Published in final edited form as:

Int J Semant Comput. 2008 ; 2(1): 5–9. doi:10.1142/S1793351X08000324.

Contributions of Sign Language Research to Gesture Understanding: What can Multimodal Computational Systems Learn from Sign Language Research

Ronnie B. Wilbur and

Linguistics Program, Purdue University, 500 Oval Drive, West Lafayette, IN 47907, USA,
wilbur@purdue.edu

Evguenia Malaia

Sign Language Linguistics Research Laboratory, Purdue University, 500 Oval Drive, West Lafayette, IN 47907, USA, emalaya@purdue.edu

Abstract

This paper considers neurological, formational and functional similarities between gestures and signed verb predicates. From analysis of verb sign movement, we offer suggestions for analyzing gestural movement (motion capture, kinematic analysis, trajectory internal structure). From analysis of verb sign distinctions, we offer suggestions for analyzing co-speech gesture functions.

Keywords

ASL; signed languages; event structure; motion analysis; gesture

1. Introduction

This paper provides an overview of neurological, functional (semantic) and formational (kinematic) similarities between gesture (in spoken languages) and sign languages, and elaborates on research methodologies used for semantic investigations in ASL event structure, which can be practical in informing semantic analysis of co-speech gesture in spoken languages.

One of our goals is to suggest ways in which computational modeling of gestures might be approached. This is of particular relevance and concern, not only because this journal is focused on semantic computing, but also because previous discussions of the topic have suggested that such modeling may not be possible. McNeill [22] considers one major problem of computational modeling applied to gestures to be the ‘global’ characteristic of gesture imagery and gesture meaning. For him, the direction of interpretation of meaning in gestures is ‘downward’. He suggests [22] that “the meanings of the ‘parts’ of the gesture are determined by the meaning of the whole. In fact, parts come into being only in the meaning landscape of the whole; they have no independent existence.” This view towards (lack of) gesture structure has led to a tradition of labeling gestures holistically, and to a purely sequential labeling of components: preparatory, stroke, and retraction/relaxation, with holds identified as an option. Martell [19] notes the hazards of such an approach: “a coding scheme that only labels the gesture or phase as a whole runs the risk of missing important variations in meaning created by subtle changes in the components.”

In contrast, we provide evidence that signs from American Sign Language (ASL), in addition to having well-documented formational components that are paralleled to some degree in

gestures (handshapes, locations, orientations, movements; see Brentari [3] for ASL, Calbris [4] *inter alia* for gestures), also have semantic components that are more extensive than previously recognized. These semantic components are represented formationally in sign production, produce clear neurological effects in perception, and reflect a well-defined system of event structure. We suggest that application of such an approach to gesture analysis might provide fruitful results that can be successfully used for computational modeling (in contrast to McNeill's [22] view that a gesture is holistic and non-compositional). Our discussion will also highlight areas where sign language research has several advantages that may not be so easily addressed in gesture research.

One situation that gives us confidence in our approach is the history of facial expression analysis in ASL. When first identified, facial expressions were treated as holistic units and given names related to the structure they tended to occur with, such as 'question face', 'relative clause marker' and 'conditional face'. Over time, it has become apparent that postures of the upper face (head, eyes, eyebrows) tend to span larger units such as clauses, whereas lower face markers tend to occur with lexical items like nouns, adjectives and verbs. It is now known that there are 14 separate articulators on the face/head, and that they can combine in a variety of ways, leading Sandler and Lillo-Martin [28] to suggest that they should be viewed much like the components of melody in intonational systems. The event structure system we describe in this paper interacts with some of these articulators, allowing us to identify certain ones as markers of certain (sub-)event types. Thus, in a span of 30 years, the perspective on facial expressions has changed dramatically from holistic unanalyzable units to compositional constructions of simpler meaningful components. Similarly, for sign language researchers, a hand configuration is a composite of at least handpart (palm, back, fingers), selected fingers (ones that move, if any), non-selected fingers (which are either open or closed), and facing of handpart [3]. Our hope is that the discussion presented here will help engender a similar change with respect to gesture structure.

To make our case, we will introduce basic event structure concepts (Sec. 2), provide a discussion of the neurological similarities between gesture and signs (Sec. 3), results of kinematic analysis of signs and neurolinguistic confirmation of that analysis (Sec. 4), and a viable event structure model (Sec. 5). We raise questions regarding prior gesture analyses that might be answered with greater semantic and kinematic precisions parallel to the studies we describe on signs.

2. Event Structure and Rationale for Application to Gestures

Generally, since Vendler [32], analyses of event structure have identified four basic types: States, Activities, Achievements and Accomplishments. States *hold* at a Reference Time. In Activities, there is something going on ('dynamic'), but there is *no goal or endpoint* at which the activity will be complete. In contrast, events (Achievements, Accomplishments) *happen* at an Event Time. Achievements are composed of Initial States and Final States, e.g. 'not broken' → 'broken', 'not sick' → 'sick'. Accomplishments are composed of two parts, an initial activity, such as painting, that leads to a final result, as in "She painted a picture" (verb of creation) and "He painted the house" (accomplishment verb).

The telic/atelic distinction is most clearly analyzed in terms of the internal structure of events. 'Telic' is understood as the property of events containing a natural conceptual endpoint. In contrast, 'atelic' events do not contain such a point and have the potential to continue indefinitely, without any change in internal structure. Events composed of a single subevent are *homogenous*, in that they may be divided into identical intervals, each of which is an instance of the event itself, i.e. 'walking' as an instance of 'walking'. Events composed of at least two subevents, one of which is the final state, are *heterogeneous* and cannot be divided

into identical intervals because the last/final interval will always be different from the others (i.e. 'dying' is not an instance of 'die').

Whether homogenous or heterogeneous, all four types of events can be coded in English as a single lexical item, or by multiple phrases and clauses. The English 'word' does not permit a transparent analysis of the event structure it represents, that is, the event structure is opaque at the surface and must be demonstrated by a variety of tests. Other languages may have separate morphemes that reflect each subevent.

The model Pustejovsky [27] proposes identifies States, Processes (Activities), and Transitions (Achievements and Accomplishments together). For him, the primary distinction is between static subevent type S(tate) and dynamic subevent type P(rocess). Transitions are composed of non-identical subevents: either $S \rightarrow S$ (Achievements) or $P \rightarrow S$ (Accomplishments) (Table 1).

We have extended this typology to include inception and causation [11]. Each dynamic event has an inception from an initial state — but these are not reflected in Table 1. Pustejovsky's typology is designed to neatly capture the homogeneity of atelic events (S and P) and the heterogeneity of telic events (Transitions to end States).

We have found Pustejovsky's three-category analysis useful for signs and suggest that it will be useful for gestures as well. For example, McNeill [20] investigated Vendler's classes and reported that only Activities and Accomplishments occurred in his gesture data (the event category assigned to the gesture was inferred from the accompanying spoken verb). With no Achievements occurring, there seems to be little reason to separate them from Accomplishments, although a larger study is warranted before a final decision can be made on the matter.

McNeill [20] also notes that the duration of Activity gestures was longer than that for Accomplishment gestures. He concludes that gestures join with speech to 'create specific aspectual perspectives'. This raises a distinction that is frequently confused — that of viewpoint perspective (or what is also known as 'upper aspect') and Aktionsart (or 'inner aspect'). Viewpoint aspect generally refers to perfectivity and imperfectivity, whereas inner aspect refers to the event/subevent structure just described.

Subevent analysis can be further applied to gestures, which are generally treated as composed of a preparatory phase, a stroke phase, and a relaxation phase. The stroke phase may have holds before or after it. If we treat S subevents (States) as represented by a hold and P subevents (Processes) as represented by a stroke, then the various combinations of subevents displayed in Table 1 can be seen as distinctive sequences of holds and strokes. We would suggest that this approach explains two of McNeill's observations. First, the absence of Achievement gestures would seem to be related to the lack of a way to show two holds without separating them by a stroke — whenever the hand/arm moves, the movement is visible and subject to attempts at interpretation. Second, the longer duration of Activity gestures compared to Accomplishment gestures would seem to be the result of the goal-directed movement in Accomplishment gestures, where in these cases, the goal is the final State hold. Without further data, it cannot be determined whether the stroke time and hold time need to be added together to compare with Activity gestures, or whether the movement is shorter in duration because it is faster or because it is just shorter as McNeill concludes. Either way, we are talking about 'inner aspect' here, not viewpoint perspective.

Similarly, Duncan [6] reports durational differences for 'perfective' (brief), 'durational' (extended), and 'progressive' (cyclically marked for manner) gestures from a study that compared Mandarin and English speakers. Mandarin has overt imperfectivity and

perfectivity markings that enabled Duncan to sort the gestures accordingly. These claims need to be revisited with inner and upper aspects separated from each other, and with tense carefully separated from aspect (for example, something in the past tense is not necessarily perfective) (cf. [2]). It is clear then that an event structure approach holds out hope for application of various sign language analytical techniques to further analysis of gestures. We turn now to further rationale for doing so.

3. Neurological Rationale for Comparing Gesture and Sign Language Research Methodologies

Current neurological research on comparison of gesture and sign demonstrates that gesture in non-signers serves as a supplementary communicative system. Willems *et al.* [37] demonstrated that co-speech gestures in spoken languages elicit increased activation in Brodmann areas (BA) 45 and 47, concluding that gesture and speech signal are integrated in a sentence-level semantic representation, which is constructed online and leads to higher activation levels in inferior frontal gyrus (IFG), especially in cases when semantics of the sentence and meaning of the gesture do not match. Given the facilitatory role of co-speech gesture, comparison of it with sign languages does not appear far-fetched, even given the fact that some neuroimaging studies found differential patterns of activation in Deaf and hearing subjects to specific kinds of stimuli. For example, Corina *et al.* [5] report that PET neuroimaging studies from passive viewing of three distinct action types elicited different patterns of neural activity in Deaf and hearing subjects. All three action types, including intransitive self-oriented actions (e.g. stretching, rubbing one's eyes, etc.), transitive object-oriented actions (e.g. opening a door, lifting a cup to the lips to drink), and abstract, symbolic action-signs used in ASL, consistently engaged a frontal/parietal/STS human action recognition system in hearing subjects. The same stimuli elicited completely different activation patterns in neural networks of Deaf subjects. Presentation of ASL videos to Deaf subjects predictably activated left-hemisphere perisylvian language areas, while the other kinds of stimuli activated middle-occipital temporal-ventral regions, known to participate in the detection of human bodies, faces, and movements. There was no activation similar to that of hearing subjects, i.e. of frontal/parietal circuitry. Corina *et al.* suggested that Deaf individuals may engage specialized neural systems allowing for rapid, online differentiation of meaningful linguistic actions from non-linguistic human movements.

However, Gallagher and Frith [9] have also demonstrated dissociative neural activation for expressive/emotional gestures (such as “I am angry”, “I do not care”) vs. instrumental gestures intended to change the behavior of others by communicating commands (“come here”, “look over there”) in hearing subjects. The perception of expressive gestures in the study elicited activity in the anterior paracingulate cortex, the amygdala and the temporal poles bilaterally and the right superior temporal sulcus. The perception of instrumental gestures, on the other hand, demonstrated that they were recognized as intentional communicative instruments: the elicited activity was left-lateralized and contained to the areas associated with language and motor initiation. Thus, the similarities between gesture and signed language, such as ASL, appear to be moderated by the perceived intentional goal (possibly akin to the phenomenon of recognition of sine-wave speech). This reasoning is corroborated by Emmorey *et al.* [7], who demonstrated that nonce signs (gestures modeling kinematic patterns of ASL, but not realized as meaningful units in it) activate left hemisphere language regions in Deaf signers, but not in non-signers, suggesting a specialized language-recognition neural subsystem in Deaf signers. On the other hand, passive viewing of ASL verbs and pantomimes (gestured stories) elicited very similar neural activation in Deaf signers, supporting the idea of similar processing for the stimuli with the same communicative goal.

The above research demonstrates that while signers have an ability to quickly dissociate sign language from non-linguistic gesture, both signers and non-signers demonstrate similar (language-processing-type) neural responses to the stimuli in IFG and especially BA45/47 whenever they recognize the communicative purpose of the gesture/sign. At the same time signers are at an advantage, being familiar with the visual modality and having it more closely associated with linguistic processing. That advantage is also the reason that nonce signs activate the language system in Deaf signers: it's the by-product of having a neural system that is tuned to recognize "communicative kinematics" of ASL. Thus, while gesture and sign use the same kinematic mechanisms, the perceived communicative goal is important for how the stimulus is processed. As co-speech gesture facilitates communicative process [10,24], the investigation of the nature of such facilitation can be informed by the research methodology used to explore communication confined to the visual modality: sign languages. The following section describes two research methodologies currently used in ASL research and briefly addresses theoretical conclusions drawn from it which appear relevant for both sign and spoken languages, as well as gesture.

4. Research on Semantics of the Predicates in ASL

The question addressed in the research below was formulated as a computational problem of parsing sign language predicates. The main features of a sign — hand-shape, movement, and location — need to be processed in parallel in order to identify the sign. Since predicates, and consequently verbs, are central to communicative discourse, our research concentrated on them. Our approach permits us to identify internal structure to the sign movement, to identify sequential subevents represented in the movement, and to demonstrate empirically both production and perception of our hypothesized components.

Verbs contain multifaceted information about both the semantics of an action and potential argument structures. As discussed in Sec. 2, linguistic theory classifies verbs according to whether the denoted action has an inherent (telic) end-point (fall, awaken), or whether it is considered homogenous, or atelic (read, worship). Telicity, a feature of predicates, is a major component of event structure, relating verbal semantics to syntactic frames. Interactions between telicity and grammatical phenomena have been described in many typologically distinct languages, including English, Dutch, Russian, Bengali, Icelandic and Scottish Gaelic (for a survey of relevant studies, see [8]). In order to facilitate automatic computational recognition of ASL predicates, we investigated their motion signatures with regard to telic endpoints (semantically telic vs. atelic predicates).

Wilbur [34] argues that sign languages systematically recruit characteristics of the physical world for morphological and semantic purposes, which she proposes in the form of the Event Visibility Hypothesis (EVH): In the predicate system, the semantics of event structure is visible in the phonological form of predicate signs. In particular, 'movement which stops at points in space maps semantically to the final State of telic events and its individual argument semantic variable'. That is, predicate signs contain morphemes that reflect the event structure they represent and have regular phonological forms by which they are recognized. These morphemes illustrate nicely the semantics-phonology interface and lexicon in a parallel architecture grammar [13,14].

As a result of this mapping, we expected to see differences in the ending movement of telic as compared to atelic signs, in particular, more rapid deceleration to a stop in telics than in atelics, reflecting the marking of the telic end-state in semantically telic predicates.

4.1. Motion capture study of kinematic semantics in ASL verbal predicates

A group of 29 telic and 21 atelic signs were randomized, and presented via Powerpoint to a native bilingual ASL signer, who produced them twice in isolation, once in the carrier phrase 'SIGN X AGAIN', and once sentence-medially 'SHE X TODAY'. The signer wore a Gypsy 3.0 wired motion capture suit and a pair of 18-sensor Metamotion Cybergloves (Fig. 1). We report the data from the marker on the right wrist, since the right hand, as the dominant one (our signer was right-handed), presents much of the information in sign language discourse (other than the fingers), while the non-dominant hand (left in our case) serves as ground or repeats the movement of the dominant one. Additionally, the wrist was included in the movement of all of the signs selected for the experiment.

The data from the motion capture suit was recorded into a Motionbuilder file and exported to an .htr file, with multiple channels of information, each corresponding to a sensor on the body. The header of the HTR contains basic information about the recording, including the axes order, rotation and calibration units, direction of gravity, and frame rate. Each channel's starting values are listed under [BasePosition], and each channel's data is listed for each frame. The data includes frame number, position along the 3d axis (X, Y, Z) in mm, and rotation angles and bone lengths for each axis (the latter two were not used for the present analysis, however, they would be useful for handshape analysis).

The data collection rate was 50 fps for the motion capture suit, and 30 fps for the video; the video also included an audio marker for the beginning of motion capture recording. Both the video and the vectors for the 3-dimensional location of the wrist marker were imported into ELAN (free annotation software, Max Planck Institute for Psycholinguistics, downloaded from <http://www.lat-mpi.eu/tools/elan/>) and aligned using the audio marker, and several points along the more visible axis (Y: top-bottom; and Z: left-right). The alignment's precision was 30 fps — the minimal rate of recording. The speed was then calculated in MATLAB software by subtracting the position of each frame on each axis from the previous position, summing their squares, taking the square root, and dividing by frame rate (see Eq. (1), below). This provided speed in m/s.

$$s(\text{tot}) = \Delta d(\text{tot}) / \Delta t = \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2} / 20 \text{ [mm/ms]} \quad (1)$$

(Total speed formula.)

Total Acceleration was calculated by subtracting the speed of each frame from the speed of the preceding frame, divided by the frame rate (see Eq. (2), below). This gives acceleration in m/s/s (or m/s²).

$$a(\text{tot}) = \Delta s(\text{tot}) / \Delta t = (s_1 - s_0) \text{ [m/s]} / 20 \text{ [ms]} = (s_1 - s_0) / 0.02 \text{ [m/s/s or m/s}^2\text{]} \quad (2)$$

(Total acceleration formula.)

The video was annotated in ELAN in three separate steps. First, the entire signed phrase, from the start of the hands moving upward until the hands were again down by the signer's sides, was marked by the first author. Then the start/stop boundaries were exported in a text file and sent to the second author, who extracted the kinematic data from the motion capture files for each phrase and then sent this back to the first author in three separate files (vertical-axis displacement, speed, acceleration). The first author then imported the y-axis data and used it in conjunction with the video to mark the beginning of each sign (following standard criterion: once the dominant hand assumed appropriate handshape) to the end of each movement (point

of contact, or maximal displacement). The third pass, also by the first author, concentrated solely on the movement of the sign, marking the first and last moving frame. For this analysis, the speed display was imported and used to help find the lowest speed at the beginning and end of the sign movement. Note that this information is used to identify only the start/stop frames (duration) of the sign movement; these start/stop frames were then exported to the second author, who extracted the following with MATLAB analysis: the number of frames (motion capture) the predicate spanned, the frame number on which peak speed occurred, and the number of frames from the peak speed to the local minimum speed. The measures were chosen to allow for maximal homogeneity among the metrics for predicate comparison: using the slope from peak speed to the local minimum allowed us to not be committed to the start of the sign, which can be difficult to define in discourse, and using the local minimal speed mitigated the effect of over-precise marking in ELAN. These data were analyzed by the first author using SPSS. The predicates which did not span more than 2 video frames were omitted from analysis, since precision for them would be too low.

The data from the right wrist marker indicate that in all environments, mean deceleration of telic signs is 1.5 to 2 times steeper than that of atelic signs, and the difference between telic and atelic signs is significant ($p < 0.05$). The data supports the hypothesis that there is a production difference reflecting the semantic distinction. ASL takes advantage of available perceptual distinctions to provide cues to the viewer regarding the semantics of the predicate. This distinction has implications for both the problem of parsing connected ASL discourse, and for the question of semantics and grammar interaction in signed and spoken languages.

4.2. Investigation of neural substrate for processing of semantic-kinematic predicate distinctions

The goal of the subsequent neuroimaging study was to investigate the neurological subsystems underlying event structure processing using a functional neuroimaging technique. Previous studies have demonstrated that left inferior frontal gyrus is activated during both syntactic and semantic processing in spoken and signed languages, leading to the hypothesis that it may be the locus of integration of syntactic and semantic information during sentence comprehension [12]. Based on the EVH framework we predicted that ASL predicate signs with distinctive semantics (telic vs. atelic) would elicit differentiated activation patterns in IFG (BA 44, 45 and 47), as contrasted using a subtraction paradigm.

The subjects were presented with visual stimuli consisting of telic and atelic ASL signs, and simple gestures as a baseline condition. The simple gestures consisted of non-intentional (non-communicative) slow movements of the upper arms being lowered from a *t*-position (straight out) to the sides of the body, and back up. The video recordings of the verbs signed for motion capture analysis, presented above, were re-cut into 28-second blocks of telics and atelics, each containing seven 2.5 s predicate videos with 1.5 s of monochrome grey between predicates. The simple gestures were similarly cut into blocks for the baseline, and the entire set was presented in a block paradigm. To control for attention and ensure semantic processing of the stimuli, the subjects had to indicate whether the action denoted by the predicate was more likely to occur inside a house, or outside, by pressing a button. Scans were acquired on a 3 T GE scanner, with FSPGR high-resolution anatomical images (FOV = 24 cm, 186 slices, slice thickness = 1 mm) acquired for each subject prior to the functional runs. Functional scans were collected using an EPI sequence (TE = 30 ms, TR = 2 s, FOV = 24 cm, FA = 90, FOV = 24 cm, 26 slices with 4 mm thickness and 0 interslice interval; 176 images per volume). For each subject a total of 704 volumes were collected over 4 runs with the same stimuli. Data were analyzed using SPM5 software, with motion correction, alignment to anatomical scans, normalization to the standard MNI space using the T1 template, and smoothing with an isotropic Gaussian filter. Preliminary analysis of the pilot data from 5 native signer subjects

demonstrates differential activation of the left inferior frontal gyrus, BA 45 ($p < 0.001$ uncorrected) in subtraction comparison of telic vs. atelic verbs.

Although data from more subjects is needed to determine the generalizability of the findings, the preliminary result supports the hypothesis that semantic information encoded in event structure of the ASL predicates is relevant for integration of semantic and syntactic information, and that parametric differences in production of the signs in the motion capture experiment are available for perception and used for integrative online semantic processing. The combined results of the motion capture and neuroimaging experiments in ASL confirm the prediction that world-view information is systematically recruited by American Sign Language for sign production, and is used by native signers in perception of predicate signs; our results also point to early interaction of syntax and semantics in human languages. It is the theory dealing with the basis for such interaction to which we now turn.

5. Conclusions

Let us return again to the implications for computational modeling McNeill [22] is pessimistic about computational modeling, and Wilcox [36] concludes that gestures are ‘unanalyzable’. We have shown that the kinematic patterns of sign formation systematically reflect fairly abstract conceptual and semantic components of event structure within the context of predicate signs. Further analysis [11] indicates that what might be called the sign ‘stroke’, parallel to the gesture stroke, has internal structure that can in fact be systematically and sequentially identified with (1) the inception of an event (e.g. loss of contact between two hands, or rapid acceleration of one hand away from the other or a location in space), (2) the dynamic portion of the event (extent/path, manner), and, if appropriate, (3) the completion of an event (the rapid deceleration we have documented above). The argument for considering there to be internal structure is that either the inception, or the completion, or both, must be perceptually *marked* (rapid acceleration or deceleration). Movement which begins and ends evenly does not permit postulation of internal subevent structure. It is likely then that some gestures are not ‘unanalyzable’, but need to be further investigated to identify possible components. Indeed, if one is to take seriously the argument that gestures are the forerunners to language, with mediation through sign languages [1], it is practically an imperative that some gestures are analytical in order to permit the development of sign *language* with its clear phonological and morphological structure (cf. [3,11]).

To fully appreciate the relevance of event structure and the compositional nature of it in sign structure as applied to gestures, we will reconsider the analysis of an oft-quoted example from McNeill’s [22] work, that of Tweety bird dropping a bowling ball toward the cat. The speaker says “and Tweety Bird runs and gets a bowling ba[ll and drops it down the drainpipe],” which is accompanied by a gesture in which the two hands move downward sharply from chest level with curved handshape that McNeill describes as “as if curved over the top of a large spherical object.” At the left bracket, the hands start to move up, then wait until the end of the word ‘drops’ before the gesture stroke — downward movement — in synchrony with the words ‘it down’, and then return to rest.

McNeill uses this example (among others) to argue that speech and gesture are co-timed so as to reflect the same idea — the ‘growth point’ of thought and speech. He notes that ‘it down’ is not “grammatical” (in the sense of being a constituent in English) but forms a growth point with the downward image in the gesture. It is McNeill’s contention that the gesture specifically excludes Tweety and his action of dropping the ball to focus on the movement of the bowling ball. Furthermore, this downward moving ball is a metaphoric representation of the ‘abstract idea of an antagonistic force’. He suggests that “the ‘it down’ growth point was this speaker’s taking up of a position in the world of her cartoon narration, her momentary ‘state of being’

materialized in the image of the bowling ball as an antagonistic force” [quotes around ‘state of being’ added for readability], an argument that he further uses as support for the treatment of gesture as a kind of language-imagery dialectic, a conflict that is resolved with thinking-for-speaking.

On the other hand, we see a more compositional interpretation. Rather than viewing the gesture as McNeill does, our event structure approach and our knowledge of sign structure leads us to recognize that the handshape described by McNeill and illustrated in pictures is amenable to a dual interpretation. That is, in addition to its meaning as the ball’s action through space, the handshape is also that of an animate agent holding the ball with his/her hands. Within the sign language research community, we distinguish the handpart from the selected fingers, the first of which indicates *agent* and the latter the *affected object* [3,15]. Thus what looks to McNeill like an embodied speaker’s mental image of an antagonistic force would be analyzed by us as an agent causing a ball-shaped object to move rapidly downward — ‘to drop it down’. The lack of synchrony with the English word ‘drops’ may be at least as much related to the rhythmic differences between speech and gesture [16,17] as it is to the fact that gesture is not a translation from English words into hand movements. Comparing the gesture to the exact words it occurs with is in some sense failing to recognize the structure of gesture as a system that co-occurs with speech, but also without it. In this regard, McNeill differs from other gesture researchers who do treat gestures without speech (A. Kendon, personal communication, 28 Feb. 2008).

Indeed, McNeill [21,22] also recognizes some gestures as ‘synthetic’, which he defines as “the fact that a single gesticulation concentrates into one symbolic form distinct meanings that might be spread across the entire surface of the accompanying sentence.” We suggest that event structure analysis of such gestures, coupled with further distinctions in movement and hand configurations, will reveal further compositionality of gesture structure.

It is ironic then that McNeill [22] himself offers an analysis along these lines but ends up rejecting it. He suggests that features might be combinations of forms and meaning such that the orientation of the hands reflects the force downward, the shape of the hands reflects the bowling ball, and the downward movement reflects the direction of the bowling ball and its force.

“In a model, such form-meaning pairs combine to create a gesture with the intended significance. To be global, however, the process wants to work from the overall meaning downward. Even if we force a model to proceed in this direction, it appears that form features need to have their own meanings in order for a global meaning to find them — but do they?”

Our answer is yes. We note first that gesture researchers have already demonstrated the relevance of such component meanings to co-speech gesture. For example, Özjürek *et al.* [26] demonstrated that the presence of ‘path’ or ‘manner’ meaning in a single gesture or multiple gestures is influenced by the typology of the spoken language it accompanies (verb-framed vs. satellite-framed), whereas McNeill and Duncan [21] report that speakers of Spanish include in their gestures meaning that is omitted from their speech. Hence, it has been shown that co-speech gesture may contain ‘path’ or ‘manner’ information in accordance with the speech that it accompanies, and also that gesture may contain information that is *not* included in the accompanying speech at all. This contrasts with McNeill’s view that “If speech and gesture coincide, they must present the same semantic information and perform the same pragmatic function,” and suggests that the meanings that are packed into gestures are not completely derivative of the speech they accompany, and that they could be independently analyzed as well as compared to the co-speech. Both Duncan [8] and McNeill [20] have shown the potential relevance of upper aspect/viewpoint/(im)perfectivity and inner aspect/Aktionsart to the duration and timing of co-speech gestures. It remains for gesture researchers to take the

next step, leaving behind the speech portion of the signal for a time, and applying event structure models that have shown utility for sign analysis to gestures themselves to determine their internal structures, parts, and meanings.

Finally, Loehr [17] notes that facial expression has not yet been systematically analyzed in conjunction with speech and other gestures, although they contribute to communicative rhythm. We note that event structure can also be used to analyze facial expression. Schalber (2006) reports that in Austrian Sign Language (ÖGS), atelic events are only found with continuous mouth gesture, that is, a single mouth posture. Telic events, in contrast, occur with either continuous or discontinuous mouth gestures, the latter being abrupt changes in mouth shape that tend to coincide with the end of the stroke of the predicate sign. Event structure then can be useful for analyses beyond just hand gestures.

However it is clear that gesture researchers have many remaining problems to overcome. For example, Martell [19] notes the difficulty that raters had agreeing on when ‘holds’ occurred in the gestures. Sign language research would appear to have two advantages over gesture researchers with respect to the analysis of ‘holds’. First, we have a long tradition of considering the minimum number of video frames necessary with no apparent movement to count as a hold to be two, based on a joint decision between the first author and S. Liddell in the early 1980s when both were measuring sign movements, hold, and syllables (although the maximum length to distinguish between a hold and a pause is not so clear cut). There is also somewhat clear agreement on how apparent movement is determined, for example with respect to background landmarks such as clothing (creases/lines if possible), and face. To date, gesture researchers do not appear to have such guidelines, and it is not obvious how they would be constructed, especially given the large number of frames Martell [19] found necessary for high inter-rater reliability. Second, as we have suggested with our event structure analysis, and its confirming production and perception studies, holds can be meaningful and intentional communicative subevents in sign languages, whereas their role in gesture has not yet been determined. These sign research advantages are further extended by (1) the long history of movement and syllable measurements in sign research, (2) the existence of the ELAN software tool for annotation which makes such measurements much easier and reliability easier to determine as an ongoing process rather than calculated post hoc, and (3) the existence of a clear model for annotation of phonological level configurations and movement given in the Prosodic Model of Brentari [3]. This latter model supports the steps taken in Martell [19] to code each moving articulator with two tracks, one for Location/Shape/Orientation, and one for Movement. These tracks could be seen to parallel the Inherent Features node and the Prosodic Features (movement) node, respectively. We hope our contribution will provide suggestions for further development of analytical tools for investigation of gesture structure.

Acknowledgments

This research is supported by NSF-RDE grant #0622900, by NIH grant #DC005241 and by NSF Linguistics grant #0345314. We thank Prof. Nicoletta Adamo-Villani and other personnel at the Purdue University Envision Center for Data Perceptualization for assistance with the motion capture, and Prof. Thomas Talavage, Dr. Greg Tamer, Javier Gonzalez Castillo, and others at the Purdue/Innervision MRI facility for assistance with the fMRI study. Discussion with Adam Kendon and other participants at the 30th German Linguistics Society Annual Meeting Workshop on Gestures: A Comparison of Signed and Spoken Languages, Bamberg, was exceptionally helpful.

References

1. Armstrong, D.; Stokoe, WC.; Wilcox, S. *Gesture and the Nature of Language*. Cambridge University Press; New York: 1995.
2. Bertinetto, PM. On a frequent misunderstanding in the temporal-aspectual domain: The ‘perfective = telic’ confusion. In: Cecchetto, C., et al., editors. *Semantic Interfaces*. CSLI Publications; Stanford: 2001.

3. Brentari, D. A Prosodic Model of Sign Language Phonology. MIT Press; Cambridge, MA: 1998.
4. Calbris G. From cutting an object to a clear cut analysis: Gesture as the representation of a preconceptual schema linking concrete actions to abstract notions. *Gesture* 2003;3(1):46.
5. Corina D, Chiu Y, Knapp H, Greenwald R, San Jose-Robertson L, Braun A. Neural correlates of human action observation in hearing and deaf subjects. *Brain Research* 2007;1152:111–129. [PubMed: 17459349]
6. Duncan S. Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture* 2002;2(2):183–206.
7. Emmorey, K.; Xu, J.; Gannon, P.; Goldin-Meadow, S.; Braun, A. Cognitive Neuroscience Society. New York: May. 2007 Perception of pantomime, American Sign Language verbs, and nonsense signs by deaf signers and hearing non-signers.
8. Folli R, Harley H. What language says about the psychology of events. *Trends in Cognitive Science* 2006;10.3
9. Gallagher H, Frith C. Differential cerebral activation during observation of expressive gestures and motor acts. *Neuropsychologia* 2004;42(13):1725–1736. [PubMed: 15351623]
10. Goldin-Meadow, S. *Hearing Gesture: How Our Hands Help Us Think*. Belknap Press of Harvard University Press; Cambridge, MA: 2003.
11. Grose, D. Doctoral dissertation. Purdue University; 2008. A Feature-Geometry Analysis of Event Structure: Evidence from English and American Sign Language.
12. Hagoort P. On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences* 2005;9(9):416–423. [PubMed: 16054419]
13. Jackendoff, R. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford University Press; 2003.
14. Jackendoff, R. Your theory of language evolution depends on your theory of language. *Linguistic Society of America Annual Meeting*; Chicago. 2008.
15. Kegl, J. Causative marking and the construal of agency in ASL, in CLS 21. Part 2. In: Eilfort, WH.; Kroeber, PD.; Peterson, KL., editors. *Parasession on Causatives and Agentivity*. University of Chicago, Chicago Linguistic Society; 1985.
16. Loehr, D. Doctoral dissertation. Georgetown University; Washington, DC: 2004. *Gesture and Intonation*.
17. Loehr D. Aspects of rhythm in gesture and speech. *Gesture* 2007;7(2):179–214.
18. Malaia, E. *Proceedings of International Workshop on Intelligent Linguistics Technologies*. CSREA Press; 2006. *Event Structure in Ontological Semantics*.
19. Martell, CH. Doctoral dissertation. University of Pennsylvania; Philadelphia, PA: 2005. *FORM: An Experimental in the Annotation of the Kinetics of Gesture*.
20. McNeill D. Aspects of aspect. *Gesture* 2003;3(1):1–17.
21. McNeill, D.; Duncan, SD. Growth points in thinking-for-speaking. In: McNeill, D., editor. *Language and Gesture*. Cambridge: Cambridge University Press; 2000. p. 141-161.
22. McNeill, D. Gesture and thought. In: Esposito, A.; Keller, E.; Marinaro, M.; Bratanic, M., editors. *The Fundamentals of Verbal and Non Verbal Communication and the Biometric Issues*. IOS Press; Amsterdam: 2007.
23. Metzinger, T.; Gallese, V. The emergence of a shared action ontology: Building blocks for a theory. In: Schalley, A.; Zaefferer, D., editors. *Ontolinguistics*. Mouton de Gruyter; p. 3-22.
24. Morsella E, Krauss RM. The role of gestures in spatial working memory and speech. *American Journal of Psychology*. in press.
25. O'Bryan, EL. Doctoral dissertation. Department of Linguistics, University of Arizona; *Event Structure in Language Comprehension*.
26. Özyürek AS, Kita S, Allen S, Furman R, Brown A. How does linguistic framing of events influence co-speech gestures? Insights from crosslinguistic variations and similarities. *Gesture* 2005;5(12): 219–240.
27. Pustejovsky J. The syntax of events. *Cognition* 1991;41:47–81. [PubMed: 1790655]
28. Sandler, W.; Lillo-Martin, D. *Sign Language and Linguistic Universal*. Cambridge University Press; 2006.

29. Schalber K. Event visibility in Austrian Sign Language (ÖGS). *Sign Language & Linguistics* 2006;9 (12):207–231.
30. Schalley, A.; Zaefferer, D. Ontolinguistics — an outline. In: Schalley, A.; Zaefferer, D., editors. *Ontolinguistics*. Mouton de Gruyter; 2006. p. 3-22.
31. Vendler, Z. *Linguistics in Philosophy*. Cornell University Press; Ithaca, NY: 1967.
32. Stein, A. Motion events in concept hierarchies. In: Schalley, A.; Zaefferer, D., editors. *Ontolinguistics*. Mouton de Gruyter; 2006. p. 3-22.
33. Wilbur RB. Representations of telicity in ASL. *Chicago Linguistic Society (CLS)* 2003;39(1):354–368.
34. Wilbur, RB. Complex predicates involving events, time and aspect: Is this why sign languages look so similar?. In: Quer, J., editor. *Signs of the Time: Selected Papers from TISLR*. Hamburg: Signum Press; 2004. p. 219-250.
35. Wilbur RB, Malaia E. Event Visibility Hypothesis: Motion capture evidence for overt marking of telicity in ASL. *Linguistic Society of America*. January;2008
36. Wilcox S. Gesture and language: Cross-linguistic and historical data from signed languages. *Gesture* 2004;4(1):43–73.
37. Willems R, Özyürek A, Hagoort P. When language meets action: The neural integration of gesture and speech. *Cerebral Cortex* 2007;17(10):2322–2333. [PubMed: 17159232]

Fig. 1.
Signer in motion capture suit.

Table 1

Subevents in Pustejovsky's framework.

Subevent type	Subevent notation	Telicity
States	S	atelic
Processes	P	atelic
Transitions: Achievements	$S \rightarrow S$	telic
Transitions: Accomplishments	$P \rightarrow S$	telic