaneously, while listening, the addressee/observer draws on both her sense of audition and vision to make sense of what the interlocutor is conveying. Compared to the various other types of multimodal communication that have come to predominate contemporary media landscapes, this kind of fusion of two modes of expression, speech and gesture, is perhaps the most basic one. Yet given the complexities of human communicative behavior, researchers wishing to investigate its specific aspects face a variety of methodological questions and challenges. These may range from data access to combining methods that do justice to both the semiotic idiosyncrasies of each modality and to the ways in which the different modes interact in a given context.

This chapter provides an example of the empirical steps involved in working with multimodal data consisting of spontaneous speech and its accompanying gestures. It is meant as a companion to the introductory chapter by Mittelberg, Farmer, and Waugh (this volume), which offers a general overview of approaches to discourse and corpus analysis. It is important to realize that the routine presented below is only one of many ways of transcribing, coding, and analyzing co-speech gesture (see also Sweetser this volume). While special attention is paid to gesture-speech co-expression, transcribing naturally occurring speech data, which represents the logical first step in multimodal data transcription, will also be discussed. Overall, the chapter reflects and advocates the trend, noticeable already for some time in various areas of linguistic inquiry, to account for bodily semiotics in the recording and analysis of spoken discourse (e.g., Cienki 1998; Clark 1996; De Stefani fc.; Duranti 1997; Goodwin 1986; Kendon 2004; Lantolf & Thorne 2006; Levinson 2003; McNeill 1992, 2005; Müller 1998, 2004a/b; Schegloff 1984; Streeck 1993, 1994; Sweetser 1998; Tao 1999).

Despite this substantial body of work and the ubiquity of gesture in human communication, manual gestures and other types of bodily semiotics, such as gaze and head