LANGUAGE AND LINGUISTICS 4.2:207-250, 2003
2003-0-004-002-000039-1

The Representation of Spatial Structure in Spoken and Signed Language: A Neural Model

Leonard Talmy
University at Buffalo, State University of New York

Linguistic research to date has determined many of the factors that structure the spatial schemas found across spoken languages. It is now feasible to determine the comprehensive system that these factors constitute. Among its features, this system has a relatively closed, universally available inventory of fundamental spatial elements that are combined to form whole schemas; it has a relatively closed set of categories that these elements appear in; and it, therefore, has a relatively closed, small number of particular elements in each category of spatial distinctions that each category can ever mark. By contrast, the structural representation of space in signed language systematically differs from that in spoken language in the direction of what appear to be the structural characteristics of scene parsing in visual perception. Among such differences, signed language can mark finer spatial distinctions with its greater inventory of structural elements, categories, and elements per category. It represents many more of these distinctions in any particular expression. It also represents these distinctions independently in the expression, not bundled together into "pre-packaged" schemas. And its spatial representations are largely iconic with visible spatial characteristics. The findings suggest that instead of some discrete whole-language module, spoken language and signed language are both based on some more limited core linguistic system that then connects with different further subsystems for the full functioning of the two different language modalities. The different behaviors of the two language modalities are plausible, given the apparently different intrinsic properties of the neural systems underlying them.

Key words: ASL, language module, sign language, spatial schema, spatial structure

1. Introduction[1]

This paper presents and relates novel perspectives on the representation of spatial structure in three domains: in spoken language, in signed language, and in the neural underpinning of these two language modalities. The analysis of spatial schemas in spoken language has now progressed to where we can catalog most of the basic elements that make them up and observe how combinations of such basic elements behave. We can

1 This version of ongoing research supersedes a prior article, Talmy (in press).

now also observe the ways in which the representation of spatial structure in signed language differs from that in spoken language. And we can accordingly propose a new neural model of our language capacity that accommodates both the similarities and the differences found in these two systems of spatial representation.

Linguistic research to date has determined many of the factors that structure the spatial schemas found across spoken languages (e.g., Gruber 1965, Fillmore 1968, Leech 1969, Clark 1973, Bennett 1975, Herskovits 1982, Jackendoff 1983, Zubin and Svorou 1984, as well as myself, Talmy 1983, 2000a, 2000b). It is now feasible to integrate these factors and to determine the comprehensive system they constitute for spatial structuring in spoken language. This system is characterized by several features. With respect to constituency, there is a relatively closed universally available inventory of fundamental spatial elements that in combination form whole schemas. There is a relatively closed set of categories that these elements appear in. And there is accordingly a relatively closed, small number of particular elements in each category of spatial distinctions that each category can ever mark. With respect to synthesis, selected elements of the inventory are combined in specific arrangements to make up the whole schemas represented by closed-class spatial forms. Each such whole schema that a closed-class form represents is thus a “pre-packaged” bundling together of certain elements in a particular arrangement. Each language has in its lexicon a relatively closed set of such pre-packaged schemas (larger than that of spatial closed-class forms, due to polysemy) that a speaker must select among in depicting a spatial scene. Finally, with respect to the whole schemas themselves, these schemas can undergo a certain set of processes that extend or deform them. Such processes are perhaps part of the overall system so that a language’s relatively closed set of spatial schemas can fit more spatial scenes.

An examination of signed language[2] shows that its structural representation of space systematically differs from that in spoken language in the direction of what appear to be the structural characteristics of scene parsing in visual perception. Such differences include the following: Signed language can mark finer spatial distinctions with its inventory of more structural elements, more categories, and more elements per category. It represents many more of these distinctions in any particular expression. It also represents these distinctions independently in the expression, not bundled together into pre-packaged schemas. And its spatial representations are largely iconic with visible spatial characteristics.

[2] I here approach signed language from the perspective of spoken language, because it is not at this point an area of my expertise. For their help with my questions on signed language, my thanks to Paul Dudis, Karen Emmorey, Samuel Hawk, Nini Hoiting, Marlon Kuntze, Scott Liddell, Stephen McCullough, Dan Slobin, Ted Suppala, Alyssa Wolf, and others—who are not responsible for my errors and oversights.

Our findings point to a new view of the neural implementation of language. They suggest that instead of some discrete whole-language module, spoken language and signed language are both based on some more limited core linguistic system responsible for their commonalities. This system then further connects with different neural subsystems for the full functioning of the two different language modalities.

When formal linguistic investigation of signed language began several decades ago, it was important to establish in the context of that time that signed language was in fact a full genuine language, and the way to do this, it seemed, was to show that it fit the prevailing model of language, the Chomskyan-Fodorian language module. Since then, however, evidence has been steadily accruing that signed language does diverge in various respects from spoken language. The modern response to such observations—far from once again calling into question whether signed language is a genuine language—should be to rethink what the general nature of language is. Part of that enterprise is undertaken in this paper.

2. Fundamental space-structuring elements and categories in spoken language

An initial main finding emerges from analysis of the spatial schemas expressed by closed-class (grammatical) forms across spoken languages. There is a relatively closed and universally available inventory of fundamental conceptual elements that recombine in various patterns to constitute those spatial schemas. These elements fall within a relatively closed set of categories, with a relatively closed, small number of elements per category.

2.1 The target of analysis

As background to this finding, spoken languages universally exhibit two different subsystems of meaning-bearing forms. One is the “open-class” or “lexical” subsystem, comprised of elements that are great in number and readily augmented—typically, the roots of nouns, verbs, and adjectives. The other is the “closed-class” or “grammatical” subsystem, consisting of forms that are relatively few in number and difficult to augment—including such bound forms as inflections and such free forms as prepositions and conjunctions. As argued in Talmy (2000a, chap. 1), these subsystems basically perform two different functions: Open-class forms largely contribute conceptual content, while closed-class forms determine conceptual structure. Accordingly, our discussion focuses on the spatial schemas represented by closed-class forms so as to examine the concepts used by language for structuring purposes.

Across spoken languages, only a portion of the closed-class subsystem regularly represents spatial schemas. We can identify the types of closed-class forms in this portion and group them according to their kind of schema.

1. Types of closed-class forms with schemas for paths or sites include:
   a. forms in construction with a nominal, such as prepositions like English *across* (as in *across the field*) or noun affixes like the Finnish illative suffix *-:n* 'into', as well as prepositional complexes such as English *in front of* or Japanese constructions with a "locative noun" like *ue* 'top surface', (as in *teeburu no ue ni* 'table GEN top at' = "on the table")
   b. forms in construction with a verb, such as verb-satellites like English *out, back* and *apart* (as in *They ran out / back / apart*)
   c. deictic determiners and adverbs such as English *this* and *here*
   d. indefinites, interrogatives, relatives, etc., such as English *everywhere / whither / wherever*)
   e. qualifiers such as English *way* and *right* (as in *It's way / right up there*)
   f. adverbials like English *home* (as in *She isn't home*).

2. Types of closed-class forms with schemas for the spatial structure of objects include:
   a. forms modifying nominals such as markers for plexity or state of boundedness, like English *-s* for multiplexing (as in *birds*) or *-ery* for debounding (as in *shrubbery*)
   b. numeral classifiers like Korean *chang* 'planar object'
   c. forms in construction with the verb, such as some Atsugewi Cause prefixes, like *cu-* 'as the result of a linear object moving axially into the Figure'.

3. Sets of closed-class forms that represent a particular component of a spatial event of motion/location (see Talmy 2000b, chaps. 1 and 2) include:
   a. the Atsugewi verb-prefix set that represents different Figures
   b. the Atsugewi verb-suffix set that represents different Grounds (together with Paths)
   c. the Atsugewi verb-prefix set that represents different Causes
   d. the Nez Perce verb-prefix set that represents different Manners

2.2 Determining the elements and categories

A particular methodology is used to determine fundamental spatial elements in language. One starts with any closed-class spatial morpheme in any language, considering the full schema that it expresses and a spatial scene that it can apply to. One then determines any factor one can change in the scene so that the morpheme no longer applies to it. Each such factor must therefore correspond to an essential element in the morpheme's schema. To illustrate, consider the English preposition *across* and the scene it refers to in *The board lay across the road*. Let us here grant the first two elements in the *across* schema (demonstrated elsewhere): (1) A Figure object (here, the board) is spatially related to a Ground object (here, the road); and (2) the Ground is ribbonal—a plane with two roughly parallel-line edges that are as long as, or longer than the distance between them. The remaining elements can then be readily demonstrated by the methodology. Thus, a third element is that the Figure is linear, generally bounded at both ends. If the board were instead replaced by a planar object, say, some wall siding, one could no longer use the original *across* preposition but would have to switch to the schematic domain of another preposition, that of *over*, as in *The wall siding lay over the road*. A fourth element is that the axes of the Figure and of the Ground are roughly perpendicular. If the board were instead aligned with the road, one could no longer use the original *across* preposition but would again have to switch to another preposition, *along*, as in *The board lay along the road*. Additionally, a fifth element of the *across* schema is that the Figure is parallel to the plane of the Ground. In the referent scene, if the board were tilted away from parallel, one would have to switch to some other locution such as *The board stuck into / out of the road*. A sixth element is that the Figure is adjacent to the plane of the Ground. If the board were lowered or raised away from adjacency, even while retaining the remaining spatial relations, one would need to switch to locutions like *The board lay (buried) in the road. / The board was (suspended) above the road*. A seventh element is that the Figure's length is at least as great as the Ground's width. If the board were replaced by something shorter, for example, a baguette, while leaving the remaining spatial relations intact, one would have to switch from *across* to *on*, as in *The baguette lay on the road*. An eighth element is that the Figure touches both edges of the Ground. If the board in the example retained all its preceding spatial properties but were shifted axially, one would have to switch to some locution like *One end of the board lay over one edge of the road*. Finally, a ninth element is that the axis of the Figure is horizontal (the plane of the Ground is typically, but not necessarily, horizontal). Thus, if one changes the original scene to that of a spear hanging on a wall, one can use *across* if the spear is horizontal, but not if it is vertical, as in *The spear hung across the wall. / The spear hung up and down on the* *wall*. Thus, from this single example, the methodology shows that at least the following elements figure in closed-class spatial schemas: a Figure and a Ground, a point, a line, a plane, a boundary (a point as boundary to a line, a line as boundary to a plane), parallelness, perpendicularity, horizontality, adjacency (contact), and relative magnitude.

In the procedure of systematically testing candidate factors for their relevance, the elements just listed have proved to be essential to the selected schema and hence, to be in the inventory of fundamental spatial elements. But it is equally necessary to note candidates that do not prove out, so as to know which potential spatial elements do not serve a structuring function in language. In the case of *across*, for example, one can probe whether the Figure, like the board in the referent scene, must be planar—rather than simply linear—and co-planar with the plane of the Ground. It can be seen, though, that this is not an essential element to the *across* schema, since this factor can be altered in the scene by standing the board on edge without any need to alter the preposition, as in *The board lay flat / stood on edge across the road*. Thus, co-planarity is not shown by *across* to be a fundamental spatial element. However, it does prove to be so in other schemas, and so in the end must be included in the inventory. This is seen for one of the schemas represented by English *over*, as in *The tapestry hung over the wall*. Here, both the Figure and Ground must be planes and co-planar with each other. If the tapestry here were changed to something linear, say, a string of beads, it is no longer appropriate to use *over* but only something like *against*, as in *The string of beads hung *over / against the wall*.

Now, another candidate element—that the Figure must be rigid, like the board in the scene—can be tested and again found to be inessential to the *across* schema, since a flexible linear object can be substituted for the board without any need to change the preposition, as seen in *The board / The cable lay across the road*. Here, however, checking this candidate factor across numerous spatial schemas in many languages might well never yield a case in which it does figure as an essential element and so would be kept off the inventory.

This methodology affords a kind of existence proof: It can demonstrate that some element does occur in the universally available inventory of structural spatial elements since it can be seen to occur in at least one closed-class spatial schema, in at least one language. The procedure is repeated numerous times across many languages to build up a sizable inventory of elements essential to spatial schemas.

The next step is to discern whether the uncovered elements comprise particular structural categories and, if so, to determine what these categories are. It can be observed that for certain sets of elements, the elements in a set are mutually incompatible—only one of them can apply at a time at some point in a schema. Such sets are here taken to be basic spatial categories. Along with their members, such categories are also part of language's fundamental conceptual structuring system for space. A representative sample of these categories is presented next.

It will be seen that these categories generally have a relatively small membership. This finding depends in part on the following methodological principles. An element proposed for the inventory should be as coarse-grained as possible—that is, no more specific than is warranted by cross-schema analysis. Correlatively, in establishing a category, care must be taken that it include only the most generic elements that have actually been determined—that is, that its membership have no finer granularity than is warranted by the element-abstraction procedure. For example, the principle of mutual incompatibility yields a spatial category of "relative orientation" between two lines or planes, a category with perhaps only two member elements (both already seen in the *across* schema): approximately parallel and approximately perpendicular. Some evidence additionally suggests an intermediary 'oblique' element as a third member of the category. Thus, some English speakers may distinguish a more perpendicular sense from a more oblique sense, respectively, for the two verb-satellites *out and off*, as in *A secondary pipe branches out / off from the main sewer line*. In any case, though, the category would have no more than these two or three members. Although finer degrees of relative orientation can be distinguished by other cognitive systems, say, in visual perception and in motor control, the conceptual structuring subsystem of language does not include anything finer than the two- or three-way distinction. The procedures of schema analysis and cross-schema comparison, together with the methodological principles of maximum granularity for elements and for category membership, can lead to a determination of the number of structurally distinguished elements ever used in language for a spatial category.

2.3 Sample categories and their member elements

The fundamental categories of spatial structure in the closed-class subsystem of spoken language fall into three classes according to the aspect of a spatial scene they pertain to: the segmentation of the scene into individual components, the properties of an individual component, and the relations of one such component to another. In a fourth class are categories of non-geometric elements frequently found in association with spatial schemas. A sampling of categories and their member elements from each of these four classes is presented next. Category names are enclosed by quotation marks here and throughout. The examples provided here are primarily drawn from English but can be readily multiplied across a diverse range of languages (see Talmy 2000a, chap. 3).

2.3.1 Categories pertaining to scene segmentation

The class designated as scene segmentation may include only one category, that of "major components of a scene", and this category may contain only three member elements: the Figure, the Ground, and a secondary Reference Object. Figure and Ground were already seen for the *across* schema. Schema comparison shows the need to recognize a third scene component, the Secondary Reference Object—in fact, two forms of it: encompassive of or external to the Figure and Ground. The English preposition *near*, as in *The lamp is near the TV* specifies the location of the Figure (the lamp) only with respect to the Ground (the TV). But localizing the Figure with the preposition *above*, as in *The lamp is above the TV*, requires knowledge not only of where the Ground object is, but also of the encompassive earth-based spatial grid, in particular, of its vertical orientation. Thus, *above* requires recognizing three components within a spatial scene, a Figure, a Ground, and a Secondary Reference Object of the encompassive type. Comparably, the schema of *past* in *John is past the border* only relates John as Figure to the border as Ground. One could say this sentence on viewing the event through binoculars from either side of the border. But *John is beyond the border* can be said only by someone on the side of the border opposite John, hence the *beyond* schema establishes a perspective point at that location as a secondary Reference Object—in this case, of the external type.

2.3.2 Categories pertaining to an individual scene component

A number of categories pertain to the characteristics of an individual spatial scene component. This is usually one of the three major components resulting from scene segmentation—the Figure, Ground, or Secondary Reference Object—but it could be others, such as the path line formed by a moving Figure. One such category is that of "dimension" with four member elements: zero dimensions for a point, one for a line, two for a plane, and three for a volume. Some English prepositions require a Ground object schematizable for only one of the four dimensional possibilities. Thus, the schema of the preposition *near* as in *near the dot* requires only that the Ground object be schematizable as a point. *Along*, as in *along the trail*, requires that the Ground object be linear. *Over* as in *a tapestry over a wall* requires a planar Ground. And *throughout*, as in *cherries throughout the jello*, requires a volumetric Ground.

A second category is that of "number" with perhaps four members: one, two, several, and many. Some English prepositions require a Ground comprising objects in one or another of these numbers. Thus, *near* requires a Ground consisting of just one object, *between* of two objects, *among* of several objects, and *amidst* of numerous objects, as in

The basketball lay near the boulder / between the boulders / among the boulders / amidst the cornstalks. The category of number appears to lack any further members—that is, closed-class spatial schemas in languages around the world seem never to incorporate any other number specifications—such as 'three' or 'even-numbered' or 'too many'.

A third category is that of "motive state", with two members: motion and stationariness. Several English prepositions mark this distinction for the Figure. Thus, in one of its senses, *at* requires a stationary Figure, as in *I stayed / *went at the library*, while *into* requires a moving Figure, as in *I went / *stayed into the library*. Other prepositions mark this same distinction for the Ground object (in conjunction with a moving Figure). Thus, *up to* requires a stationary Ground (here, the deer), as in *The lion ran up to the deer*, while *after* requires a moving Ground as in *The lion ran after the deer*. Apparently no spatial schemas mark such additional distinctions as motion at a fast vs. slow rate, or being located at rest vs. remaining located fixedly.

A fourth category is that of "state of boundedness" with two members: bounded and unbounded. The English preposition *along* requires that the path of a moving Figure be unbounded, as shown by its compatibility with a temporal phrase in *for* but not *in*, as in *I walked along the pier for 10 minutes / *in 20 minutes*. But the spatial locution *the length of* requires a bounded path, as in *I walked the length of the pier in 20 Minutes / *for 10 minutes*.[3] While some spatial schemas have the bounded element at one end of a line and the unbounded element at the other end, apparently no spatial schema marks any distinctions other than the two cited states of boundedness. For example, there is no cline of gradually increasing boundedness, nor a gradient transition, although just such a "clinal boundary" appears elsewhere in our cognition, as in geographic perception or conception, e.g., in the gradient demarcation between full forest and full meadowland (Mark and Smith, under review).

Continuing the sampling of this class, a fifth category is that of "directedness" with two members: basic and reversed. A schema can require one or the other of these elements for an encompassive Ground object, as seen for the English prepositions in *The axon grew along / against the chemical gradient*, or for the Atsugewi verb-satellites for (moving) 'downstream' and 'upstream'. Or it can require one of the member elements for an encompassive Secondary Reference Object (here, the line), as in *Mary is ahead of / behind John in line*.

A sixth category is "type of geometry" with two members: rectilinear and radial. This category can apply to an encompassive Secondary Reference Object to yield

[3] As it happens, most motion prepositions in English have a polysemous range that covers both the unbounded and the bounded sense. Thus, *through* as in *I walked through the tunnel for 10 minutes* refers to traversing an unbounded portion of the tunnel's length, whereas in *I walked through the tunnel in 20 minutes*, it refers to traversing the entire bounded length.

reference frames of the two geometric types. Thus, in a subtle effect, the English verb-satellite *away*, as in *The boat drifted further and further away / out from the island*, tends to suggest a rectilinear reference frame in which one might picture the boat moving rightward along a corridor or sea lane with the island on the left (as if along the x-axis of a Cartesian grid). But *out* tends to suggest a radial reference frame in which the boat is seen moving from a center point along a radius through a continuum of concentric circles. In the type-of-geometry category, the radial-geometry member can involve motion about a center, along a radius, or along a periphery. The first of these is the basis for a further category, that of "orientation of spin axis", with two members: vertical and horizontal. The English verb-satellites *around* and *over* specify motion of the Figure about a vertical or horizontal spin axis, respectively, as in *The pole spun around / toppled over* and in *I turned the pail around / over*.

An eighth category is "phase of matter", with three main members, solid, liquid, and empty space, and perhaps a fourth member, fire. Thus, among the dozen or so Atsugewi verb-satellites that subdivide the semantic range of English *into* plus a Ground object, the suffix *-ik's* specifies motion horizontally into solid matter (as chopping an ax into a tree trunk), *-ic't* specifies motion into liquid, *-ipsnu* specifies motion into the empty space of a volumetric enclosure, and *-caw* specifies motion into a fire. The phase of matter category even figures in some English prepositions, albeit covertly. Thus, *in* can apply to a Ground object of any phase of matter, whereas *inside* can apply only to one with empty space, as seen in *The rock is in / inside the box; in / *inside the ground; in / *inside the puddle of water; in / *inside the fire*.

A final category in this sampled series is that of "state of consolidation" with apparently two members: compact (precisional) and diffuse (approximative). The English locative prepositions *at* and *around* distinguish these two concepts, respectively, for the area surrounding a Ground object, as in *The other hiker will be waiting for you at / around the landmark*. The two deictic adverbs in *The hiker will be waiting for you there / thereabouts* mark the same distinction (unless *there* is better considered neutral to the distinction). And in Malagasy (Imai 2002), two locative adverbs for 'here' mark this distinction, with *eto* for 'here within this bounded region', typically indicated with a pointing finger, and *ety* for 'here spread over this unbounded region', typically indicated with a sweep of the hand. In addition to this sampling, some ten or so further categories pertaining to properties of an individual schema component, each category with a small number of fixed contrasts, can be readily identified.

2.3.3 Categories pertaining to the relation of one scene component to another

Another class of categories pertains to the relations that one scene component can bear to another. One such category was described earlier, that of "relative orientation", with two or three members: parallel, perpendicular, and perhaps oblique. A second such category is that of "degree of remove" of one scene component from another. This category appears to have four or five members, two with contact between the components—coincidence and adjacency—and two or three without contact—proximal, perhaps medial, and distal remove. Some pairwise contrasts in English reveal one or another of these member elements for a Figure relating to a Ground. Thus, the locution *in the front of*, as in *The carousel is in the front of the fairground*, expresses coincidence, since the carousel as Figure is represented as being located in a *part* of the fairground as Ground. But *in front of* (without a *the*) as in *The carousel is in front of the fairground*, indicates proximality, since the carousel is now located outside the fairground and near but not touching it. The distinction between proximal and distal can be teased out by noting that *in front of* can only represent a proximal but not a distal degree of remove, as seen in the fact that one can say *The carousel is 20 feet in front of the fairground*, but not, **The carousel is 20 miles in front of the fairground*, whereas *above* allows both proximal and distal degrees of remove, as seen in *The hawk is 1 foot / 1 mile above the table*. The distinction between adjacency and proximality is shown by the prepositions *on* and *over*, as in *The fly is on / over the table*. Need for a fifth category member of 'medial degree of remove' might come from languages with a 'here / there / yonder' kind of distinction in their deictic adverbs or demonstratives.

A third category in this series is that of "degree of dispersion" with two members: sparse and dense. To begin with, English can represent a set of multiple Figures, say, Ø-dimensional peas, as adjacent to or coincident with a 1-, 2-, or 3-dimensional Ground, say, with a knife, a tabletop, or aspic, in a way neutral to the presence or absence of dispersion, as in *There are peas on the knife; on the table; in the aspic*. But in representing dispersion as present, English can (or must) indicate its degree. Thus, a sparse degree of dispersion is indicated by the addition of the locution *here and there*, optionally together with certain preposition shifts, as in *There are peas here and there on / along the knife; on / over the table; in the aspic*. And for a dense degree of dispersion, English has the three specialized forms *all along, all over* and *throughout*, as seen in *There are peas all along the knife; all over the table; throughout the aspic*.

A fourth category is that of "path contour" with perhaps some four members: straight, arced, circular, and meandering. Some English prepositions require one or another of these contour elements for the path of a Figure moving relative to a Ground. Thus, *across* indicates a straight path, as seen in *I drove across the plateau / *hill*, while *over*—in its usage referring to a single path line—indicates an arced contour, as in *I drove over the hill / *plateau*. In one of its senses, *around* indicates a roughly circular path, as in *I walked around the maypole*, and *about* indicates a meandering contour, as in *I walked about the town*. Some ten or so additional categories for relating one scene component to another, again each with its own small number of member contrasts, can be readily identified.

2.3.4 Non-geometric categories

All the preceding elements and their categories have broadly involved geometric characteristics of spatial scenes or the objects within them—that is, they have been genuinely spatial. But a number of non-geometric elements are recurrently found in association with otherwise geometric schemas. One category of such elements is that of "force dynamics" (see Talmy 2000a, chap. 7) with two members: present and absent. Thus, geometrically, the English prepositions *on* and *against* both represent a Figure in adjacent contact with a Ground, but in addition, *on* indicates that the Figure is supported against the pull of gravity through that contact, while *against* indicates that it is not, as seen in *The poster is on / *against the wall* and *The floating helium balloon is against / *on the wall*. Cutting the conceptualization of force somewhat differently (Melissa Bowerman, personal communication), the preposition *op* in the Dutch of the Netherlands indicates a Figure supported comfortably in a natural rest state through its contact with a Ground, whereas *aan* indicates that the Figure is being actively maintained against gravity through contact with the Ground, so that flesh is said to be "op" the bones of a live person but "aan" the bones of a dead person.

A second non-geometric category is that of "accompanying cognitive/affective state", though its extent of membership is not clear. One recurrent member, however, is the attitude toward something that it is unknown, mysterious, or risky. Perhaps in combination with elements of inaccessibility or non-visibility, this category member is associated with the Figure's location in the otherwise spatial indications of the English preposition *beyond*, whereas it is absent from the parallel locution *on the other side of*, as in *He is beyond / on the other side of the border* (both these locutions—unlike *past* seen above—are otherwise equivalent in establishing a viewpoint location as an external Secondary Reference Object).

A third non-geometric category—in the class that relates one scene component to another—is that of "relative priority", with two members: co-equal and main/ancillary. The English verb-satellites *together* and *along* both indicate joint participation, as seen in *I jog together / along with him*. But *together* indicates that the Figure and the Ground are co-equal partners in the activity, whereas *along* indicates that the Figure entity is ancillary to the Ground entity, who would be assumed to engage in the activity even if alone (see 2000b, chap. 3).

2.4 Properties of the inventory

By our methodology, the universally available inventory of structural spatial elements includes all elements that appear in at least one closed-class spatial schema in at least one language. These elements may indeed be equivalent in their sheer availability for use in schemas. But beyond that, they appear to differ in their frequency of occurrence across schemas and languages, ranging from very common to very rare. Accordingly, the inventory of elements—and perhaps also that of categories—may have the property of being hierarchical, with entries running from most- to least-frequent. Such a hierarchy suggests asking whether the elements in the inventory, the categories in the inventory, and the elements in each category form fully closed memberships. That is, does the hierarchy end at a sharp lower boundary or trail off indefinitely? With many schemas and languages already examined, our sampling method may have yielded all the commoner elements and categories, but as the process slows down in the discovery of the rarer forms, will it asymptotically approach some complete constituency and distinctional limit in the inventory, or will it be able to go on uncovering sporadic novel forms as they develop in the course of language change?

The latter seems likelier. Exotic elements with perhaps unique occurrence in one or a few schemas in just one language can be noted, including in English. Thus, in referring to location at the interior of a wholly or partly enclosed vehicle, the prepositions *in* and *on* distinguish whether the vehicle lacks or possesses a walkway. Thus, one is *in* a car but *on* a bus, *in* a helicopter but *on* a plane, *in* a grain car but *on* a train, and *in* a rowboat but *on* a ship. Further, Fillmore has observed that this *on* also requires that the vehicle be currently in use as transport: *The children were playing in / *on the abandoned bus in the junkyard*. Thus, schema analysis in English reveals the element "(partly) enclosed vehicle with a walkway currently in use as transport". This is surely one of the rarer elements in schemas around the world, and its existence, along with that of various others that can be found, suggests that indefinitely many more of them can sporadically arise.

In addition to being only relatively closed at its hierarchically lower end, the inventory may include some categories whose membership seems not to settle down to a small fixed set. One such category may be that of "intrinsic parts". Frequently encountered are the five member elements 'front', 'side', 'back', 'top', and 'bottom', as found in the English prepositions in *The cat lay before / beside / behind / atop / beneath the TV*. But languages like Mixtec seem to distinguish a rather different set of intrinsic parts in their spatial schemas (Brugmann and Macaulay 1986), while Makah distinguishes many more and finer parts, such as with its verb suffixes for 'at the ankle' and 'at the groin' (Matthew Davidson, personal communication).

Apart from any such fuzzy lower boundary or non-coälescing categories, there does appear to exist a graduated inventory of basic spatial elements and categories that is universally available and, in particular, is relatively closed. Bowerman has raised the main challenge to this notion (see, e.g., Bowerman 1989). She notes, for example, that at the same time that children acquiring English learn its *in/on* distinction, children acquiring Korean learn its distinction between *kkita* ‘put [Figure] in a snug fit with [Ground]’ and *nehta* ‘put [Figure] in a loose fit with [Ground]’ she argues that since the elements ‘snug fit’ and ‘loose fit’ are presumably rare among spatial schemas across languages, they do not come from any preset inventory, one that might plausibly be innate, but rather are learned from the open-ended semantics of the adult language. My reply is that the spatial schemas of genuinely closed-class forms in Korean may well still be built from the proposed inventory elements, and that the forms she cites are actually open-class verbs. Open-class semantics—whether for space or other domains —seems to involve a different cognitive subsystem, drawing from finer discriminations within a broader perceptual/conceptual sphere. The Korean verbs are perhaps learned at the same age as English space-related open-class verbs like *squeeze*. Thus, English-acquiring children probably understand that *squeeze* involves centripetal pressure from encircling or bi-/multilaterally placed Antagonists [typically the arm(s) or hand(s)] against an Agonist that resists the pressure but yields down to some smaller compass where it blocks further pressure, and hence that one can squeeze a teddy bear, a tube of toothpaste, or a rubber ball, but not a piece of string or sheet of paper, juice or sugar or the air, a tabletop or the corner of a building. Thus, Bowerman’s challenge may be directed at the wrong target, leaving the proposed roughly preset inventory of basic spatial building blocks intact.

2.5 Basic elements assembled into whole schemas

The procedure so far has been analytic, starting with the whole spatial schemas expressed by closed-class forms and abstracting from them an inventory of fundamental spatial elements. But the investigation must also include a synthetic procedure: examining the ways in which individual spatial elements are assembled to constitute whole schemas. Something of such an assembly was implicit in the initial discussion of the *across* schema. But an explicit example here can better illustrate this part of the investigation.

Consider the schema represented by the English preposition *past* as in *The ball sailed past my head at exactly 3 p.m.* This schema is built out of the following fundamental spatial elements (from the indicated categories) in the indicated arrangements and relationships: There are two main scene components (members of the “major scene components” category), a Figure and a Ground (here, the ball and my head, respectively). The Figure is schematizable as a Ø-dimensional point (a member element of the "dimension" category). This Figure-point is moving (a member element of the "motive state" category). Hence it forms a one-dimensional line (a member of the "dimension" category). This line constitutes the Figure's "path". The Ground is also schematizable as a Ø-dimensional point (a member of the "dimension" category). There is a point P at a proximal remove (a member of the "degree of remove" category) from the Ground point, forming a 1-dimensional line with it (a member of the "dimension" category). This line is parallel (a member of the "relative orientation" category) to the horizontal plane (a member of the "intrinsic parts" category) of the earth-based grid (a member of the "major scene components" category). The Figure's path is perpendicular (a member of the "relative orientation" category) to this line. The Figure's path is also parallel to the horizontal plane of the earth-based grid. If the Ground object has a front, side, and back (members of the "intrinsic parts" category), then point P is proximal to the side part. A non-boundary point (a member of the "state of boundedness" category) of the Figure's path becomes coincident (a member of the "degree of remove" category) with point P at a certain point in time.

Note that here the Figure's path must be specified as passing through a point proximal to the Ground because if it instead passed through the Ground point, one would switch from the preposition *past* to *into*, as in *The ball sailed into my head*, and if it instead passed through some distal point, one might rather say something like *The ball sailed along some ways away from my head*. And the Figure's path must be specified both as horizontal and as located at the side portion of the Ground, because (for example here) if the ball were either falling vertically or traveling horizontally at my front, one would no longer say that it sailed "past" my head.

The least understood aspect of the present investigation is what well-formedness conditions, if any, may govern the legality of such combinations. As of yet, no obvious principles based, say, on geometric simplicity, symmetry, consistency, or the like, are seen to control the patterns in which basic elements assemble into whole schemas. On the one hand, some seemingly Byzantine combinations—like the schemas seen above for *across* and *past*—occur with some regularity across languages. On the other hand, much simpler combinations seem never to occur as closed-class schemas. For example, one could imagine assembling elements into the following schema: down into a surround that is radially proximal to a center point. One could even invent a preposition *apit* to represent this schema. This could then be used, say, in *I poured water apit my house* to refer to my pouring water down into a nearby hole dug in the field around my house. But such schemas are not found. Similarly, a number of schematic distinctions in, for example, the domain of rotation are regularly marked by signed languages, as seen below, and could readily be represented with the inventory elements available to spoken languages, yet they largely do not occur. It could be argued that the spoken language schemas are simply the spatial structures most often encountered in everyday activity. But that would not explain why the additional sign-language schemas—presumably also reflective of everyday experience—do not show up in spoken languages. Besides, the different sets of spatial schemas found in different spoken languages are diverse enough from each other that arguing on the basis of the determinative force of everyday experience is problematic. Something else is at work, but it is not yet clear what that is.

2.6 Properties and processes applying to whole spatial schemas

It was just seen that selected elements of the inventory are combined in specific arrangements to make up the whole schemas represented by closed-class spatial forms. Each such whole schema is thus a "pre-packaged" bundling together of certain elements in a particular arrangement. Each language has in its lexicon a relatively closed set of such pre-packaged schemas—a set larger than that of its spatial closed-class forms, because of polysemy. A speaker of the language must select among these schemas in depicting a spatial scene. We now observe that such schemas, though composite, have a certain unitary status in their own right, and that certain quite general properties and processes can apply to them. In particular, certain properties and processes allow a schema represented by a closed-class form to generalize to a whole family of schemas. In the case of a generalizing *property*, all the schemas of a family are of equal priority. On the other hand, a generalizing *process* acts on a schema that is somehow basic, and either extends or deforms it to yield non-basic schemas; see Talmy 2000a, chaps. 1 and 3, 2000b, chap. 5. Such properties and processes are perhaps part of the overall spoken-language system so that any language's relatively closed set of spatial closed-class forms and the schemas that they basically represent can be used to match more spatial structures in a wider range of scenes.

Looking first at generalizing properties of spatial schemas, one such property is that they exhibit a topological or topology-like neutrality to certain factors of Euclidean geometry. Thus, they are magnitude neutral, as seen in such facts as that the *across* schema can apply to a situation of any size, as in *The ant crawled across my palm / The bus drove across the country*. Further, they are largely shape-neutral, as seen by such facts as that, while the *through* schema requires that the Figure form a path with linear extent, it lets that line take any contour, as in *I zig-zagged / circled through the woods*. And they are bulk-neutral, as seen by such facts as that the *along* schema requires a linear Ground without constraint on the Ground's radial extension, as in *The caterpillar crawled up along the filament / tree trunk*. Thus, while holding to their specific constraints, schemas can vary freely in other respects and so cover a range of spatial configurations.

Among the generalizing processes that extend schemas, one is that of "extendibility from the prototype", which can actually serve as an alternative interpretation for some forms of neutrality, otherwise just treated under generalizing properties. Thus, in the case of shape, as for the *through* schema above, this schema could alternatively be conceived of as prototypically involving a strait path line for the Figure, one that can then be bent to any contour. And, in the case of bulk, as for the *along* schema above, this schema could be thought prototypically to involve a purely 1-dimensional line that then can be radially inflated.

Another such process is “extendibility in ungoverned dimensions”. By this process, a scene component of dimensionality N in the basic form of a schema can generally be raised in dimensionality to form a line, plane, or volume aligned in a way not conflicting with the schema’s other requirements. To illustrate, it was seen earlier under the “type of geometry” category that the English verb-satellite *out* has a schema involving a point Figure moving along a radius away from a center point through a continuum of concentric circles, as in *The boat sailed further and further out from the island*. This schema, with the Figure idealizable as a point, is the basic form. But the same satellite can be used when this Figure-point is extended to form a 1-dimensional line along a radius, as in *The caravan of boats sailed further and further out from the island*. And the *out* can again be used if the Figure-point were instead extended as a 1-dimensional line forming a concentric circle, as in *A circular ripple spread out from where the pebble fell into the water*. In turn, such a concentric circle could be extended to fill in the interior plane, as in *The oil spread out over the water from where it spilled*. Alternatively, the concentric circle could have been extended in the vertical dimension to form a cylinder, as in *A ring of fire spread out as an advancing wall of flames*. Or again, the circle could have been extended to form a spherical shell, as in *The balloon I blew into slowly puffed out*. And such a shell can be extended to fill in the interior volume, as in *The leavened dough slowly puffed out*. Thus, the same form *out* serves for this series of geometric extensions without any need to switch to some different form.

One more schema-extending process is “extendibility across motive states”. A schema basic for one motive-state-and-Figure geometry can in general be systematically extended to another motive-state-and-Figure geometry. For example, a closed-class form whose most basic schema pertains to a point Figure moving to form a path can generally serve as well to represent the related schema with a stationary linear Figure in the same location as the path. Thus, probably the most basic *across* schema is actually for a moving point Figure, as in *The gopher ran across the road*. By the present process, this schema can extend to the static linear Figure schema first seen in *The board lay across the road*. All the spatial properties uncovered for that static schema hold as well for the present basic dynamic schema, which in fact is the schema in which these properties originally arise.

Among the generalizing processes that deform a schema, one is that of “stretching”, which allows a slight relaxing of one of the normal constraints. Thus, in the *across* schema, where the Ground plane is either a ribbon with a long and short axis or a square with equal axes, a static linear Figure or the path of a moving point Figure must be aligned with the short Ground axis or with one of its equal axes. Accordingly, one can say *I swam across the canal* and *I swam across the square pool* when moving from one side to the other, but one cannot say **I swam across the canal* when moving from one end of the canal to the other. But, by moderately stretching one axis length relative to the other, one might just about be able to say *I swam across the pool* when moving from one end to the other of a slightly oblong pool.

Another schema deforming process is that of "feature cancellation", in which a particular complex of elements in the basic schema is omitted. Thus, the preposition *across* can be used in *The shopping cart rolled across the boulevard and was hit by an oncoming car*, even though one feature of the schema—'terminal point coincides with the distal edge of the Ground ribbon'—is canceled from the Figure's path. Further, both this feature and the feature 'beginning point coincides with the proximal edge of the Ground ribbon' are canceled in *The tumbleweed rolled across the prairie for an hour*. Thus, the spoken language system includes a number of generalizing properties and processes that allow the otherwise relatively closed set of abstracted or basic schemas represented in the lexicon of any single language to be applicable to a much wider range of spatial configurations.

3. Spatial structuring in signed language

All the preceding findings on the linguistic structuring of space have been based on the patterns found in spoken languages. The inquiry into the fundamental concept-structuring system of language leads naturally to investigating its character in another major body of linguistic realization, signed language. The value of extending the inquiry in this way would be to discover whether the spatial structuring system is the same or is different in certain respects across the two language modalities, with either discovery having major consequences for cognitive theory.

In this research extension, a problematic issue is exactly what to compare between spoken and signed language. The two language systems appear to subdivide into somewhat different sets of subsystems. Thus, heuristically, the generalized spoken language system can be thought to consist of an open-class or lexical subsystem (generally representing conceptual content); a closed-class or grammatical subsystem (generally representing conceptual structure); a gradient subsystem of "vocal dynamics" (including loudness, pitch, timbre, rate, distinctness, unit separation); and an accompanying somatic subsystem (including facial expression, gesture, and "body language"). On the other hand, by one provisional proposal, the generalized sign language system might instead divide up into the following: a subsystem of lexical forms (including noun, verb, and adjective signs); an "inflectional" subsystem (including modulations of lexical signs for person, aspect); a subsystem of size-and-shape specifiers (or SASS's; a subsystem of so-called "classifier expressions"; a gestural subsystem (along a gradient of incorporation into the preceding subsystems); a subsystem of face, head, and torso representations; a gradient subsystem of "bodily dynamics" (including amplitude, rate, distinctness, unit separation); and an associated or overlaid somatic subsystem (including further facial expression and "body language"). In particular here, the subsystem of classifier expressions—which is apparently present in all signed languages—is a formally distinct subsystem dedicated solely to the schematic structural representation of objects moving or located with respect to each other in space (see Liddell in press, Emmorey in press). Each classifier expression, perhaps generally corresponding to a clause in spoken language, represents a so-conceived event of motion or location.[4]

The research program of comparing the representation of spatial structure across the two language modalities ultimately requires considering the two whole systems and all their subsystems. But the initial comparison—the one adopted here—should be between those portions of each system most directly involved with the representation of spatial structure. In spoken language, this is that part of the closed-class subsystem that represents spatial structure and, in signed language, it is the subsystem of classifier constructions. Spelled out, the shared properties that make this initial comparison apt include the following. First, of course, both subsystems represent objects relating to each other in space. Second, in terms of the functional distinction between "structure" and "content" described earlier, each of the subsystems is squarely on the structural side. In fact, analogous structure-content contrasts occur. Thus, the English closed-class form *into* represents the concept of a path that begins outside and ends inside an enclosure in terms of schematic structure, in contrast with the open-class verb *enter* that represents the same concept in terms of substantive content (see Talmy 2000a, chap. 1 for this structure-content distinction). Comparably, any of the formations within a classifier expression for such an outside-to-inside path represents it in terms of its schematic structure, in contrast with the unrelated lexical verb sign that can be glossed as 'enter'. Third, in each subsystem, a schematic structural form within an expression can in general be semantically elaborated by a content form that joins or replaces it within the same expression. Thus, in the English sentence *I drove it (—the motorcycle—) in (-to the shed)*

[4] The "classifier" label for this subsystem—originally chosen because its constructions largely include a classifier-like handshape—can be misleading, since it names the whole expression complex for just one of its components. A more apt term might be the "Motion-event subsystem".

the parenthesized forms optionally elaborate on the otherwise schematically represented Figure and Ground. Comparably, in the ASL sentence *(SHED) (MOTORCYCLE) vehicle-move-into-enclosure*, the optionally signed forms within parentheses elaborate on the otherwise schematic Figure and Ground representations within the hyphenated classifier expression.

To illustrate the classifier system, a spatial event that English could express as *The car drove past the tree* could be expressed in ASL as follows: The signer's dominant hand, used to represent the Figure object, here has a "3 handshape" (index and middle fingers extended forward, thumb up) to represent a land vehicle. The non-dominant hand, used to represent the Ground object, here involves an upright "5 handshape" (forearm held upright with the five fingers extended upward and spread apart) to represent a tree. The dominant hand is moved horizontally across the signer's torso and past the non-dominant forearm. Further though, this basic form could be modified or augmented to represent additional particulars of the referent spatial event. Thus, the dominant hand can show additional characteristics of the path. For example, the hand could move along a curved path to indicate that the road being followed was curved, it could slant upward to represent an uphill course, or both could be shown together. The dominant hand can additionally show the manner of the motion. For example, as it moves along, it could oscillate up and down to indicate a bumpy ride, or move quickly to indicate a swift pace, or both could be shown together, as well as with the preceding two-path properties. And the dominant hand can show additional relationships of the Figure to the Ground. For example, it could pass nearer or farther from the non-dominant hand to indicate the car's distance from the tree when passing it; it could make the approach toward the non-dominant hand longer (or shorter) than the trailing portion of the path to represent the comparable relationship between the car's path and the tree; or it could show both of these together or, indeed, with all the preceding additional characteristics.

The essential finding of how signed language differs from spoken language is that it more closely parallels what appear to be the structural characteristics of scene parsing in visual perception. This difference can be observed in two venues, the universally available spatial inventory and the spatial expression. These two venues are discussed next in turn.

3.1 In the inventory

The inventory of forms for representing spatial structure available to the classifier subsystem of signed language has a greater total number of fundamental elements, a greater number of categories, and generally a greater number of elements per category than the spoken-language, closed-class inventory. While many of the categories and their members seem to correspond across the two inventories, the signed-planguage inventory has an additional number of categories and member elements not present in the spoken-language inventory. Comparing the membership of the corresponding categories in terms of discrete elements, the number of basic elements per category in signed language actually exhibits a range: from being the same as that for spoken language to being very much greater. Further, though, while the membership of some categories in signed language may well consist of discrete elements, that of others appears to be gradient. Here, any procedure of tallying some fixed number of discrete elements in a category must give way to determining the approximate fineness of distinctions that can be practicably made for that category. So while some corresponding categories across the two language modalities may otherwise be quite comparable, their memberships can be of different types, discrete vs. analog. Altogether then, given its greater number of categories, generally larger membership per category, and a frequently gradient type of membership, the inventory of forms for building a schematic spatial representation available to the classifier subsystem of signed language is more extensive and finer than for the closed-class subsystem of spoken language. This greater extensiveness and finer granularity of spatial distinctions seems more comparable to that of spatial parsing in visual perception.

The following are some spatial categories in common across the two language modalities, but with increasing disparity in size of membership. First, some categories appear to be quite comparable across the two modalities. Thus, both the closed-class subsystem of spoken language and the classifier subsystem of signed language structurally segment a scene into the same three components, a Figure, a Ground, and a secondary Reference Object. Both subsystems represent the category of dimensionality with the same four members—a point, a line, a plane, and a volume. And both mark the same two degrees of boundedness: bounded and unbounded.

For certain categories, signed language has just a slightly greater membership than does spoken language. Thus, for motive state, signed language structurally represents not only moving and being located, but also remaining fixedly located—a concept that spoken languages typically represent in verbs, but not in their spatial preposition-like forms.

For some other spatial categories, signed language has a moderately greater membership than spoken language. In some of these categories, the membership is probably gradient, but without the capacity to represent many fine distinctions clearly. Thus, signed language can apparently mark moderately more degrees of remove than spoken language's four or five members in this category. It can also apparently distinguish moderately more path lengths than the two—short and long—that spoken language marks structurally (as in English *The bug flew right / way up there*). And while spoken language can mark at most three distinctions of relative orientation —parallel, perpendicular, and oblique—signed language can distinguish a moderately greater number, for example, in the elevation of a path's angle above the horizontal, or in the angle of the Figure's axes to that of the Ground (e.g., in the placement of a rod against a wall).

Finally, there are some categories for which signed language has an indefinitely greater membership than spoken language. Thus, while spoken language structurally distinguishes some four path contours as seen in section 2.3.3, signed language can represent perhaps indefinitely many more, including zigzags, spirals, and ricochets. And for the category "locus within referent space", spoken language can structurally distinguish perhaps at most three loci relative to the speaker's location—'here', 'there', and 'yonder'—whereas sign language can distinguish indefinitely many more within sign space.

Apart from membership differences across common categories, signed language represents some categories not found in spoken language. One such category is the relative lengths of a Figure's path before and after encounter with the Ground. Or again, signed language can represent not only the category of "degree of dispersion" (which spoken language was seen to represent in section 2.3.3), but also the category "pattern of distribution". Thus, in representing multiple Figure objects dispersed over a planar surface, it could in addition structurally indicate that these Figure objects are linear (as with dry spaghetti over a table) and are arrayed in parallel alignment, crisscrossing, or in a jumble.

This difference in the number of structurally marked, spatial category and element distinctions between spoken and signed language can be highlighted with a closer analysis of a single spatial domain, that of rotational motion. As seen earlier, the closed-class subsystem in spoken language basically represents only one category within this domain, that of "orientation of spin axis", and within this category distinguishes only two member elements, vertical and horizontal. These two member elements are expressed, for example, by the English verb-satellites *around* and *over* as in *The pole spun around / toppled over*. ASL, by contrast, distinguishes more degrees of spin axis orientation and, in addition, marks several further categories within the domain of rotation. Thus, it represents the category of "amount of rotation" and within this category can readily distinguish, say, whether the arc of a Figure's path is less than, exactly, more than, or many times one full circuit. These are differences that English might offer for inference only from the time signature, as in *I ran around the house for 20 seconds / in one minute / for two minutes / for hours*, while using the same single spatial form *around* for all these cases. Further, while English would continue using just *around* and *over*, ASL further represents the category of "relation of the spin axis to an object's geometry" and marks many distinctions within this category. Thus, it can structurally mark the spin axis as being located at the center of the turning object—as well as whether this object is planar like a CD disk, linear like a propeller, or an aligned cylinder like a pencil spinning on its point. It distinguishes this from the spin axis located at the boundary of the object—as well as whether the object is linear like the "hammer" swung around in a hammer toss, a transverse plane like a swinging gate, or a parallel plane like a swung cape. And it further distinguishes these from the spin axis located at a point external to the object —as well as whether the object is point-like, like the earth around the sun, or linear, like a spinning hoop. Finally, ASL can structurally represent the category of "uniformity of rotation" with its two member elements, uniform and non-uniform, where English could mark this distinction only with an open-class form, like the verbs in *The hanging rope spun / twisted around*, while once again continuing with the same single structural, closed-class form *around*. Thus, while spoken language structurally marks only a minimal distinction of spin-axis orientation throughout all these geometrically distinct forms of rotation, signed language marks more categories as well as finer distinctions within them, and a number of these appear to be distinguished as well by visual parsing of rotational movement.

To expand on the issue of gradience, numerous spatial categories in the classifier subsystem of signed language—for example, many of the 30 spatial categories listed in section 3.2.3.1—are gradient in character. Spoken language has a bit of this, as where the vowel length in the English word *way* can be varied continuously.[5] But the preponderant norm is the use of discrete spatial elements, typically incorporated into distinct morphemes. For example, insofar as they represent degree of remove, the separate forms in the series *on / next to / near / away from* represent increasing distance in what can be considered quantal jumps. That is, the closed-class subsystem of spoken language is a type of cognitive system whose basic organizing principle is that of the recombination of discrete elements (i.e., the basic conceptual elements whose combinations, in turn, comprise the meanings of discrete morphemic forms). By contrast, the classifier subsystem of signed language is the kind of cognitive system whose basic organizing principle largely involves gradience, much as would seem to be the case as well for the visual and motor systems. In fact, within a classifier expression, the gradience of motor control and of visual perception are placed in sync with each other (for the signer and the addressee, respectively), and conjointly put in the service of the linguistic system.

[5] See sec. 3.2.1.3.

While this section provides evidence that the classifier subsystem in signed language diverges from the schematizing of spoken language in the direction of visual parsing, one must further observe that the classifier subsystem is also not "simply" a gestural system wholly iconic with visual perception. Rather, it incorporates much of the discrete, categorial, symbolic, and metaphoric character that is otherwise familiar from the organization of spoken language. Thus, as already seen above, spatial representation in the classifier subsystem does fall into categories, and some of these categories contain only a few discrete members—in fact, several of these are much the same as in spoken language. Second, the handshapes functioning as classifiers for the Figure, manipulator, instrument, or Ground within classifier expressions are themselves discrete (non-gradient) members of a relatively closed set. Third, many of the hand movements in classifier expressions represent particular concepts or metaconcepts and do not mimic actual visible movements of the represented objects. Here is a small sample of this property. After one lowers one's two extended fingers to represent a knife dipping into peanut butter—or all one's extended fingers in a curve to represent a scoop dipping into coffee beans—one curls back the fingertips while moving back up to represent the instrument's "holding" the Figure, even though the instrument in question physically does nothing of the sort. Or again, the free fall of a Figure is represented not only by a downward motion of the dominant hand in its classifier handshape, but also by an accompanying rotation of the hand—whether or not the Figure in fact rotated in just that way during its fall. As another example, a Figure is shown as simply located at a spot in space by the dominant hand in its classifier handshape being placed relaxedly at a spot in signing space, and as remaining fixedly at its spot by the hand's being placed tensely and with a slight final jiggle, even though these two conceptualizations of the temporal character of a Figure's location are visually indistinguishable. Or, further, a (so-conceivedly) random spatial distribution of a mass or multiplex Figure along a line, over a plane, or through a volume is represented by the Figure hand being placed with a loose non-concerted motion, typically three times, at uneven spacings within the relevant n-dimensional area, even though that particular spacing of three exemplars may not correspond to the actual visible distribution. And finally, a classifier hand's type of movement can indicate whether this movement represents the actual path of the Figure, or is to be discounted. Thus, the two flat hands held with palms toward the signer, fingertips joined, can be moved steadily away to represent a wall's being slid progressively outward (as to expand a room), or instead can be moved in a quick up-and-down arc to a point further away to represent a wall relocated to a further spot, whatever its path from the starting location. That is, the latter quick arc movement represents a meta-concept: that the path followed by the hands does not represent the Figure's actual path and is to be disregarded from calculations of iconicity. All in all, then, the classifier subsystem presents itself as a genuine linguistic system, but one having more extensive homology with the visual structuring system than spoken language has.

3.2 In the expression

The second venue, that of any single spatial expression, exhibits further respects in which signed language differs from spoken language in the apparent direction of visual scene parsing. Several of these are outlined next.

3.2.1 Iconic representation in the expression

Spatial representation in signed classifier expressions is iconic with scene parsing in visual perception in at least the following four respects.

3.2.1.1 Iconic clustering of elements and categories

The structural elements of a scene of motion are clustered together in the classifier subsystem's representation of them in signed language more as they seem to be clustered in perception. When one views a motion event, such as a car driving bumpily along a curve past a tree, it is perceptually the same single object, the car, that exhibits all of the following characteristics: It has certain object properties as a Figure; it moves; it has a manner of motion; it describes a path of a particular contour; and it relates to other surrounding objects (the Ground) in its path of motion. The Ground object or objects are perceived as separate. Correspondingly, the classifier subsystem maintains exactly this pattern of clustering. It is the same single hand, the dominant hand, that exhibits the Figure characteristics, motion, manner, path contour, and relations to a Ground object. The other hand, the non-dominant, separately represents the Ground object.

All spoken languages diverge to a greater or less extent from this visual fidelity. Thus, consider one English counterpart of the event, the sentence *The car bumped along past the tree*. Here, the subject nominal, *the car*, separately represents the Figure object by itself. The verb complex clusters together the representations of the verb and the satellite: The verb *bumped* represents both the fact of motion and the manner of motion together, while its sister constituent, the satellite *along* represents the presence of a path of translational motion. The prepositional phrase clusters together the preposition *past*, representing the path conformation, and its sister constituent, the nominal *the tree*, representing the Ground object. In fact, it remains a mystery at this point in the investigation as to why all spoken languages using a preposition-like constituent to indicate path always conjoin it with the Ground nominal and basically never with the

Figure nominal,[6] even though the Figure is what executes the path, and is so represented in the classifier construction of signed language.

3.2.1.2 Iconic representation of object vs. action

The classifier subsystem of signed language appears to be iconic with visual parsing not only in its clustering of spatial elements and categories, as just seen, but largely also in its representation of them. For example, it marks one basic category opposition, that between an entity and its activity, by using an object like the hand to represent an object, and motion of the hand to represent motion of the object. More specifically, the hand or other body part represents a structural entity (such as the Figure)—with the body part's configuration representing the identity or other properties of the entity—while movements or positionings of the body part represent properties of the entity's motion, location, or orientation. For example, the hand could be shaped flat to represent a planar object (e.g., a sheet of paper), or rounded to represent a cup-shaped object. And, as seen, any such handshape as Figure could be moved along a variety of trajectories that represent particular path contours.

But an alternative to this arrangement could be imagined. The handshape could represent the path of a Figure—e.g., a fist to represent a stationary location, the outstretched fingers held flat together to represent a straight line path, the fingers in a curved plane for a curved path, and the fingers alternately forward and backward for a zigzag path. Meanwhile, the hand movement could represent the Figure's shape—e.g., the hand moving in a circle to represent a round Figure and in a straight line for a linear Figure. However, no such mapping of referents to their representations is found.[7] Rather, the mapping in signed language is visually iconic: It assigns the representation of a material object in a scene to a material object in a classifier complex, for example, the hand, and the representation of the movements of that object in the scene to the movements of the hand.

No such iconic correspondence is found in spoken language. Thus, while material objects are prototypically expressed by nouns in English, they are instead prototypically represented by verb roots in Atsugewi (see Talmy 2000b, chap. 1). And while path

---

[6] As the only apparent exception, a "demoted Figure" (see Talmy 2000b, chap. 1) can acquire either of two "demotion particles"—e.g., English *with* and *of*—that mark whether the Figure's path had a "TO" or a "FROM" vector, as seen in *The fuel tank slowly filled with gas / drained of its gas*.

[7] The size and shape specifiers (SASS's) in signed languages do permit movement of the hands to trace out an object's contours, but the hands cannot at the same time adopt a shape representing the object's path.

configurations are prototypically represented in Spanish by verbs, this is done by prepositions and satellites in English.

3.2.1.3 Iconic representation of further particular categories

Finer forms of iconicity are also found within each branch of the broad opposition of entity/activity. In fact, most of the spatial categories listed in section 3.2.3.1 that a classifier expression can represent are largely iconic with visual parsing. Thus, an entity's form is often represented by the form of the hand(s), its size by the compass of the hand(s), and its number by the number of digits or hands extended. And, among many other categories in the list, an entity's motive state, path contour, path length, manner of motion, and rate of motion are separately represented by corresponding behaviors of the hand(s).

Spoken language, again, has only a bit of comparable iconicity. As examples, path length can be iconically represented in English by the vowel length of *way*, as in *The bird flew waay / waaaay / waaaaaay up there*. Path length can also be semi-iconically represented by the number of iterations, as in *The bird flew up / up up / up up up and away*. Perhaps the number of an entity can be represented in some spoken language by a closed-class reduplication. But the great majority of spoken closed-class representations show no such iconicity.

3.2.1.4 Iconic representation of the temporal progression of a trajectory

The classifier subsystem is also iconic with visual parsing in its representation of temporal progression, specifically, that of a Figure's path trajectory. For example, when an ASL classifier expression represents "The car drove past the tree", the "past" path is shown by the Figure hand progressing from the nearer side of the Ground arm to a point beside it and then on to its further side, much like the path progression one would see on viewing an actual car passing a tree. By contrast, nothing in any single closed-class path morpheme in a spoken language corresponds to such a progression. Thus, the *past* in *The car drove past the tree* is structurally a single indivisible linguistic unit, a morpheme, whose form represents no motion ahead in space. Iconicity of this sort can appear in spoken language only where a complex path is treated as a sequence of subparts, each with its own morphemic representation, as in *I reached my hand down around behind the clothes hamper to get the vacuum cleaner*.

3.2.2 A narrow time-space aperture in the expression

Another way that the classifier expression in signed language may be more like visual perception is that it appears to be largely limited to representing a narrow time-space aperture. The tentative principle is that a classifier complex readily represents what would appear within a narrow scope of space and time if one were to zoom in with one's scope of perception around a Figure object, but little outside that narrowed scope. Hence, a classifier expression readily represents the Figure object as to its shape or type, any manipulator or instrument immediately adjacent to the Figure, the Figure's current state of Motion (motion or locatedness), the contour or direction of a moving Figure's path, and any Manner exhibited by the Figure as it moves. However, a classifier expression can little represent related factors occurring outside the current time, such as a prior cause or a follow-up consequence. And it can little represent even concurrent factors if they lie outside the immediate spatial ambit of the Figure, factors like the ongoing causal activity of an intentional Agent or other external instrumentality.

By contrast, spoken languages can largely represent such non-local spatio-temporal factors within a single clause. In particular, such representation occurs readily in satellite-framed languages such as English (see Talmy 2000b, chaps. 1 and 3). In representing a Motion event, this type of language regularly employs the satellite constituent (e.g., the verb particle in English) to represent the Path, and the main verb to represent a "co-event". The co-event is ancillary to the main Motion event and relates to it as its precursor, enabler, cause, manner, concomitant, consequence, or the like.

Satellite-framed languages can certainly use this format to represent within-aperture situations that can also be represented by a classifier complex. Thus, English can say within a single clause—and ASL can sign within a single classifier expression —a motion event in which the Figure is moved by an adjacent manipulator, as in *I pinched some moss up off the rock* and *I pulled the pitcher along the counter*, or in which the Figure is moved by an adjacent instrument, as in *I scooped jelly beans up into the bag*. The same holds for a situation in which a moving Figure exhibits a concurrent Manner, as in *The cork bobbed past the seaweed*.

But English can go on to use this same one-clause format to include the representation of co-events outside the aperture, either temporally or spatially. Thus, temporally, English can include the representation of a prior causal event, as in *I kicked the football over the goalpost* (first I kicked the ball, then it moved over the goalpost). And it can represent a subsequent event, as in *They locked the prisoner into his cell* (first they put him in, then they locked it). But ASL cannot represent such temporally extended event complexes within a single classifier expression. Thus, it can represent the former sentence with a succession of two classifier expressions: first, flicking the middle finger of the dominant hand across the other hand's upturned palm to represent the component event of kicking an object, and next moving the extended index finger of the dominant hand axially along a line through the space formed by the up-pointing index and little fingers of the non-dominant hand, representing the component event of the ball's passing over the goalpost. But it cannot represent the whole event complex within a single expression—say, by flicking one's middle finger against the other hand whose extended index finger then moves off axially along a line.

Further, English can use the same single-clause format to represent events with spatial scope beyond a narrow aperture, for example, an Agent's concurrent causal activity outside any direct manipulation of the Figure, as in *I walked / ran / drove / flew the memo to the home office*. Again, ASL cannot represent the whole event complex of, say, *I ran the memo to the home office* within a single classifier expression. Thus, it could not, say, adopt the classifier for holding a thin, flat object (thumb pressed against flat fingers) with the dominant hand and placing this atop the non-dominant hand while moving forward with it as it shows alternating strokes of two downward pointed fingers to indicate running (or concurrently with any other indication of running). Instead, a sequence of two expressions would likely be used; for example, first one for taking a memo, then one for a person speeding along.[8]

Although the unacceptable examples above have been contrived, they nevertheless show that it is physically feasible for a signed language to represent factors related to the Figure's Motion outside its immediate space-time ambit. Accordingly, the fact that signed languages, unlike spoken languages, do avoid such representations may follow from deeper structural causes, such as a greater fidelity to the characteristics of visual perception.

However apt, though, such an account leaves some facts still needing explanation. Thus, on the one hand, it makes sense that the aperture of a classifier expression is limited temporally to the present moment—this accords with our usual understanding of visual perception. But it is not clear why the aperture is also limited spatially. Visual perception is limited spatially to a narrow scope only when attention is being focused, but is otherwise able to process a wide-scoped array. Why then should classifier expressions avoid such wide spatial scope as well? Further, sign languages *can* include representation of the Ground object within a single classifier expression (typically with the non-dominant hand), even where that object is not adjacent to the Figure.

---

[8] The behavior here of ASL cannot be explained away on the grounds that it is simply structured like a verb-framed language, since such spoken languages typically *can* represent concurrent Manner outside a narrow aperture, in effect saying something like: *I walking / running / driving / flying carried the memo to the home office*.

3.2.3 More independent distinctions representable in the expression

This third property of classifier expressions has two related aspects—the large number of different elements and categories that can be represented together, and their independent variability—and these are next treated in turn.

3.2.3.1 Many more elements / categories representable within a single expression

Although the spatio-temporal aperture that can be represented within a single classifier expression may be small compared to that in a spoken-language clause, the number of distinct factors within that aperture that can be represented is enormously greater. In fact, perhaps the most striking difference between the signed and the spoken representation of space in the expression is that the classifier system in signed language permits the representation of a vastly greater number of distinct spatial categories simultaneously and independently. A spoken language like English can separately represent only up to four or five different spatial categories with closed-class forms in a single clause. As illustrated in the sentence *The bat flew way back up into its niche in the cavern*, the verb is followed in turn by: a slot for indication of path length (with three members: “zero” for ‘neutral’, *way* for ‘relatively long’, *right* for ‘relatively short’); a slot for state of return (with two members: “zero” for ‘neutral’, *back* for ‘return’); a slot for displacement within the earth-frame (with four members: “zero” for ‘neutral’, *up* for ‘positive vertical displacement’, *down* for ‘negative vertical displacement’, *over* for ‘horizontal displacement’); a slot for geometric conformation (with many members, including *in, across, past*); and perhaps a slot for motive state and vector (with two members: “zero” for ‘neutral between location AT and motion TO’ as seen in *in / on*, and *-to* for ‘motion TO’ as seen in *into / onto*). Even a polysynthetic language like Atsugewi has closed-class slots within a single clause for only up to six spatial categories: path conformation combined with Ground type, path length, vector, deixis, state of return, and cause or manner. In contrast, by one tentative count, ASL has provision for the separate indication of thirty different spatial categories. These categories do exhibit certain coöccurrence restrictions, they differ in obligatoriness or optionality, and it is unlikely—perhaps impossible—for all thirty of them to be represented at once. Nevertheless, a sizable number of them can be represented in a single classifier expression and varied independently there. The table below lists the spatial categories that I have provisionally identified as available for concurrent independent representation. The guiding principle for positing a category has been that its elements are mutually exclusive: different elements in the same category cannot be represented together in the same classifier expression. If certain elements can be concurrently represented, they belong to different categories. Following this principle has, on the one hand, involved joining together what some sign language analyses have treated as separate factors. For example, the first category below covers equally the representation of Figure, instrument, or manipulator (handling classifier), since these three kinds of elements apparently cannot be separately represented in a single expression—one or another of them must be selected. On the other hand, the principle requires making distinctions within some categories that spoken languages treat as uniform. Thus, the single "manner" category of English must be subdivided into a category of "divertive manner" (e.g., moving along with an up-down bump) and a category of "dynamic manner" (e.g., moving along rapidly) because these two factors can be represented concurrently and varied independently.

A. Entity properties
   1. identity (form or semantic category) of Figure / instrument / manipulator
   2. identity (form or semantic category) of Ground
   3. magnitude of some major entity dimension
   4. magnitude of a transverse dimension
   5. number of entities
B. Orientation properties
   6. an entity's rotatedness about its left-right axis ("pitch")
   7. an entity's rotatedness about its front-back axis ("roll")
   8. a. an entity's rotatedness about its top-bottom axis ("yaw")
      b. an entity's rotatedness relative to its path of forward motion
C. Locus properties
   9. locus within sign space
D. Motion properties
   10. motive state (moving / resting / fixed)
   11. internal motion (e.g., expansion/contraction, form change, wriggle, swirling)
   12. confined motion (e.g., straight oscillation, rotary oscillation, rotation, local wander)
   13. translational motion
E. Path properties
   14. state of continuity (unbroken / saltatory)
   15. contour of path
   16. state of boundedness (bounded / unbounded)
   17. length of path
   18. vertical height
   19. horizontal distance from signer
   20. left-right positioning
   21. up-down angle ("elevation")

22. left-right angle ("direction")
23. transitions between motion and stationariness (e.g., normal, decelerated, abrupt as from impact)

F. Manner properties
24. divertive manner
25. dynamic manner

G. Relations of Figure or Path to Ground
26. path's conformation relative to Ground
27. relative lengths of path before and after encounter with Ground
28. Figure's path relative to the Path of a moving Ground
29. Figure's proximity to Ground
30. Figure's orientation relative to Ground

It seems probable that something more on the order of this number of spatial categories are concurrently analyzed out by visual processing on viewing a scene than the much smaller number present in even the most extreme spoken language patterns.

3.2.3.2 Elements / categories independently variable in the expression—Not in pre-packaged schemas

The signed-spoken language difference just presented was mainly considered for the sheer number of distinct spatial categories that can be represented together in a single classifier expression. Now, though, we stress the corollary: their independent variability. That is, apart from certain constraints involving coöccurrence and obligatoriness in a classifier expression, a signer can generally select a category for inclusion independently of other categories, and select a member element within each category independently of other selections. For example, a classifier expression can separately include and independently vary a path's contour, length, vertical angle, horizontal angle, speed, accompanying manner, and relation to Ground object.

By contrast, it was seen earlier that spoken languages largely bundle together a choice of spatial member elements within a selection of spatial categories for representation within the single complex schema that is associated with a closed-class morpheme. The lexicon of each spoken language will have available a certain number of such "pre-packaged" spatial schemas, and the speaker must generally choose from among those to represent a spatial scene, even where the fit is not exact. The system of generalizing properties and processes seen in section 2.6 that apply to the set of basic schemas in the lexicon (including their plastic extension and deformation) may exist to compensate for the pre-packaging and closed stock of the schemas in any spoken language. Thus, what are largely semantic components within a single morpheme in spoken language correspond to what can be considered separate individually controllable morphemes in the signed classifier expression.

The apparent general lack in classifier expressions of pre-packaging, of a fixed set of discrete basic schemas, or of a system for generalizing, extending, or deforming such basic schemas may well accord with comparable characteristics of visual parsing. That is, the visual processing of a viewed scene may tend toward the independent assessment of spatial factors without much prepacketing of associated factors or of their plastic alteration. If shown to be the case, then signed language will once again prove to be closer to perceptual spatial structuring than spoken language is.

4. Cognitive and neural implications of spoken / signed language differences

The preceding comparison of the space-structuring subsystems of spoken and of signed language has shown a number of respects in which these are similar and in which they are different. It can be theorized that their common characteristics are the product of a single neural system, what can be assumed to be the core language system, while each set of distinct characteristics results from the activity of some further distinct neural system. These ideas are outlined next.

4.1 Where signed and spoken language are alike

We can first summarize and partly extend the properties above found to hold both in the closed-class subsystem of spoken language and in the classifier subsystem of signed language. Both subsystems can represent multifarious and subtly distinct spatial situations—that is, situations of objects moving or located with respect to each other in space. Both represent such spatial situations schematically and structurally. Both have basic elements that in combination make up the structural schematizations. Both group their basic elements within certain categories that themselves represent particular categories of spatial structure. Both have certain conditions on the combination of basic elements and categories into a full structural schematization. Both have conditions on the coöccurrence and sequencing of such schematizations within a larger spatial expression. Both permit semantic amplification of certain elements or parts of a schematization by open-class or lexical forms outside the schema. And in both subsystems, a spatial situation can often be conceptualized in more than one way, so that it is amenable to alternative schematizations.

4.2 Where spoken and signed language differ

Beside the preceding commonalities, though, the two language modalities have been seen to differ in a number of respects. First, they appear to divide up into somewhat different sets of subsystems without clear one-to-one match-ups. Accordingly, the spatial portion of the spoken-language, closed-class subsystem and the classifier subsystem of signed language may not be exactly corresponding counterparts, but only those parts of the two language modalities closest to each other in the representation of schematic spatial structure. Second, within this initial comparison, the classifier subsystem seems closer to the structural characteristics of visual parsing than the closed-class subsystem in all of the following ways: It has more basic elements, categories, and elements per category in its schematic representation of spatial structure. Its category membership exhibits much more gradient representation, in addition to discrete representation. Its elements and categories exhibit more iconicity with the visual in the pattern in which they are clustered in an expression, in their observance of an object/action distinction, in their physical realization, and in their progression through time. It can represent only a narrow temporal aperture in an expression (and only a narrow spatial aperture as well, though this difference from spoken language might not reflect visual fidelity). It can represent many more distinct elements and categories together in a single expression. It can more readily select categories and category elements independently of each other for representation in an expression. And it avoids pre-packaged, category-element combinations as well as generalizations of their range and processes for their extension or deformation.

4.3 A new neural model

In its strong reading, the Fodor-Chomsky model relevant here is of a complete, inviolate language module in the brain, one that performs all and only the functions of language without influence from outside itself—a specifically linguistic "organ". But the evidence assembled here challenges such a model. What has here been found is that two different linguistic systems, the spoken and the signed, both of them undeniably forms of human language, share extensive similarities but—crucially—also exhibit substantial differences in structure and organization. A new neural model can be proposed that is sensitive to this finding. We can posit a "core" language system in the brain, more limited in scope than the Fodor-Chomsky module, that is responsible for the properties and performs the functions found to be in common across both the spoken and the signed modalities. In representing at least spatial structure, this core system would then further connect with two different, outside brain systems responsible, respectively, for the properties and functions specific to each of the two language modalities. It would thus be the interaction of the core linguistic system with one of the outside systems that would underlie the full functioning of each of the two language modalities.

The particular properties and functions that the core language system would provide would include all the spoken-signed language properties in section 4.1 specific to spatial representation, though presumably in a more generic form. Thus, the core language system might have provision for all of the following. It might use individual unit concepts as the basis for representing broader conceptual content. It might group individual concepts into categories. It might associate individual concepts with overt physical representations, whether vocal or manual. It might combine individual concepts—and their physical representations—under certain constraints to represent a conceptual complex. And it might establish a subset of individual concepts as the basic schematic concepts that, in combinations, represent conceptual structure.

When in use for signed language, this core language system might then further connect with particular parts of the neural system for visual perception. I have previously called attention to the already great overlap of structural properties between spoken language and visual perception (see Talmy 2000a, chap. 2), which might speak to some neural connection already in place between the core language system and the visual system. Accordingly, the proposal here is that in the case of signed language, still further connections are brought into play, ones that might underlie the finer granularity, iconicity, gradience, and aperture limitations we have seen in signed spatial representations.

When in use for spoken language, the core language system might further connect with a putative neural system responsible for some of the characteristics present in spoken spatial representations, but absent from signed ones. These could include the packeting of spatial elements into a stable, closed set of patterned combinations, and a system for generalizing, extending, and deforming the packets. It is not clear why such a further system might otherwise exist but, speculatively, one might look to see if any comparable operations hold, say, for motor control. A cognitive capacity may have evolved (perhaps already in early animals) for the formation and maintenance of certain general-outline motor patterns suited to regularly encountered types of circumstances—that is, for the packeting of motor schemas—as well as for their modification and tailoring to the particular details of a specific occasion.

The present proposal of a more limited core language system connecting with outlying subsystems for full language function seems more consonant with contemporary neuro-scientific findings that relatively smaller neural assemblies link up in larger combinations in the subservience of any particular cognitive function. In turn, the proposed core language system might itself be found to consist of an association and interaction of still smaller units of neural organization, many of which might in turn participate in subserving more than just language functions.

4.4 Cognitive and neural plausibility

We can further briefly consider why a neural arrangement of this sort might be plausible. As a precursor, note that each of the two language modalities must be characterized in terms of the combination both of a particular form of stimulus production and of the perception of that stimulus type: vocal-auditory for spoken language and manual-visual for signed language (where "manual" is meant to cover bodily movements more broadly). Each of these two production-perception modalities has certain basic properties of structure and organization, some of which differ across the two. The differences pertinent here are in the type of iconicity, the degree of parallelism, and the type of representation.

4.4.1 Type of iconicity

The types of iconicity at hand in each of the two language modalities can account for the types of representational subsystems present in them. The concept of iconicity operative here is that of *available relevant iconicity*. This refers to the characteristics available in a modality that are iconic with referential areas of greater relevance to communication. "Iconicity" here can be understood to apply to the relationship between a representation and what it represents: They must both be realized in the same physical domain (for example, spatial, temporal, qualitative) and must exhibit a correspondence of degree or kind in that domain. "Relevant" here applies to referential areas that occur in communication more frequently, more pervasively, and more ramifiedly (for reasons that themselves can be separately examined).

One referential area of evidently great relevance is that of objects moving or located with respect to each other in space. The manual-visual modality of signed language includes among its characteristics the motion or location of objects (the hands) with respect to each other in space. And this manual system in fact exhibits the two conditions for iconicity. It is a representation that is realized in space, and what it represents is that same spatial domain. And the representations and the represented are in a relation of correspondence in degree or kind. For example, within the classifier subsystem, the more toward the vertically upward that a hand moves, the more toward the vertically upward that the angle is of the object's path being represented—not, say, the more downward or the more circular. Likewise, the faster a hand moves, the faster the motion of the object represented—not, say, the slower or the larger. By contrast, the vocal-auditory modality of spoken language does not have this form of spatial realization available among its characteristics. Whereas auditory perception by itself can determine the locations and paths of sound emitters, vocal production is fixed in place, and so cannot iconically represent such phenomena. If "throwing one's voice" were somehow a genuine physical option, spoken language might well have formed a subsystem for the iconic representation of objects following particular paths through space or occupying particular sites within space. But since such spatial localization is absent in the joint production-perception modality of spoken language, it is not available there for use as an iconic representation.

Spoken language does have other available characteristics that could have served for iconic representation. But some of these—for example, vocal timbre—are realized in a physical domain that does not also constitute a referential area of great relevance to communication. Thus, while a speaker can effect a different vocal timbre to mimic another speaker or to convey a certain attitude, this is typically only an occasional practice, not a pervasive communicative necessity, and has in fact not become part of spoken language's systematic organization. Perhaps curiously, spoken language does have other characteristics—ones in the temporal domain: rate of speed and length of interval—that could have served as a communicatively relevant iconic subsystem, but these have not entered into linguistic structure. If they had, it might have been obligatory, for example, to utter each of the three successive phrases in the sentence *The pen lay on the table, rolled to the edge, and fell off* progressively faster in iconic correspondence with the speed of the three events depicted. Or one might have had to introduce pauses between the three phrases in the sentence *I entered, sat down, and fell asleep*—in fact, pauses longer than the utterance time of each phrase—to iconically represent the duration of the phases in the depicted situation. It is not clear why this form of available and seemingly relevant iconicity was not adopted into the spoken language modality, but one explanation will be suggested below in section 4.4.3.

Having seen that the manual-visual modality has at least one available and relevant type of iconicity—spatial localization—and that the vocal-auditory modality has neither this nor much of any other relevant type of iconicity available, we could round out the picture by noting a type of iconicity not available in the manual-visual modality. It was pointed out above that, although auditory perception is attuned to the spatial domain, the combined vocal-auditory modality is not so attuned. In a parallel way, although visual perception is well attuned to surface texture, the combined manual-visual modality is not. This is because hand shapes and movements are not realized in the physical domain of texture. Accordingly, the manual-visual modality cannot represent texture iconically, and no subsystem attempting such textural representation is found in signed language.

The conclusion of the present consideration is that where a production-perception modality has available a communicatively relevant form of iconic representation, that form may become a structural part of the modality's functional system, and the neural connections that can subserve this function may be established. This conclusion accounts for why signed language has the additional subsystems that it does—ones not present in spoken language. Signed language's classifier subsystem and the subsystem of size-and-shape-specifiers specifically represent the location, motion, form, and size of objects in space—comprising a physical domain that is part of the manual-visual modality. Where signed language is not representing such spatial factors, it relies on other subsystems—principally, the lexical and inflectional subsystems—whose properties are much closer to those of spoken language.

4.4.2 Degree of parallelism

The second set of properties that differ across the two language modalities can be called "degree of parallelism". This is the number of independently variable factors that can be produced and perceived concurrently, that is, "in parallel". For the vocal-auditory modality of spoken language, the main independently variable factors would seem to be on the order of five in number: phonetic quality, pitch, loudness, rate, and timbre. By contrast, the number of independently variable factors in the manual-visual modality of signed language would appear to be more on the order of thirty, as listed in section 3.2.3.1. Again, while auditory perception alone might add several further independent factors to the former list—for example, the location in space of a stationary sound and the direction of a moving sound—and while visual perception alone could add at least color and texture to the latter list, the joint production-perception modalities would seem to have roughly the indicated degrees of parallelism. Perhaps this difference can simply be regarded as due to the nature of the respective mediums and the kinds of neural processing they afford. Given this difference, though, it is reasonable that signed language, at least within its classifier subsystem, takes advantage of this expanded set of dimensions.

4.4.3 Representational type

The third set of property differences across the two language modalities is in representational type. This involves the following three factors: granularity, categoriality, and recombinance. The granularity of a dimension pertains to the size of the components occupying that dimension (perhaps relative to the size of the whole dimension). With sufficiently fine granularity, the dimension can be considered gradient; otherwise, it can be considered to consist of discrete chunked elements. The second factor of categoriality pertains to whether a coarse-grained, chunked element in a dimension is considered to be simply a discrete step along the dimension or a qualitatively distinct category in its own right. And the third factor of recombinance pertains to whether the categories in a dimension occur there solely with their own identities and at sites relevant to that identity or can also recombine in different arrangements to constitute new higher-level entities.

In the classifier subsystem of signed language, many of the independently variable dimensions have a fine degree of granularity, in effect behaving as gradient continua for both motoric production and visual perception. For example, it seems that over a roughly continuous range, a signer can vary the locus of a hand within sign space, the contour, length, and speed of a path of motion, and the distance between Figure and Ground. By contrast, the handshapes that represent the Figure (or Manipulator, Instrument, or Ground) are for the most part organized into categories—hence, the term "classifiers". Thus, ASL has two distinct handshapes to represent a ground vehicle and an aircraft, but the first handshape cannot be continuously morphed into the second handshape to represent a series of hybrid machines that progress in design from ground vehicles to aircraft. However, certain classifier handshapes do allow a representation of magnitude, apparently not over a continuous range, but only with at most three values. As an example, with the thumb and forefinger of each hand extended and curved to form a semicircle, the two hands can be held touching, slightly separated, or much separated to represent a planar circular Figure object that is small, medium, or large. Such forms should perhaps be regarded as coarse-grained steps along a dimension, rather than as distinct categories. (Note with respect to section 4.4.1 that, even under this interpretation, the separation of the two hands exhibits iconic correspondence with the disk diameter it represents, so that neither the concept of iconicity generally nor that of correspondence of degree in particular requires gradience.) With regard to recombinance, relatively little in the classifier subsystem seems to involve this representational type.

Within spoken language, the set of dimensions that I term "vocal dynamics" —including pitch, loudness, rate, and timbre—exhibit fine granularity and so act as gradients both in vocal production and in auditory perception. If any coarser discrete steps occur in these dimensions, they do not seem to behave as categories. And there is certainly no recombinance. Vocal dynamics—which can convey speaker affect and attitude—seems to be an older inherited system antedating the evolution of the central structural system of human language.

By contrast with the other systems considered so far for their representational type, though, the central structural system of spoken language relies heavily on categories and their recombination. As is well known, this type of representation occurs at two levels. To begin with, phonetic quality is not treated as varying continuously over a range, but rather as segmented into discrete phonetic categories: the phonetic units or phonemes of a language. At the first level of recombinance, then, selections of such phonetic units from a language's inventory are arranged in particular sequences to constitute distinct higher-level units, the morphemes. Thus, there is no continuous phonetic transition from, say, "red" through "rej" to "reg" to represent the color shift from red through orange to yellow; rather, the three phonetically unrelated morphemes just mentioned are used. In turn, selections of morphemes from a language's inventory are arranged in accordance with that language's principles of morphosyntax to constitute still higher-level entities, sentences. Significantly, except for the contribution of vocal dynamics, the conveying of all conceptual content in spoken language, no matter the kind of content, is accomplished by this same format of two tiers of the recombining of discrete categories. This consistent reliance on the same single format for the central structural system of spoken language may account for why temporal iconicity—as noted in section 4.4.1—did not become part of this central structural system.

The classifier subsystem of signed language is for the most part not structured in terms of this single-format recombinant system of spoken language. First, there are a number of distinct "formats": The Figure type is represented by a handshape; the Figure's path by a linear movement of the hand; the Figure's Manner by quick hand motions outside this linear path; the Figure's angle relative to the path of motion by the angle at which the hand is held; the distance between the Figure and the Ground by the distance between the dominant and the non-dominant hand; etc. Second, the representations in these different formats largely combine with each other in a single compatible arrangement and do not shift their relations with each other to represent novel relationships among the referents. The one part of this subsystem that does seem to work this way is the classifier handshapes themselves; for example, in showing an animal moving past a vehicle or a vehicle moving past an animal by reversing the handshapes of the dominant and non-dominant hands.

In addition to spoken language's being based on discrete recombinance at the level of morphemes into sentences, the subsystems of signed language that above were suggested as being governed by the neural core language system—principally, the lexical and inflectional subsystems—do exhibit recombinance at the level of morphemes into sentences. Thus, the operation of recombinance may be one of the characteristics of the neural core language system. The question arises whether, as it evolved, the core language system could have adopted this recombinant form of organization from other cognitive systems already extant. Other cognitive systems do at least exhibit the first two representational types discussed in this section. Thus, both visual perception and auditory perception include the recognition of gradients, for example, that of the location of an object in space, as well as the recognition of discrete categories, for example, the identity of an object. But it is not clear whether these perceptual modalities exhibit the third representational type, recombinance. Perhaps visual perception can concurrently maintain the recognition of the identities of several objects as they move about relative to each other in space. But such repositionings may not yield novel higher-level arrays and hence constitute a form of recombinance. Perhaps recombinance can yet be found in visual perception—for example, in the patterns in which the perception of surfaces, edges, and vertices combine in different patterns to be perceived as different objects. Or again, perhaps motor patterns will be found to involve recombinations of some set of basic “motor units”.

However, if recombinance is in fact not found in other cognitive modalities, or is minor there, then the evolving core language system must have developed recombinance newly or ramified it into playing a major role. If that is the case, there is the question of why it might have been advantageous or necessary for that to happen. Part of the answer may lie in the following consideration. Apart from cognitive systems involved with communication across organisms, the sphere of most cognitive modalities is entirely internal to a single organism. The connection between one part of a cognitive modality and another part (for example, within the visual processing system) is neural, which has the properties of being massively parallel and of high-transmission fidelity over the relatively short distances involved. By contrast, the vocal-auditory modality between two different organisms has available only a few dimensions in parallel (again, mainly phonetic quality, pitch, loudness, timbre, and rate), and the fidelity of transmission over the relatively longer distances through the air can be low. It may be that whereas a predominantly analog representational system served for intra-organismal transmission, the newly evolving inter-organismal system had to become, as it were, “digital” to afford sufficient fidelity for the transmission of more than just a few communicative concepts. If the new language system had originally evolved as a manual-visual modality, the greater degree of parallelism and perhaps finer granularity available in that modality might well have permitted a continuation of the more analog representational type. But the utilization of the vocal-auditory channel, which lacks these properties, may have necessitated the development of something like a two-tiered discrete recombinant system. In turn, once evolved, this new representational type now carries over to those subsystems of signed language (the lexical and the inflectional) based on the core language system, whereas other subsystems of signed language instead have returned to the earlier analog representational type of the manual and visual modalities.

References

Bennett, David C. 1975. *Spatial and Temporal Uses of English Prepositions: An Essay in Stratificational Semantics.* London: Longman.

Bowerman, Melissa. 1989. Learning a semantic system: What role do cognitive predispositions play? *The Teachability of Language*, ed. by Mabel L. Rice and Richard L. Schiefelbusch. Baltimore: P. H. Brookes Pub. Co.

Brugmann, Claudia, and Monica Macaulay. 1986. Interacting semantic systems: Mixtec expressions of location. *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*, 315-328. Berkeley: Berkeley Linguistics Society.

Clark, Herb. 1973. Space, time, semantics, and the child. *Cognitive Development and the Acquisition of Language*, ed. by Timothy E. Moore. New York: Academic Press.

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

Fillmore, Charles. 1968. The case for case. *Universals in Linguistic Theory*, ed. by Emmon Bach and Robert T. Harms. New York: Holt, Rinehart and Winston.

Gruber, Jeffrey S. 1965. *Studies in Lexical Relations*. Cambridge: MIT dissertation. Reprinted as part of *Lexical Structures in Syntax and Semantics*, 1976. Amsterdam: North-Holland.

Herskovits, Annette. 1982. *Space and the Prepositions in English: Regularities and Irregularities in a Complex Domain*. Stanford: Stanford University dissertation.

Imai, Shingo. 2002. *Spatial Deixis: How Demonstratives Divide Space*. Buffalo: University at Buffalo, State University of New York dissertation.

Jackendoff, Ray. 1983. *Semantics and Cognition*. Cambridge: MIT Press.

Leech, Geoffrey. 1969. *Towards a Semantic Description of English*. New York: Longman Press.

Liddell, Scott. (in press). Sources of meaning in ASL classifier predicates. *Perspectives on Classifier Constructions in Sign Language*, ed. by Karen Emmorey. Mahwah, NJ: Lawrence Erlbaum.

Mark, David M., and Barry Smith. (under review). A science of topography: Bridging the qualitative-quantitative divide. Submitted to *Geographic Information Science and Mountain Geomorphology Chichester*, ed. by Michael P. Bishop and Jack Shroder. England: Springer-Praxis. Chapter under review.

Talmy, Leonard. 1983. How language structures space. *Spatial Orientation: Theory, Research, and Application*, ed. by Herbert L. Pick, Jr. and Linda P. Acredolo. New York: Plenum Press.

Talmy, Leonard. 2000a. *Toward a Cognitive Semantics*, vol. 1: *Concept Structuring Systems*. Cambridge: MIT Press.

Talmy, Leonard. 2000b. *Toward a Cognitive Semantics*, vol. 2: *Typology and Process in Concept Structuring*. Cambridge: MIT Press.

Talmy, Leonard. (in press). The representation of spatial structure in spoken and signed language. *Perspectives on Classifier Constructions in Sign Language*, ed. by Karen Emmorey. Mahwah, NJ: Lawrence Erlbaum.

Zubin, David, and Soteria Svorou. 1984. Orientation and gestalt: Conceptual organizing principles in the lexicalization of space. With S. Choi. *Papers from the Parasession on Lexical Semantics*, ed. by David Testen, Veena Mishra and Joseph Drogo. Chicago: Chicago Linguistic Society.

[Received 14 October 2002; revised 13 December 2002; accepted 16 December 2002]

Linguistics Department
609 Baldy Hall
State University of New York
Buffalo, NY 14260
USA
talmy@ascu.buffalo.edu

口語與手語空間結構的表達：腦神經模式

Leonard Talmy
University at Buffalo, State University of New York

前人的研究已確定了不同口語中構成空間基模的許多因素，本文擬進一步仔細研究這些因素組成的系統。口語的空間系統有幾個重要的屬性。首先，這個系統有一個各語言共通、相對有限的集合，包含了可以組成整個空間基模的基本元素。其次，這些基本元素出現在相對有限的幾個類別中。最後，每個類別只包含了相對有限的少數元素，因而只能標誌少數的空間區分。相對的，在視覺感知景像處理的結構特徵方面，手語裡空間的表達結構與口語有系統性的差異。舉例來說，手語有較多的基本元素、較多的類別，每個類別也包含較多的元素，因而可以標誌更細微的空間差異。手語可在一個片語中表達更多的空間區分。同時，這些區分有獨立的表達方式，不是透過事先設定好的基模。此外，手語的空間表徵跟視覺的空間特徵也有很大程度的對應。這些發現顯示，口語跟手語並非基於腦中某個單一的語言模組，而可能是由一個共同的核心語言系統跟與之連結的、不同的子系統分別負責口語或手語的語言功能。由於口語、手語這兩種語言模式背後的腦神經系統有本質上的差異，其行爲表現自然有所不同。

關鍵詞：美國手語，語言模組，手語，空間基模，空間結構