
Spatial Cognition and Computation **2:** 157–180, 2000.

Using space to describe space: Perspective in speech, sign, and gesture

KAREN EMMOREY[1], BARBARA TVERSKY[2] and HOLLY A. TAYLOR[3]
[1]*Laboratory for Cognitive Neuroscience, The Salk Institute for Biological Studies, 10010 North Torrey Pines Rd., La Jolla, CA 92037, U.S.A. (Fax: (858) 452-7052, E-mail: emmorey@salk.edu);* [2]*Stanford University;* [3]*Tufts University*

Received 11 October 2000; accepted 30 May 2001

Abstract. Describing the location of a landmark in a scene typically requires taking a perspective. Descriptions of scenes with several landmarks use either a route perspective, where the viewpoint is within the scene or a survey perspective, where the viewpoint is outside, or a mixture of both. Parallel to this, American Sign Language (ASL) uses two spatial formats, viewer space, in which the described space is conceived of as in front of the speaker, or diagrammatic space, in which the described space is conceived of as from outside, usually above. In the present study, speakers of English or ASL described one of two memorized maps. ASL signers were more likely to adopt a survey perspective than English speakers, indicating that language modality can influence perspective choice. In ASL, descriptions from a survey perspective used diagrammatic space, whereas descriptions from a route perspective used viewer space. In English, iconic gestures accompanying route descriptions used the full 3-D space, similar to viewer space, while gestures accompanying survey descriptions used a 2-D horizontal or vertical plane similar to diagrammatic space. Thus, the two modes of experiencing environments, from within and from without, are expressed naturally in speech, sign, and gesture.

Key words: gesture, perspective, sign language, spatial description, spatial language

Introduction

Where's the train station? That question, or one like it, is among the first in any traveler's guide to a local language. Answering the question, describing the location of a landmark, typically requires taking a perspective (e.g., "The train station is on Bahnhoff Strasse, north of the bus terminal" or "You go down Bahnhoff Strasse, and you'll see the bus terminal on your left and then the train station"). How I tell you where the train station is depends on how I think about the environment or how I think you're thinking about the environment. In describing environments with many landmarks, such as a convention center or a town, speakers of spoken languages such as English or Dutch generally take either a *route* or a *survey* perspective, or a mixture of

the two (Taylor and Tversky 1996). In a route perspective, the viewpoint is within the scene, and the addressee is taken on a mental tour of the environment, so the viewpoint changes with the addressee's position. Landmarks are described relative to the current imagined position of the addressee in terms of "left," "right," "front," and "back." For example, "After you turn right on Michigan Avenue from Randolph, you will see the Chicago Symphony on your right and the Art Institute on your left." In a survey perspective, the imagined viewpoint is stationary and outside and above the environment. Landmarks are described relative to each other in terms of "north," "south," "east," and "west" (Perrig and Kintsch 1985; Taylor and Tversky 1992a, b, 1996). For example, "The Museum of Natural History is west of Central Park and the Metropolitan Museum is on the east side of the Park, a little north of the Natural History Museum." Occasionally, speakers may adopt a *gaze* perspective, which can be viewed as an amalgam of route and survey perspectives. In a gaze perspective, speakers take a point of view just at the edge of a scene, for example, an entrance to a room; and they describe the locations of objects to other objects relative to that point of view, e.g., "the lamp is left of the reading chair" (Ehrich and Koster 1983). Gaze perspective descriptions are relatively rare for large-scale environments and thus won't be discussed here.

Route and survey descriptions

Route and survey perspectives differ with respect to 1) point of view (moving within the scene vs. fixed above the scene), 2) reference object (the addressee vs. another object/landmark), and 3) reference terms (right-left-front-back vs. north-south-east-west). Route and survey perspectives also correspond to two natural ways of experiencing an environment. A route perspective corresponds to experiencing an environment from within, by navigating it, and a survey perspective corresponds to viewing an environment from a single outside point at a height, such as a tree or a hill. In a survey description, the viewer is clearly outside the environment, looking on it almost as an object with parts; whereas in a route description, the viewer is immersed within the environment. The correspondence of the description perspective to natural ways of experiencing environments may account for the consistency of perspective that appears in many spatial descriptions (Taylor and Tversky 1992a).

For spoken languages, which perspective is adopted seems to depend at least in part on features of the environment. In research eliciting descriptions of a number of environments learned from maps or by exploration and varying on many features, participants tended to use relatively more survey perspectives when the environment had features on several size scales and

when there were several routes through the environment. Participants used relatively more route perspectives when environmental features were on a single size scale and when there was only a single natural route through the environment (Taylor and Tversky 1996). In about half the cases, participants mixed perspectives; frequently, perspective switching occurred where portions of the environments matched the ideal conditions for one or the other perspective.

Expression of spatial perspective is especially interesting to study in a signed language because of a unique resource afforded by the visual-manual modality: the use of space to represent space. In fact, American Sign Language (ASL), the primary language of the Deaf[1] in most of North America, uses two different spatial formats to express locations and spatial relations between objects (Emmorey and Falgier 1999). *Signing space* is the term used for the three-dimensional space in front of the signer, extending from the waist to the forehead, where signs can be articulated. In the *diagrammatic* spatial format, signing space represents a 2-D or 3-D map-like model of the environment and takes a fixed bird's eye view from above the environment. Signing space is either a low horizontal plane (tabletop) or a vertical plane (blackboard). In the *viewer* spatial format, signing space reflects the individual's current 3-D view of the environment, so the viewpoint can change as turns are described.[2]

The two spatial formats proposed for ASL bear resemblance to the two perspectives that appear in spoken language descriptions of environments. The essential quality separating route and survey perspectives and use of viewer and diagrammatic space seems to be whether the viewer is conceived of as immersed in an environment or whether the viewer is conceived of as outside an environment, looking at it as an object. These two modes of conceptual relationship between a viewpoint and an environment have been evident in previous work on perspective (Bryant et al. 1992; Bryant and Tversky 1999; Franklin et al. 1992). In these experiments, internal or immersed and external perspectives could be induced in several ways, by the nature of the description of viewpoint, by instructions, by conveying the environment through diagrams or models. In each case, the immersed or internal viewpoint led to different consequences compared to the external viewpoint, specifically with respect to the time required to respond to direction probes (e.g., above, behind, etc.).

Comparison of English, ASL, and gesture

Given the similarities of the two ASL spatial formats to the two frequent spoken perspectives, and given their correspondence to two basic ways of experiencing the world, two questions arise. First, does the use of diagrammatic and viewer spatial formats in ASL coincide with the choice of survey

and route perspective descriptions by signers? And second, do the same environmental features that elicit survey or route perspective descriptions for English speakers, elicit the same perspective choices for ASL signers?

The comparison of spoken and signed languages also invites the study of gesture, where, as for signed language, space can also be used to describe space, though not in any conventionalized way. In fact, spatial descriptions elicit relatively large numbers of gestures (Iverson and Goldin-Meadow 1998). Of course, in spoken languages, gestures are not the primary means of communication. Nor are gestures universally used, to the frustration of those who study them. Three aspects of gesture are of special interest to the study of spatial descriptions. First, describing environments is likely to elicit a number of iconic gestures, gestures that schematically depict the situation they are intended to convey (McNeill 1992). Even more than that, describing environments may elicit gestures that construct models. By model, we mean three or more successive gestures that are used to convey structural features of the environment. In research in which an expert explained the workings of a lock to a novice, experts used gestures to construct models of the components and dynamics of the locks (Engle 1998). The third and perhaps most interesting aspect of gestures investigated here is gesture space (i.e., the space in front of a speaker where gestures are made). Studying gestures produced for spatial descriptions allows us to examine the spatial format of gestures. For example, do the spatial features attributed to diagrammatic and viewer spatial formats in ASL apply to gesture space? What is the relation between the perspective indicated by spoken descriptions and the nature of iconic gestures that accompany these descriptions? This study examines the correspondence between the spoken perspective and the gesture space, and the possible relations of gesture space to spatial format in signing.

To investigate how space is used to describe space in sign and in gestures accompanying speech, both ASL signers and English speakers studied one of two maps of two environments, a town or a convention center (see Figure 1). They then described the environment from memory. In a previous study using written English, Taylor and Tversky (1992b, 1996) found that the town elicited primarily survey perspective descriptions, while the convention center elicited primarily route perspective descriptions.

Method

Participants

Forty Deaf ASL signers participated in the study (26 females; 14 males). Twenty-seven signers had Deaf families and learned ASL from birth. Ten signers acquired ASL between age two and seven, and three acquired ASL

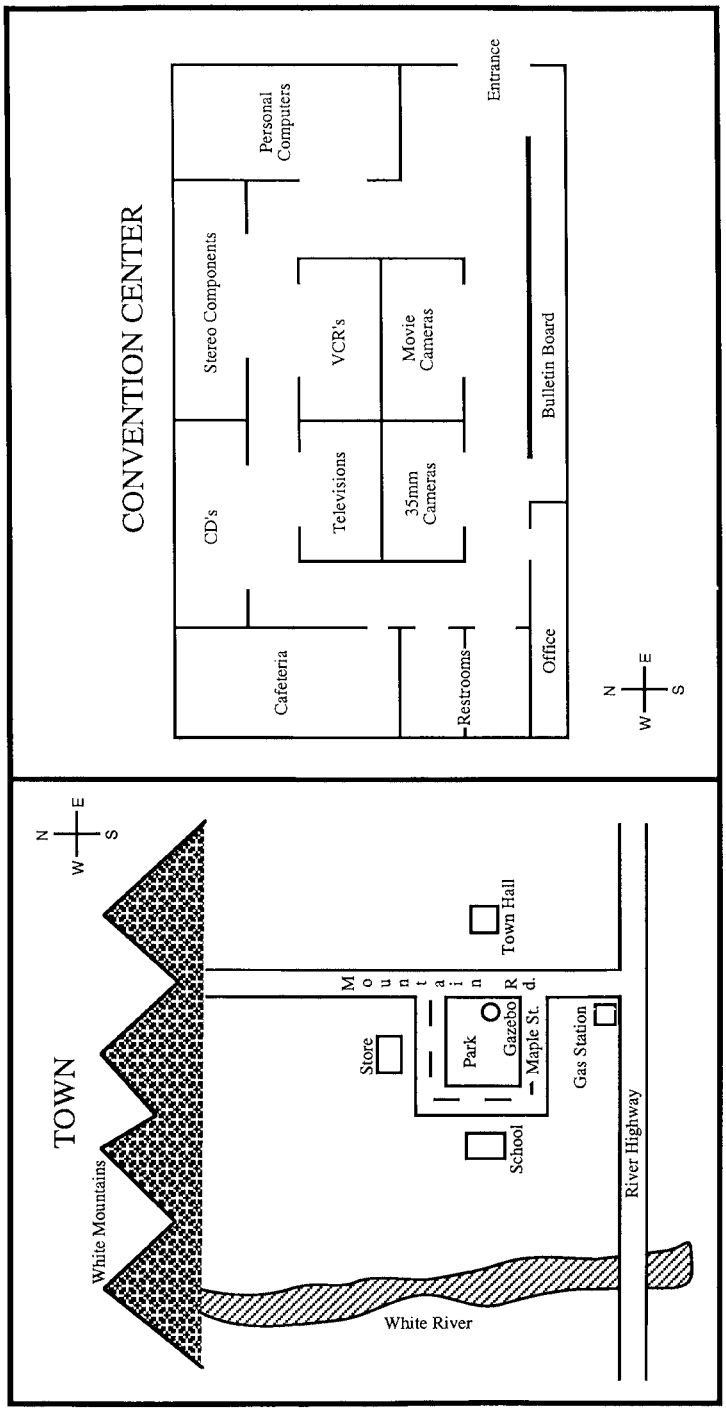

Figure 1. Maps of the town and the convention center from Taylor and Tversky (1992, 1996). Used with permission.

after age fourteen. Thirty-eight signers were deaf at birth or became deaf before 1 year of age, and two signers became deaf at or before age three. Signers participated in the study either at Gallaudet University, The Salk Institute, California State University at Northridge, or at Deaf Community Services in San Diego. All participants were college educated. Forty hearing English speakers also participated in the study (21 females, 19 males); thirty-nine hearing participants were students at Stanford University, and one was a student at the University of California, San Diego.

Materials and procedure

Half of the participants were given the map of the town (Figure 1a), and half were given the map of the convention center (Figure 1b). They were asked to study the map until they had memorized it. Participants were told to describe the environment so that if someone unfamiliar with the area were shown the videotape of their description, they would know what the environment (town/convention center) looked like and where all the landmarks were. The instructions were given either in ASL by a Deaf native signer or in English by a native English speaker.

Results and discussion

Each description was judged as adopting a route perspective, a survey perspective, or a mixed perspective. For the ASL descriptions, two Deaf native signers were asked to decide if the description felt more like a "tour", a bird's eye view description, or a mixture of both. They were shown examples of written English route and survey descriptions from Taylor and Tversky (1996) to familiarize them with how English descriptions were coded. The ASL coders agreed on 88% of judgments for ASL descriptions. When disagreements occurred, the signers discussed the description and came to an agreement. For the English descriptions, two of the authors (HT and BT) independently rated each description and agreed on all of the judgements. The results are shown in Table 1.

The perspectives adopted by ASL signers differed from those of English speaking participants. ASL signers were significantly more likely to adopt a survey perspective when describing the convention center, whereas English participants preferred a route perspective ($X^2 = 8.98$, $p < 0.05$). For the town, English and ASL participants did not differ significantly in perspective choice.

Why do ASL signers use more survey descriptions than English speakers for the convention center? One possibility is that signers prefer survey

Table 1. Perspective choice by ASL signers and English speakers

	Route	Mixed	Survey
Convention Center			
English (N = 20)	11	6	3
ASL (N = 20)	3	5	12
Town			
English (N = 20)	3	6	11
ASL (N = 20)	7	4	9

perspectives in general, perhaps because signing space can be used so effectively to represent a map. That is, participants can locate landmarks on a horizontal plane in signing space in a manner that is isomorphic to the locations of landmarks on a map (in fact, this is how signing space is utilized for survey descriptions). However, ASL signers appear to choose either route or mixed descriptions when describing environments that they have actually experienced themselves. In a pilot study, ASL signers described either their houses (N = 8) or the locations of the dormitories on the Gallaudet Campus (N = 5). Only one person produced a description with a survey perspective. Thus, the difference between English speakers and ASL signers does not appear to be due to an unqualified preference for the use of a survey perspective.

The nature of the linguistic system may affect which spatial perspective is chosen. The fact that signers studied a map may have influenced how they structured signing space within their descriptions, which in turn may have affected their perspective choice. That is, a mental representation of the map may be more easily expressed using a horizontal plane in signing space, and this type of spatial format may be more compatible with a survey perspective, as expressed in ASL. There is no such natural correspondence in English. However, the apparent preference for survey perspectives in ASL requires further substantiation.

It is possible that the differences between ASL and English regarding perspective choice are due to differences in linguistic judgment criteria used by the ASL and English judges. To determine whether ASL and English participants used similar language, we examined the use of motion verbs and spatial terms. Although ASL signers tended to rely on classifier constructions (see below) and the topographic use of signing space in their environment descriptions, signers did produce some lexical spatial terms. We compare the use of these terms with their English counterparts.

Table 2. Mean number of English and ASL motion verbs and relational terms per discourse type

	Motion verbs	Left-Right terms	Cardinal directions
ASL			
Route (N = 10)	9.1 (8.0)[a]	3.7 (4.2)	1.2 (2.1)
Mixed (N = 9)	3.8 (2.8)	1.3 (1.6)	0.75 (1.4)
Survey (N = 21)	1.5 (1.2)	0.14 (0.35)	1.0 (1.8)
English			
Route (N = 14)	13.7 (6.6)	11.4 (5.5)	1.6 (1.7)
Mixed (N = 12)	8.8 (8.2)	4.6 (4.8)	12.5 (6.3)
Survey (N = 14)	0.5 (1.2)	1.5 (3.2)	10.1 (4.7)

[a]Standard deviations are given in parentheses.

Previously, Taylor and Tversky (1996) found that written English route descriptions contained significantly more "active" verbs (primarily motion verbs) and more terms that related a landmark to the viewer (e.g. *left, right*); whereas, English survey descriptions contained more "stative" verbs (i.e., existential verbs) and more relational terms that related a landmark to the environment (e.g., *North, South*). Since stative verbs such as the copula (forms of *to be*) or verbs like *stand* or *lie* are rarely (if ever) used to express locative relations in ASL, we did not attempt to count these verb forms in ASL.[3]

Motion verbs

ASL expresses motion with both classifier predicates in which the handshape represents the moving object and with lexical verbs such as DRIVE, PASS, or WALK. We counted the occurrence of both of these verb types for each ASL description. For English, only motion verbs that described the hypothetical movements of "you" were counted (e.g., fictive motion verbs like *run* in "Mountain Road runs North-South" were not included). The mean number of motion verbs for each description within a discourse category is shown in Table 2.

Route descriptions contained significantly more motion verbs than survey descriptions for ASL signers ($t(30) = 4.4$, $p < 0.01$) and for English speakers ($t(26) = 7.4$, $p < 0.001$). The most frequent ASL motion verbs were DRIVE, VEHICLE-MOVE (using the ASL vehicle classifier), PASS, CROSS (as in "cross over the river"), and TURN-LEFT, RIGHT.[4] These verbs were most often used in route descriptions of the town. The most frequent English motion verbs were *walk*, *turn*, and *go* with path direction particles (e.g., *go*

straight, turn left, go West, walk back, go through, etc.). English speakers also used a greater variety of lexical verbs (e.g.*continue, head, pass, cross, enter, make a right/left, drive, follow, travel*, and *proceed*). Motion verbs were rarely found in survey descriptions for either ASL or English.

Usage of the lexical relational terms left/right

As predicted, English speakers who provided route descriptions were significantly more likely to use the terms *left* and *right* compared to speakers producing survey descriptions ($t(26) = 5.8$, $p < 0.001$). Similarly, ASL route descriptions were significantly more likely to contain the signs LEFT, RIGHT (or LEFT-TURN, RIGHT-TURN) than survey descriptions ($t(30) = 4.01$, $p < 0.01$). However, the use of these signs was relatively rare, even within route descriptions (see Table 2).

Usage of north, south, east, west

English survey descriptions contained significantly more references to cardinal directions than route descriptions ($t(26) = 6.4$, $p < 0.001$). In contrast, the use of ASL cardinal direction signs did not differ for survey and route descriptions (see Table 2). However, this may be due to a floor effect, since signers produced so few cardinal direction signs.

In summary, these data indicate that similar criteria were used by the ASL and English judges when determining the perspective type of a given description. The only difference may be that the ASL judges could not rely on the use of cardinal direction terms to identify survey descriptions; rather, a survey perspective in ASL would be indicated by a lack of both motion verbs and relational left/right terms. In addition, the data from English speakers indicates that language mode (writing vs. speaking) does not affect the basic spatial language that characterizes route vs. survey perspectives. Like the written descriptions from Tversky and Taylor's (1996), spoken descriptions with a route perspective contained more viewer-relational terms and more motion verbs; whereas, spoken descriptions with a survey perspective contained more cardinal direction terms.

Signed descriptions are faster than spoken ones

One other finding of interest with respect to the comparison of spoken English and ASL is that ASL signers took significantly less time (37% less) to describe the environments compared to English speakers (69 vs. 110 seconds, respectively; $t(78) = 4.03$, $p < 0.001$). This difference does not appear to be due to a speed-accuracy trade-off. Signers and speakers did not differ in the number of landmarks omitted from their descriptions ($t(78) = 1.06$, n.s.). This

result replicates Emmorey's (1996) finding of shorter description times for ASL signers when describing the spatial layout of rooms in a doll house. Emmorey (1996) hypothesized that the use of signing space to represent physical space and the use of the hands to represent objects may result in the relatively rapid and efficient expression of spatial relations in ASL. For nonspatial language, the proposition rate is equal for ASL signers and English speakers, but individual ASL signs take about twice as long to articulate as English words (Bellugi and Fischer 1972).

Lexical comparison of ASL and English

Finally, the data in Table 2 indicate that the *lexical* encoding of spatial perspective within a discourse is similar for both ASL and English. However, for ASL signers, lexical spatial terms are not the primary mechanism for expressing spatial perspective, as attested by the rarity of these terms. For English speakers, directional terms such *left/right* and *East/West* are frequent, but they are hard to produce and comprehend (e.g., Franklin and Tversky 1990; Scholl and Egeth 1981). For example, some speakers specify a landmark as *near* another in order to avoid the use of directional terms (Tversky 1996). The use of signing space in ASL generally obviates the need for directional terms because signers have a more direct way of encoding directional information, e.g., by articulating signs to the left or to the right in signing space itself. We next examine those aspects of spatial language that are unique to signed languages.

Perspective choice and spatial format in ASL

When describing environments in ASL, the identity of each landmark is generally indicated by a lexical sign (e.g., BULLETIN-BOARD, SCHOOL, STORE), and the location of the landmarks, their orientation, and their spatial relation with respect to one another are indicated by where the appropriate accompanying "classifier" sign is articulated in the space in front of the signer (see Emmorey, in press). For example, the B handshape (a flat hand) is the handshape used to refer to rectangular, flat, surface-prominent objects, such as a bulletin board. Signers can also use pointing signs to indicate the locations of landmarks in signing space. Where English uses a linear string of prepositions and adjunct phrases to express spatial relations, ASL uses the visual layout displayed by various signs positioned in signing space. This visual layout can utilize either a diagrammatic or a viewer spatial format. Table 3 summarizes the properties associated with these two spatial format types (Emmorey and Falgier 1999).

Table 3. Properties associated with two spatial formats in ASL

Diagrammatic space	Viewer space
• Signing space represents a map-like model of the environment	• Signing space reflects an individual's view of the environment at a particular point in time and space
• Space can have either a 2-D "map" format or a 3-D "model" format	• Signing space is 3-D (normal-sized scale)
• The vantage point does not change (generally a bird's eye view)	• Vantage point can change
• Relatively low horizontal signing space or a vertical plane	• Relatively high horizontal signing space

To determine the relation between the use of spatial formats and perspective choice, for each ASL description, a native signer with linguistic training determined whether the location of each landmark was described using viewer space or diagrammatic space, according to the criteria listed in Table 3. When a survey perspective was adopted, 91% of landmarks were located using a diagrammatic spatial format. When a route perspective was adopted, 88% of landmarks were located using viewer space.

Figure 2 provides an example of the use of diagrammatic space, illustrating the pointing signs produced by one participant to indicate the locations of the outer rooms of the convention center. The locations within signing space map isomorphically to the locations of the rooms on the convention center map (Figure 1b). This particular signer is unusual because she did not rotate the map. That is, most signers (80%) "shifted" the map so that the entrance was located at the chest and the bulletin board extended outward on the left of signing space. This pattern may reflect a convention for spatial descriptions of buildings (and rooms) in ASL: position a main entrance at the front of the body. In fact, most sketch maps are drawn with the starting point at the bottom the page, ignoring cardinal directions, as if the user could shrink and walk out on the map (Tversky 1981; Tversky 2000).

Viewer space is used when signers conceptualize the environment as present, describing the scene as if they were viewing the landmarks and other surrounding elements. Signers know that their addressees cannot see the environment, and therefore the description is not the same as if both discourse participants were simultaneously observing the environment. When both signers can view the environment, they use a type of *shared space* (Emmorey, in press).

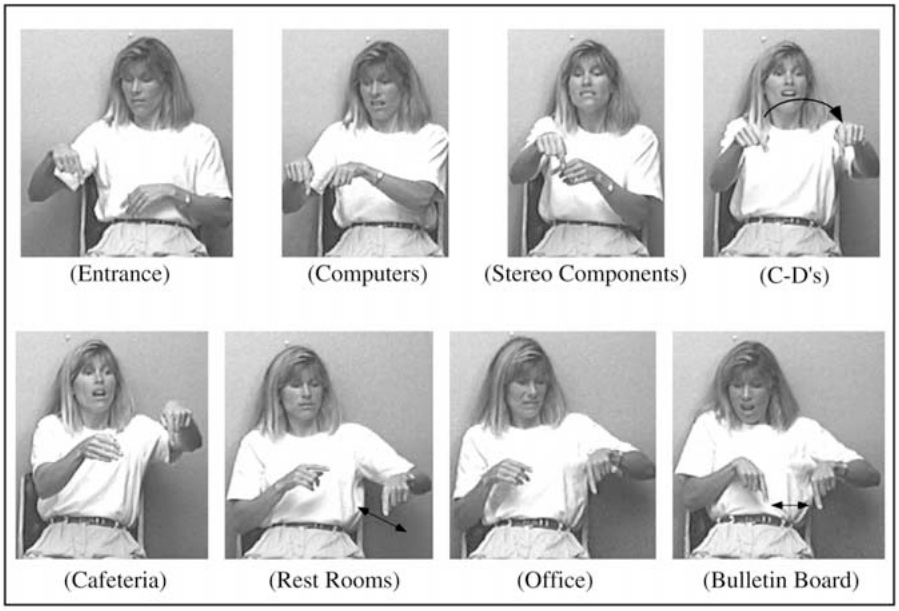

Figure 2. Illustration of the diagrammatic spatial format in ASL. The figure shows pointing signs used in a survey description of the outer rooms of the convention center. The intervening lexical signs are not shown, and the lexical and pointing signs for the inner rooms are also omitted. The rooms associated with the pointing signs are given in parentheses.

Figure 3 illustrates a route description of the convention center and shows the pointing and classifier signs used to indicate the locations of the outer rooms. In contrast to diagrammatic space, the locations in viewer space map to what the signer would observe if she moved along the corridor. The vantage point is not fixed, but changes with motion through space. For example, the signer indicates that the CD room would be in front of her (as she stands next to the Cafeteria), but later she indicates that the Personal Computer room is in front of her because she has described going around the corner (cf. the convention center map in Figure 1). The spatial meaning of signing space *changes* with the description. Now compare the location of these rooms as described by the signer in Figure 2; in this case, signing space represents a model of the entire convention center, and the spatial relationship among locations does not change during the description. Finally, note the relatively high signing plane used in the description shown in Figure 3. When signers describe environments as if they are moving through them, they tend to articulate signs at a relatively high plane, thereby indicating that the description reflects their own imagined view of the environment (see also Lucas and Valli, 1990).

In summary, ASL signers structured signing space differently depending on whether they adopted a route or survey perspective. The majority of land-

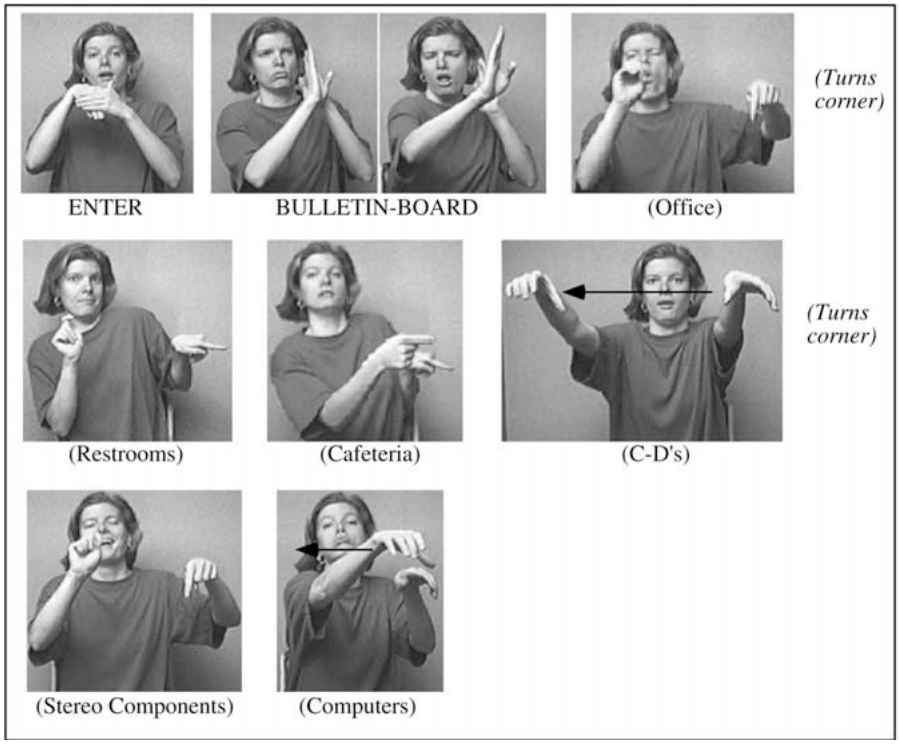

Figure 3. Illustration of the viewer spatial format in ASL. The figure shows pointing and other signs used in a route description of the outer rooms of the convention center. The intervening lexical signs for the rooms are not shown, and the lexical and pointing signs for the inner rooms are also omitted. The rooms associated with the pointing signs are given in parentheses.

marks were described using viewer space for route descriptions, but when a survey perspective was adopted, landmarks were most often described using diagrammatic space.

Comparing gesture accompanying speech and ASL signs

We next examine the similarities and differences between the spatial gestures used by speakers (i.e., those gestures that convey spatial information, rather than beat gestures or abstract metaphorical gestures) and classifier constructions produced by ASL signers when describing the same environments. As with any study of spontaneous gestures, some individuals produced many gestures and others close to none. For this reason, analyses of gestures are not like analyses of other responses that are generated by all participants. Two coders determined whether gestures were iconic or non-iconic and whether a string of gestures created a model. A model was defined as three or more gestures that referred consistently to the same spatial structure, building it

up in segments. By consistently, we mean gestures referring to connecting entities that use the same viewpoint and size scale. These criteria are similar to those of Engle (1998). The coders agreed on 70% of the judgements and came to agreement on the others.

Participants produced frequent non-iconic gestures, e.g., a gesture in which both palms are turned up, as if to present something to the listener (the "conduit" metaphor in which the discourse itself is presented as an object to the listener, see McNeill, 1992) or a gesture with the circular motion that accompanied a search for words. However, our comparative analysis will only concern the iconic gestures produced by speakers.

Comparing gestures and signs. First, there are many critical differences between gestures that accompany speech and sign language (see Emmorey, 1999, for a review; see also McNeill, 1992). Some of the most important distinctions are listed below:

- Signs have a sublexical (phonological) structure, not found in gestures. That is, signs, but not gestures, are governed by phonological rules such as assimilation and have both a segmental and a syllabic structure.
- Signs participate in a hierarchical constraint based system of generative lexical processes, whereas gestures are generally holistic without internal structure.
- Signs are combined into sentences subject to language specific and universal constraints on grammatical form; gestures occur in combination only in special circumstances, and successive gestures do not form a larger hierarchical structure with internal ordering constraints.
- Unlike signs, gestures are timed to coincide with a structural element (the prosodic stress peak) of a separate co-expressive system (i.e., speech).
- Gestures, unlike signs, are relatively idiosyncratic with no agreed upon standard of form (excluding emblems like "thumbs up").

Certainly, these distinctions hold for the comparison of spatial gestures and ASL classifier constructions, which are used to specify spatial information. Nevertheless, iconic gestures produced by speakers during their spatial descriptions exhibited some similarities with ASL classifier constructions. Twenty-six participants used either their fingers (17) or hands (9) to trace paths or shapes. An example of a gesture that traces a path is shown below in Figure 6, gesture (b); and an example of a gesture that indicates the extent of an object (the bulletin board) is shown in Figure 7 below, gesture (c).[5] Parallel to these gesture types, ASL has a type of classifier construction in which the hands trace the extent of an object or trace a path of motion (Supalla, 1986; Valli and Lucas, 1995). Examples of such constructions can be found

in the description of the extent of the Bulletin Board in Figure 2 and in the description of the C-D and Computer rooms in Figure 3.

Iconic gestures were similar to ASL classifier constructions in that the hand could represent a landmark or an object. Areas, such as rooms, were frequently indicated by either forming the thumb and index finger into a "U" (8 participants) or by forming both hands into an open rectangle, fingers together, apart from the thumbs (8 participants) or by moving open hands in parallel (6 participants). This "rectangular" hand configuration is similar to the first syllable of the ASL sign ROOM, but the U-type handshape could never be used to represent a room in ASL. The gesture handshapes and some of their functions (e.g., to represent an object or to trace a path) were similar to classifier constructions in ASL, but of course for ASL, linguistic constraints determined which handshape could be chosen to represent a particular object.

Although the majority of gestures produced by speakers were singleton gestures, twenty-five participants produced at least one string of three or more gestures that constituted a model. Overall, 91 models were produced, and these were roughly distributed across route, survey, and mixed perspective types. Some of the models involved a long sequence of gestures; for example, one participant describing the convention center, conveyed the locations of bulletin board, office, rest rooms, cafeteria, CD displays, stereo displays, computer displays, and the entrance in a continuous coherent sequence of 15 gestures.

One similarity between gesture models and signed descriptions was the use of "anchor" gestures or signs when describing the location of multiple landmarks. Figure 4 provides an example from an ASL description of the town. After tracing the outline of Maple Street, the signer leaves his index finger in space, which serves as an anchor point representing the end of Maple street. He then locates the park, school and store in signing space while maintaining the point to the north end of Maple street. The signer appears to be deliberately holding this pointing sign because he does not release the point when making the two-handed sign SCHOOL (the citation form of this sign is made with both hands in a 5 hand configuration).

Similarly, many of the speakers (44%) produced gesture strings in which one hand appears to represent an anchor landmark (and its location), and the other hand indicates the location of other landmarks. In the example shown in Figure 5, the speaker's left hand produces the anchor gesture representing the location of the school in the town and is held throughout the speech. The speaker represents the location of the school on the map with his left hand, which is held in space while he describes the location of the store and the town hall. The gestures that accompany these descriptions are produced at the appropriate locations in gesture space with respect to an imagined map with the school positioned on the left (see Figure 1a).

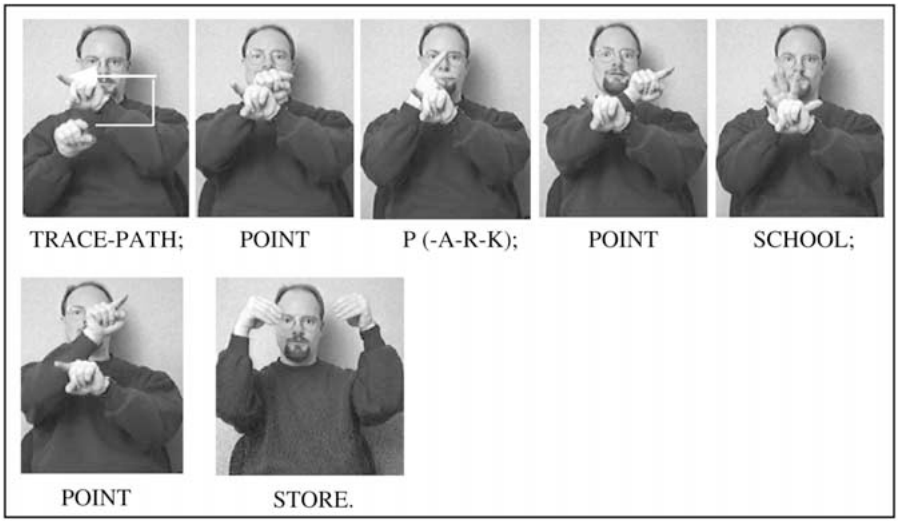

Figure 4. Illustration of the use of anchor signs within the vertical plane in signing space. English translation: "(The street) is like this. Here's a park; here's a school; and here's a store."

Thus, the gestures of speakers resembled some aspects of ASL classifier constructions in that the hand was used to represent a landmark, and speakers co-ordinated their hands in space during their descriptions to represent the relative locations of one landmark to another, creating a type of model of the environment in gesture space. Gestures differed from signs with respect to consistency (e.g., not all speakers gestured or produced gesture strings) and linguistic constraints (e.g., choice of handshape).

Comparing gesture space and signing space. Both signers and speakers use sign/gesture space to represent locations; yet, there are some interesting differences. ASL signers never used the vertical plane in signing space when describing the convention center. This constraint may be due to the fact that the vertical plane is limited to two dimensions in ASL, and the use of the sign ROOM to describe the convention center invokes a three dimensional space – therefore, signers used only the horizontal plane which can be three dimensional (representing a model, rather than a map, in diagrammatic space). English speakers were not constrained by such linguistic requirements, and 40% of the speakers produced gestures using a vertical plane parallel to the body when describing the convention center. The vertical plane seemed to be adopted to make the model created by gesture visible to listeners, and an example is provided in Figure 6. The spatial layout is in the vertical plane (e.g., the speaker's gestures indicate that the Cafeteria is above the restrooms in space).

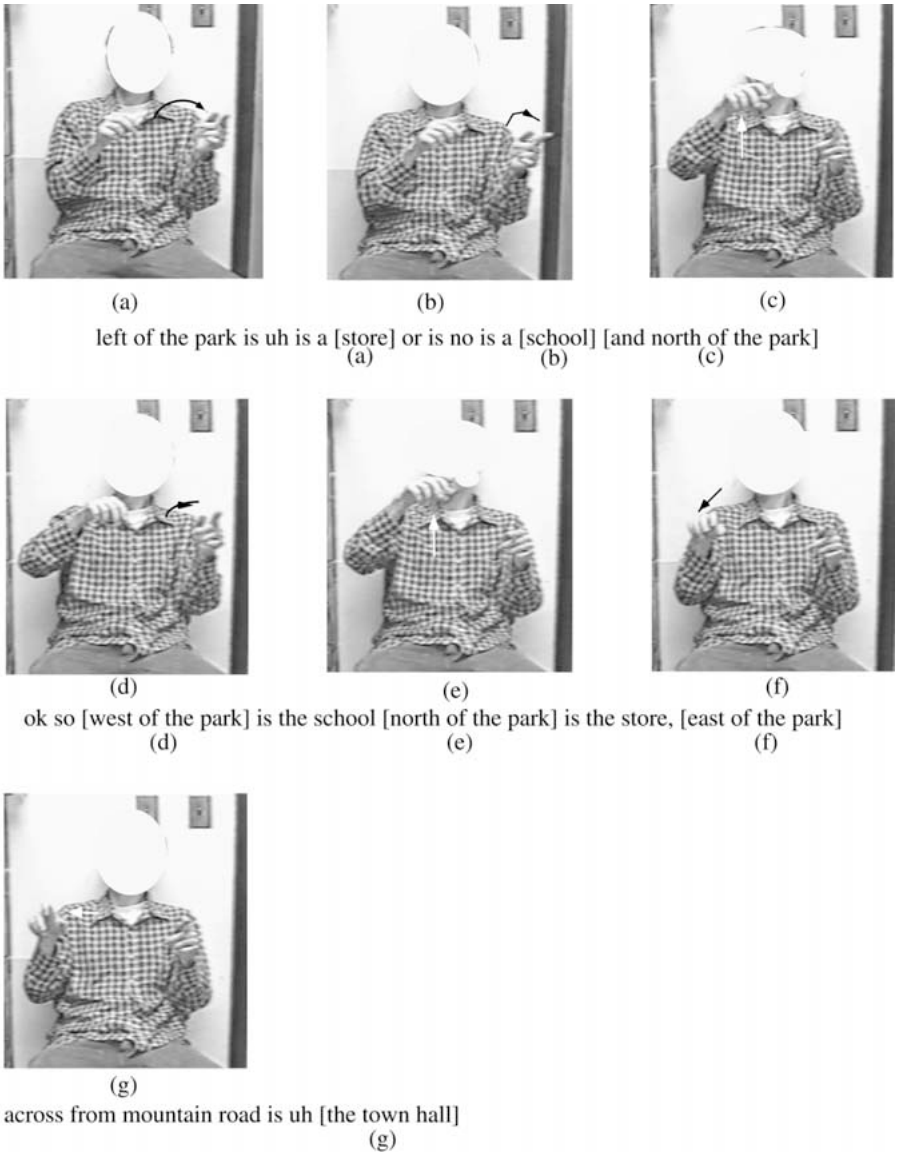

Figure 5. Illustration of the gestures accompanying a description of the town by speaker 31. The extent of the gesture is shown by enclosing the concurrent speech in square brackets. The lower case letter underneath each section of bracketed text corresponds to the accompanying gesture shown above the text. This example illustrates the use of an anchor gesture.

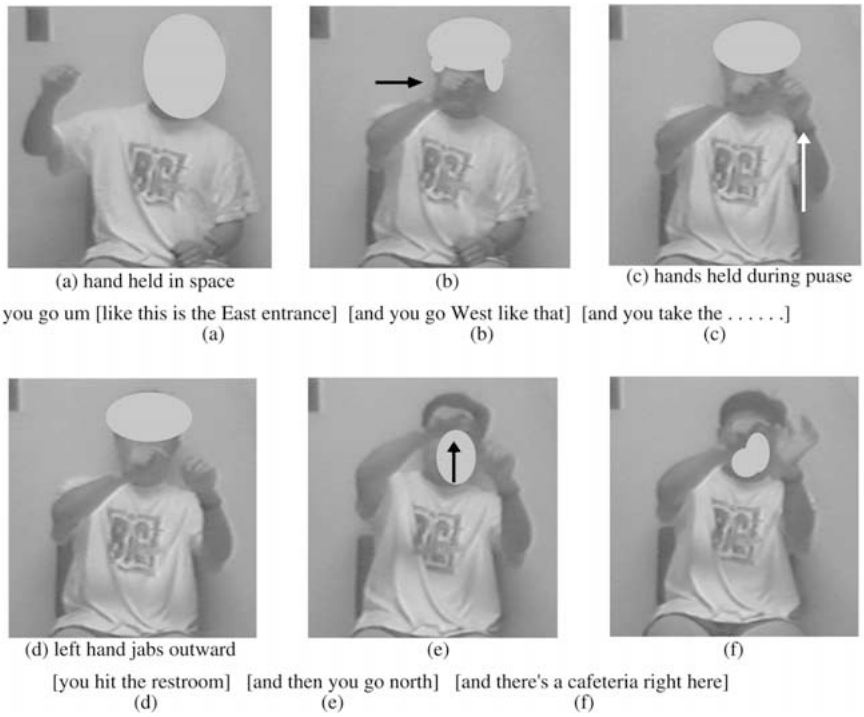

Figure 6. Illustration of the gestures accompanying a description of the convention center by speaker 3. The extent of the gesture is shown by enclosing the concurrent speech in square brackets. The lower case letter underneath each section of bracketed text corresponds to the accompanying gesture shown above the text. This example illustrates the use of the vertical plane when describing the convention center.

One striking similarity between gesturing and signing is that speakers also appeared to structure gesture space using a type of diagrammatic vs. viewer spatial format. To determine whether speakers consistently matched their use of gesture space to the verbal perspective, two new coders who were naive to the study first analyzed all gestures. They coded gestures without the soundtrack, so that there were no clues to the linguistically encoded perspective. Each hand gesture was marked as tabletop, blackboard, viewer-like, or other. Tabletop and blackboard gestures were characterized by the use of a 2-D horizontal or vertical plane, which corresponds roughly to the use of diagrammatic space in ASL (see Figures 5 and 6). The viewer-like gestures were characterized as gestures that appeared as though the speaker were in the environment indicating directions from their viewpoint (see Figure 7). Viewer-like gestures generally moved away from the speaker and used the full 3-D space (rather than a 2-D plane) which corresponds roughly to the use of viewer space in ASL. Gestures that were coded as "other" could

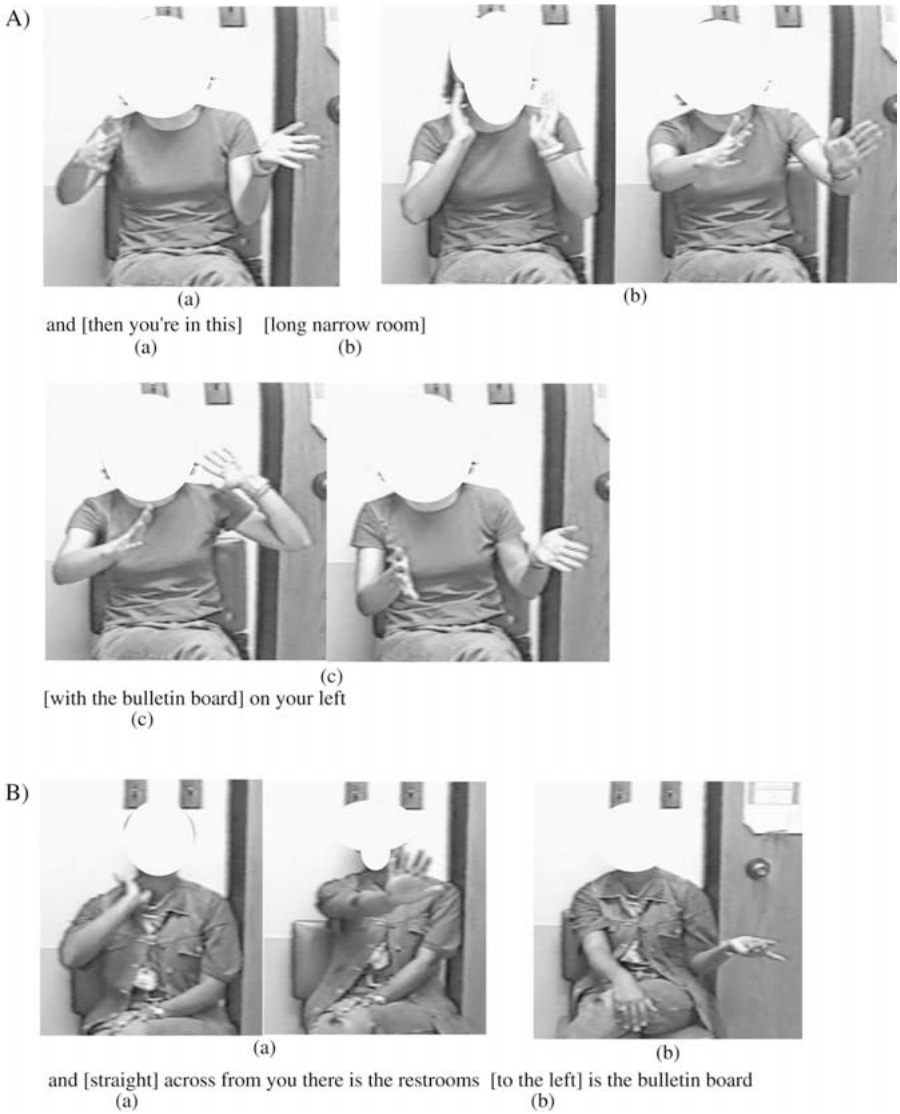

Figure 7. Illustration of the gestures accompanying a description of the convention center by A) speaker 38 and B) speaker 26. The extent of the gesture is shown by enclosing the concurrent speech in square brackets. The lower case letter underneath each section of bracketed text corresponds to the accompanying gesture shown above the text. These examples illustrate the use of viewer space for gesture.

not be unambiguously coded into one of the other three categories. Coders agreed on 80.5% of judgements, and disagreements were eliminated from the analysis. These gestures were then compared with their accompanying spoken description to examine the correspondence between the use of gesture space and perspective choice. To avoid any ambiguities, descriptions that used mixed perspectives were excluded from the analysis. Also excluded were participants who did not gesture. This left 10 route descriptions and 6 survey descriptions.

Gesture space and description perspective were highly related. For survey descriptions, 50.6% of the gestures used diagrammatic space and 23.3% used viewer space. In contrast, for route descriptions, 54.7% of the gestures used viewer space and 17.1% of the gestures used diagrammatic space (interaction between description perspective and gesture space, $F(1,14) = 7.07$, $p < 0.02$).

The examples in Figures 5 and 6 illustrate of the use of diagrammatic space for gestures in which the gesture space represents a map-like model of the environment being described. The examples shown in Figure 7 illustrate the use of viewer space for gesture in which the speakers appear to imagine the environment as present in front of them rather than as a map, and their gestures reflect this conceptualization. In both examples, the speakers produce gestures that seem to refer to a three-dimensional imagined environment rather than to a two-dimensional map.

These variations in gestures and gesture space are not unlike McNeill's (1992) analysis of character vs. observer viewpoint or "voice" in gesture:

> The use of space also differs for these voices [character vs. observer] . . . With the character voice the space envelops the narrator – it is a space for the enactment of the character, and includes the locus of the speaker at its center. With an observer's voice, in contrast, the narrative space is localized in front of the narrator – as if it were an imaginary stage or screen – and in this space the narrator moves the relatively undifferentiated figures (blobs). (p. 190)

Thus "character viewpoint" may reflect a description of the space in which the speaker imagines herself as in the environment, and "observer viewpoint" characterizes gesture space that reflects a view of the environment as a map or diagram.

Correspondence between gesture space and signing space. Combining the gesture space results with the ASL space results yields a provocative correspondence. In both spontaneous gesture and signing, the space forward from the body, the space of viewing and action, is more likely to accompany a route perspective, where the traveler is described as facing, usually moving forward, and viewing the environment from within. In these cases, the spatial

relations of the landmarks are described relative to the viewer. Similarly, for both spontaneous gesture and ASL, the depictive space typically of the table-top, but also of the black-board, is more likely to accompany a survey perspective, where the viewpoint is stationary and from above, and the spatial relations between landmarks are described relative to the reference frame of the external environment rather than the viewer. The correspondence suggests that the use of gesture and signing space may arise naturally from the two most common ways in which humans experience their environment.

General discussion

In describing environments with several landmarks, users of spoken languages frequently adopt one of two perspectives, either within the environment surrounded by it, or outside the environment observing it in entirety. For a route description, the viewpoint is embedded within an environment, changing direction within it. Landmarks are described relative to the changing point of view in terms of *right*, *left*, *front*, and *back*. For a survey perspective, the viewpoint is external to the environment and above it. Landmarks are described relative to each other in terms of *north*, *south*, *east*, and *west*. In a previous corpus, speakers mixed perspectives about half the time, usually without signaling (Taylor and Tversky 1992a).

A special feature of signed languages is that they can use space itself to convey spatial information. Indeed, in ASL, there are two dominant modes of describing objects in space. When viewer space is adopted, it is as if the speaker were traveling through the space, indicating turns but always facing forward, and locating the landmarks relative to the body. When diagrammatic space is adopted, it is as if the speaker were above the environment seeing it as a whole, and locating landmarks relative to each other. These two modes bear striking similarities to the two spoken perspectives, route to viewer space and survey to diagrammatic space. Does the special feature of using space to convey space affect perspective?

Research using a broad spectrum of environments (including indoor, outdoor, large and small, public and private) has shown that English speakers tended to use relatively more survey perspectives when there were several paths through the environment and when landmarks were on several size scales (Taylor and Tversky 1992a, 1996). Among the environments, that of a map of a town elicited the highest proportion of survey descriptions and the map of a convention center elicited the highest proportion of route descriptions. These findings were replicated here for English speakers, but not for ASL signers. Both speakers of English and of ASL preferred survey perspectives for describing the town. However, for the convention center, English speakers preferred a route perspective whereas ASL signers

preferred a survey perspective. This indicates a bias in ASL relative to English toward a survey perspective. We propose two possible and non-contradictory explanations for the bias toward survey perspectives. One reason might be an overall preference for the use of diagrammatic space to convey complex spatial information. Since the diagrammatic spatial format is most consistent with a survey perspective, ASL signers are more likely than English speakers to adopt a survey perspective when describing environments. Another reason may be that the environments were learned from maps. The spatial transformation from a map to diagrammatic space may be so direct as to be compelling in signing. Whatever the explanation for the bias toward survey perspective in ASL, this research makes clear that features of the language as well as features of the environment affect linguistic perspective choice.

The comparison of iconic gestures and ASL signs revealed some intriguing similarities and differences. Similar to ASL classifier constructions, several speakers used their hands to represent landmark objects, and the spatial relation between landmarks was depicted by where their hands were positioned in space. For both ASL and gesture, one hand can serve as an anchor landmark, and other landmarks can be located in space with respect to this landmark. However, signing, unlike gesturing, is subject to linguistic constraints on handshape choice for representing objects, on the choice of a vertical or horizontal plane in signing space, and on the height of signing within a description.

Finally, the spatial perspectives observed for spoken language map naturally to the use of spatial formats in signed language: moving within an environment maps to viewer space and observing an environment from without maps to diagrammatic space. The gestures of English speakers complete this elegant correspondence between conceptual perspective and use of space. When the linguistic description was from within the environment (a route perspective), speakers produced gestures within a 3-D space as if they were in the environment. When the description was from an external viewpoint (a survey perspective), speakers produced gestures along a 2-D horizontal or vertical plane as if they were drawing on a chalkboard or tabletop. Gestures, unlike signed language, do not have a conventional format; they are optional and not used at all by some speakers. Thus, the correlation of gesture space (and spatial formats in ASL) with linguistic perspective suggests that there is a natural correspondence between how speakers and signers conceptualize the space around their bodies as they communicate with sign, speech, and gesture. This correspondence further suggests that the use of viewer and diagrammatic formats in ASL may derive

from the same natural correspondence, another example of using space to describe space.

Acknowledgements

This research was supported in part by NSF grants SBR-9510963 and SBR-9809002 and NIH grant HD13249 awarded to Karen Emmorey and by a Varenius seed grant awarded to Barbara Tversky and Holly A. Taylor. We thank Steve McCullough, Brenda Falgier, and Kevin Clark for their help in analyzing the ASL descriptions; Bonita Ewan, Brenda Falgier, and Steve McCullough for their help collecting the ASL data; Kulin Tantod for his help collecting the spoken English data; and Eurie Chung, Rachel Groner, David Megathlin, Shelley Wynecoop, Nu Vo, and Jesse Zolna for their help analyzing the gesture data. Portions of the ASL data were described previously in Emmorey and Falgier (1999).

Notes

[1] Following convention, lower case *deaf* refers to audiological status, while upper case *Deaf* is used when sign language use and/or membership in the Deaf community is at issue.
[2] Diagrammatic space is very similar to Liddell's notion of *token space* (Liddell 1995) and to Schick's *model space* (Schick 1990). Viewer space is parallel to Liddell's *surrogate space* and Schick's *real world space*.
[3] The ASL and English verbs and directional terms were counted by the first author, KE, a linguist and fluent signer.
[4] Underlining indicates an initialized sign.
[5] The hearing participants' faces are masked to protect their privacy. The Deaf participants shown in Figures 2–4 gave permission for their images to be published.

References

Bellugi, U. and Fischer, S. (1972). A Comparison of Sign Language and Spoken Language, *Cognition* 1: 173–200.

Bryant, D.J. and Tversky, B. (1999). Mental Representations of Spatial Relations from Diagrams and Models, *Journal of Experimental Psychology: Learning, Memory and Cognition* 25: 137–156.

Bryant, D.J., Tversky, B. and Franklin, N. (1992). Internal and External Spatial Frameworks for Representing Described Scenes, *Journal of Memory and Language* 31: 74–98.

Ehrich, V. and Koster, C. (1983). Discourse Organization and Sentence Form: The Structure of Foom Descriptions in Dutch. *Discourse Processes* 6: 169–195.

Emmorey, K. (1996). The Confluence of Space and Language in Signed Languages. In P. Bloom, M. Peterson, L. Nadel and M. Garrett (eds.), *Language and Space* (pp. 171–209). Cambridge, MA: MIT Press.

Emmorey, K. (in press). *Language, Cognition, and the Brain: Insights from Sign Language Research*. Mahwah, NJ: Lawrence Erlbaum Associates.

Emmorey, K. (1999). Do Signers Gesture? In L. Messing and R. Campbell (eds.), *Gesture, Speech, and Sign*. New York: Oxford University Press.

Emmorey, K. and Falgier, B. (1999). Talking about Space with Space: Describing Environments in ASL. In E.A. Winston (ed.), *Story Telling and Conversations: Discourse in Deaf Communities* (pp. 3–26). Washington, D.C.: Gallaudet University Press.

Engle, R.A. (1998). *Constructing Spatially-Anchored Models with Multimodal Explanations.* Poster presented at the 106th Meetings of the American Psychological Association, San Francisco, CA.

Franklin, N., Tversky, B. and Coon, V. (1992). Switching Points of View in Spatial Mental Models Acquired from Text, *Memory and Cognition* 20: 507–518.

Franklin, N. and Tversky, B. (1990). Searching Imagined Environments, *Journal of Experimental Psychology: General* 119: 63–76.

Iverson, J. and Goldin-Meadow, S. (1998). Why People Gesture When They Speak, *Nature* 396(6709): 228.

Levelt, W. (1982). Linearization in Describing Spatial Networks. In S. Peters and E. Saarinen (eds.), *Processes, Beliefs, and Questions* (pp. 199–220). Reidel Publishing Company.

Levinson, S. (1996). Frames of Reference and Molyneux's Question: Crosslinguistic Evidence. In P. Bloom, M. Peterson, L. Nadel and M. Garrett (eds.), *Language and Space* (pp. 109–170). Cambridge, MA: MIT Press.

Lucas, C. and Valli, C. (1990). Predicates of Perceived Motion in ASL. In S.D. Fischer and P. Siple (eds.), *Theoretical Issues in Sign Language Research, Volume 1: Linguistics* (pp. 153–166). The University of Chicago Press: Chicago, IL.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. The University of Chicago Press.

Perrig, W. and Kintsch, W. (1985). Propositional and Situational Representations of Text. *Journal of Memory and Language* 24: 503–518.

Scholl, M.J. and Egeth, H.E. (1981). Right-Left Confusion in the Adult: A Verbal Labeling Effect, *Memory and Cognition* 9: 339–350.

Supalla, T. (1986). The Classifier System in American Sign Language. In C. Craig (ed.), *Noun Classification and Categorization*. Amsterdam: John Benjamins North America.

Taylor, H.A. and Tversky, B. (1992a). Descriptions and Depictions of Environments, *Memory and Cognition* 20: 483–496.

Taylor H.A. and Tversky, B. (1992b). Spatial Mental Models Derived from Survey and Route Descriptions, *Journal of Memory and Language* 31: 261–292.

Taylor, H.A. and Tversky, B. (1996). Perspective in Spatial Descriptions, *Journal of Memory and Language* 35(3): 371–391.

Tversky, B. (1981). Distortions in Memory for Maps, *Cognitive Psychology* 13: 407–433.

Tversky, B. (1996). Spatial Perspective in Descriptions. In P. Bloom, M.A. Peterson, L. Nadel and M.F. Garrett (eds.), *Language and Space* (pp. 463–492). Cambridge, MA: MIT Press.

Tversky, B. (2000). How Space Structures Language. In C. Freksa, C. Brauer, W., Habel and K.F. Wender (eds.), *Spatial Cognition II: Integrating Abstract theories, Empirical Studies, Formal Models, and Practical Applications* (pp. 72–79). Berlin: Springer.

Valli, C. and Lucas, C. (1995). *Linguistics of American Sign Language: An introduction*, 2nd edn. Washington, DC: Gallaudet University Press.