

CHAPTER 1

INTRODUCTION

The field of Linguistics has long avoided the semiotic complexity of human communication. There may have been many reasons for this over time, not the least of which has been the desire to produce a coherent *product* of analysis- the ‘grammar’. Unfortunately, the reflection back of ‘grammar’ onto the *subject* of investigation, actual language in use, has typically resulted in the Procrustean move to eliminate all of the inconvenient facts which the analysis failed to capture, and to then declare the remainder to be the one and true subject of the field. What ‘grammar’ has covered has varied over time, especially as the technologies of investigation have changed, and this tendency continues to this very day. But this Procrustean solution need no longer apply to ‘non-arbitrary’ qualities of speech communication. In “Quest for the essence of language”, Roman Jakobson (1965) suggests that the crucial lesson linguists may derive from the semiotic studies of both Peirce and Saussure is that the symbolic¹ (or ‘arbitrary’) nature of the linguistic sign in its typical Saussurian formulation in no way excludes it from incorporating other semiotic potentials. In Jakobson’s case, the linguistic “shifter” constituted the formal exemplification of this fact *par excellence*, since such forms are

¹Symbolic in the sense that the relation between the “form” and “meaning” of the linguistic item is established by virtue of social convention. The existence of such a relation does not preclude the existence of other relations therein, established by virtue of similarity or contiguity (as in the case of linguistic “shifters”).

simultaneously symbolic and indexical. Nonetheless, he was quite convinced that the iconic dimension of semiosis was also pervasive within the very structure of language, particularly in the constitution of the syntagmatic axis of language, as has been taken up in work in “Natural Syntax” (Haiman 1985, Tai 1985, 1993, etc.) and morphology (Bybee 1985), but also in the paradigmatic constitution of language, as exemplified in the existence of “family resemblance” relations between paradigmatically organized lexical forms (see Bolinger 1965, also Markel and Hamp 1960 on “partial morphemes”). Of course, he was not alone in the field in considering the essentially multifaceted nature of human language communication; Charles Hockett, perhaps best remembered for his check-list of the “design features” of communicative systems (Hockett 1960), and the corollary notion that “duality of patterning” was THE crucial feature that distinguishes human communication from other systems of (animal) communication², nonetheless acknowledged, particularly in work published late in his career (1978, 1987), that much of the richness of the human communicative system is due to its semiotic complexity, and that the symbolic, socially-constituted, rule-governed system of linguistic form-meaning complexes (duality of patterning) is situated within a rich semiotic web in which iconic and indexical relations between form and meaning are pervasively present³ In the

²It should be clear that “definitional” features “specific” to human communication, such as those in Hockett’s checklist, can never be taken to exclude the presence, in human communication, of other features not specific to the human system, since there is ample evidence to the contrary. Curiously, in a very real sense, this is precisely what many reductionist approaches to language amount to: the removal from the scope of investigation of any and every facet of communication, whether systematic or not, which is not subsumable under the duality of patterning model taken to be the exclusive capacity of the human (not animal) speaker.

³It seems appropriate to add that Hockett, firmly rooted in the American structuralist tradition, was keenly aware of the social embeddedness of language, and always considered Linguistics to be a subfield of Anthropology. Notably, he was a principal collaborator in the “Natural History of an Interview” project.

research presented here, I include within the scope of investigation the communicative complexity of the actional contexts within which language exists in use, namely action realized in both the auditory and visual mediums of communication in normal face-to-face interaction, focusing most particularly on the relation of contemporaneous “manual”⁴ gesture and speech. Specifically, I seek to show that the communicative complex of speaking involves evolving distributions of information and semiotic salience across the visual/auditory modalities, but crucially in a tightly integrated fashion. In so doing, I endeavor to extend the insights of previous linguistic work that has attended to the semiotic complexity of linguistic utterances, when understood as those structures that engage the Saussurian code (especially Jakobson 1965, 1980, and Hockett 1960, 1987). I will also build upon more recent work examining on the one hand the integration of iconically and indexically grounded information present in the structure of language forms in use (Fillmore, 1981, 1982a, 1982b, Chafe, 1979, 1994), and on the other the actional context within which language use exists (especially Bolinger, 1965, 1975, Clark 1992, 1996). Grounding it all is the ever-expanding body of work specifically focused on the relationship of speech to gesture (McNeill, 1987, 1992, 2000, Kendon, 1972, 1980, 2004, Butterworth and Beattie 1978, Feyereisen and de Lannoy, 1991, and Krauss et al. 1996, inter alia). In sum, I argue that modality-bounded⁵ models of language production are misleading at best, and that expanding the scope of investigation to include visual and

⁴Gestural articulation is not limited to hand movements. Much of the data I will examine includes significant articulatory complexes engaging head and upper body movement, which must not be excluded from consideration.

⁵By “modality-bounded”, I mean models that arbitrarily (in my view) restrict the scope of investigation to phenomena occurring within the primary modality in which the Saussurian code exists, whether it be the oral-auditory modality for spoken languages or the kinesic-visual modality for signed languages.

auditory phenomena together is an essential step in developing the groundwork for a comprehensive model of human communication that avoids sacrificing its essential semiotic complexity.

The modality specific properties of visual-manual action

If one considers the particular properties of the visual-manual modality that have no analogue in the oral-auditory mode⁶, the particularly rich communicative potential of systems which engage both modalities simultaneously becomes extremely easy to grasp. I claim, and will try to show, that language is structured in such a way as to take advantage of this capacity for multi-modal communication, and information in speaking, at least face-to-face, is typically distributed across modalities. The specific semiotic potentials of the different modes of expression/perception are resources which human language takes full advantage of. The pragmatic efficacy of such distributed information in face-to-face conversation should be self-evident.

Specific hypotheses

I will endeavor to show that through the generation of coherent gestural representations across discourse stretches, speakers are in effect creating self-generating indexical fields, within which gestural action may take on the function of marking both referential

⁶In particular, manual communication has the potential to simultaneously represent configured elements at the same structural level, such as two elements co-present in a specific schema. Certainly the oral-auditory channel also has the potential to represent simultaneous information without “leaking” into the visual modality, but typically, such information must be hierarchically structured in some way, such as the English combination of declarative sentence and rising intonation = query. One might also consider the genesis of such qualitative expressions as “fugly” in this vein. Dwight Bolinger (1986, 1989) wrote extensively on these sorts of overlaid representations present strictly within the oral/auditory channel of communication.

continuity (cohesion) and difference. Gestural representations are not only developed in temporal synchrony with speech, but also overtly demonstrate conceptual organization above the level of the simple utterance, through the maintenance of coherent representation over stretches of evolving discourse. Hence in gesture we may often see what in speech we must infer. Provided we attend to the evolving pattern of gesticulation accompanying speech, we will often have access to overt rather than simply pragmatic or anaphoric cues to the elements in play from moment to moment at precisely those points in a discourse when anaphors and minimally referring forms are most likely to occur: when the elements so indexed are most salient. Since the gestural schematic usually begins in conjunction with overtly referring forms, the cohesive nature of extended gesticulatory patterns makes them potential carriers of referential information, but with discourse salience marked by presence rather than absence⁷. Recurring features, such as those David McNeill refers to in developing his notion of the gesture “catchment”, are not the only sorts of gestural markers that may assume a referential or other discourse function. Complex information structures may be realized in patterns of gestural oppositions, but which crucially fit together (are ‘coherent’ in my sense) into a schematic representation of the conceptualization “at hand” in the discourse.

⁷I find it intriguing that gesture sequences so often cut right across the various hierarchies proposed to account for referent accessibility in discourse, such as Givón (1983)’s Topicality hierarchy. The gestural correlate of his Topicality hierarchy (which as per usual high saliency with minimal referring forms) is that in gesture “referential maintenance” is accomplished- by no additional effort. The trick is to understand that this is (not) measurable from a moving baseline- namely whatever effort is needed to maintain the gestural position- so actually gesture and speech are not that different after all. Changing reference typically requires additional effort, which will be visible in configurational change in gesture. The trick to distinguishing between gestural reference switching and the transformation of a configuration keeping the same elements in play is in attending to the timing relations between expressions in the two ‘channels’- the micro-alignment of gesture with speech.

Chapters to come:

The second chapter will examine some relevant prior work in gesture and speech research, then in chapter three, I will show how both American English speakers very familiar with the Warner Brothers cartoon genre and Mandarin Chinese speakers much less familiar with the same consistently anchor left-right information about events in the cartoon to an egocentric origo (thus matching the representation they had recently seen on a video screen) in gesture, while almost never including such information in co-temporal speech. This demonstrates that speakers habitually resort to structured spatial organization in gesture whether it occurs overtly in the lexical choices they make or not. In this speaking context, there is no particular benefit to be derived from including the viewer's perspective within the narrative, which is what overt mention of left-right information would do; but there is a substantial benefit from being consistent, since it allows the gestural sequences to take on a degree of schematic integrity, opening the way for them to play a role in the construction of discourse cohesion and coherence.

In chapter four, I show how speakers engaged in describing a route, a spatial memory task that does not necessarily engage short-term visual memory, produce coherent constructions in gesture, again often incorporating spatial organization not overtly expressed in lexical forms. In these data, left-right information is often crucial, but other information concerning landmark features, for instance, while not overtly be expressed in speech, may be gesturally realized in such a way that it becomes accessible to the listener (and researcher)- namely by being embedded within the matrix of current and prior gesticulation and speech. The particular nature of this spatial task results in interesting

clusters of detail at transition points in the route- places where the virtual origo must be reoriented in space, and where the gesture and speech complex must together reach a successful expression of a reorientation that may not be easily described in words. In these cases, the participants often shift their overt attention into the gestural domain, with gesture taking on the leading role in moving the discourse forward, often, but not by any means always, anchored by overt speech indexical expressions.

Then, in chapter five, I show how speakers engaged in a different spatial task, describing a living space, develop extremely elaborate gestural representations of the space as they speak. Of special interest are the shifts in scale of representation, and the layering of multiple schemas in gesture as they shift between describing the general layout of the space and particular elements of interest within them. An interesting side issue is that these data seem to show that perhaps there is less to the distinction between “layout” and “tour” strategies for describing space in actual language use than apparent in Linde and Labov (1975), who did not include gesticulation as part of their investigation.

In chapter six I summarize some of the phenomena observed in the various data sets, and then in the last extended data chapter, number seven, I shift to the examination of an academic lecture performance, and uncover therein the generation, in gesture, of a complex schematic representation of the relations between the parts of the theoretical construct that the speaker is seeking to convey to his audience, and show how it is of a piece with the gestural constructions seen in the cartoon narrative and the spatial tasks. I use this to suggest how McNeill’s notion of the “catchment” can be subsumed under a broader notion of gestural coherence, and that the schema-generating potential of gesticulation, something that only becomes apparent if you look above the level of the

single utterance, may be one of the most important functions of gesticulation during speech.

In chapter eight I return to the issue of speaker's intention and consider how gesture during speaking might be usefully related to that notion.

Finally in chapter nine I draw some conclusions about the relationship of gesture and speech, and suggest that we are now at a stage of technological development where it makes no sense for linguists not to study language in its actual context of use, for it is there that the motivations for language structures become 'visible'.

General Hypothesis

Spatial and kinesthetic cognition plays a specific role in underpinning the coordinated action of speaking and gesturing. The investigative approach that I adopt endeavors to show that gestural representation is indexically regimented⁸ within fluent speech through the spatial/kinesic organization of the "gesture space", and that the 'deictic field'⁹ of language use is also constrained by the same perceptual/actional field-anchored organization which underlies gestural production. This organization can be both local¹⁰

⁸"Regimented" in the sense that regardless of the degree of communicative focus on such gestural representation, it is constrained by the inherent indexicality of the spatial imagery it evokes. For a related use of the word "regime" see McNeill and Pedelty 1995).

⁹Karl Bühler uses the term 'Zeigfeld' to refer to the immediate context of utterance formation, that which informs and constrains the use of indexical and other speech forms (the 'Symbolfeld'). This was translated as 'deictic field' in Bühler (1983) [1934], but also as 'demonstrative field' elsewhere (e.g. Hanks 2005).

¹⁰Local regimentation involves the construction of coherent spatial representations with respect to referential structure. Most of the examples in the space and routes display this sort of structure.

and global¹¹, more than one organizing regime may contemporaneously exist within a discourse sequence¹² and apparently contradictory schemas (or indexical frameworks)¹³ can coexist within discursive sequences, when licensed by hierarchical relations between them¹⁴

¹¹Global regimentation involves the development of meta-structure in the gesture space; this is particularly salient in the academic lecture considered in chapter 7.

¹²For example, the construction of superimposed gestural referring fields in the linguistic examples associated with a. “the architecture of the theory of language” and b. “the model of language processing” in chapter 7.

¹³Easily apprehended examples are found in the coexistence of apparently contradictory spatial schemas in the gestural productions associated with the living space descriptions in chapter 5. Consider the recurring overlap of schemas that either mark the location of the room with respect to the other rooms, or mark the internal layout of the room itself. At points where two scales/degrees of proximity are simultaneously in play, scale shifts serve to license the mismatching representations. This is but one of several recurring patterns of overlay observed in the data.

¹⁴This is to be understood in the sense that the element(s) in play within each schema are related with element(s) in the other(s). This phenomenon is similar to what Scott Liddell (1995) alludes to as he develops the notions of “Surrogate space” and “Token Space” in his research on American Sign Language, and is highly suggestive of work in the “Mental Spaces” tradition found in Fauconnier (1984, 1997). It is surprising that little work in the ‘Mental Spaces’ framework has addressed the issue of the multi-modal engagement of mental spaces, despite the increasing amount of work therein focused on either oral language or sign language.

