Visual form and event semantics predict transitivity in silent gestures: Evidence for compositionality

Abstract

Silent gesture is not considered to be linguistic, on par with spoken and sign languages. It is claimed that silent gestures, unlike language, represent events holistically, without combinatorial structure. However, recent research has demonstrated that gesturers use consistent strategies when representing objects and events, and that there are behavioral and clinically relevant limits on what form a gesture may take to effect a particular meaning. This systematicity challenges a holistic interpretation of silent gesture, which predicts that there should be no stable form-meaning correspondence across event representations. Here, we demonstrate to the contrary that untrained gesturers systematically manipulate the form of their gestures when representing events with and without a theme (e.g., *Someone popped the balloon* vs. *Someone walked*), i.e., transitive and intransitive events. We elicited silent gestures and annotated them for visual features active in coding transitivity distinctions in sign languages. We trained linear support vector machines to make item-by-item transitivity predictions based on these features. Prediction accuracy was good across the entire dataset, thus demonstrating that systematicity in silent gesture can be explained with recourse to subunits. We argue that handshape features are constructs co-opted from cognitive systems subserving manual action production and comprehension for communicative purposes, which may integrate into the linguistic system of emerging sign languages. We further suggest that non-signers tend to map event participants to each hand, a strategy found across genetically and geographically distinct sign languages, suggesting the strategy's cognitive foundation.

Keywords: silent gesture, sign language, argument structure, event semantics, compositionality, language emergence

1 Introduction

When forced to carry the full communicative load of an utterance, gesture—in this case, called *silent gesture*—has been argued to take on language-like properties (Goldin-Meadow & Brentari, 2017). Specifically, silent gestures have a discrete form, such that individual word-like units are readily discernible. Further, these units are often arranged in a consistent constituent order (e.g., Goldin-Meadow et al., 1996; Hall et al., 2014), thus demonstrating combinatorial structure at the syntactic level. This makes them distinct from other types of gestures, which may be reduced in form and meaning, and which do not readily combine into larger structures due to their co-occurrence with or complementation of speech (McNeill, 1992).

Despite these 'language-like' properties, most descriptive work on gesture has nevertheless argued for its holistic interpretation (McNeill, 1992, 2005, and subsequent; Goldin-Meadow et al. 1996; Kendon, 2004; Wilcox, 2004; Duncan, 2005; Hostetter & Alibali, 2008; Arbib, 2010; McNeill et al., 2010; Müller, 2017, Motamedi et al., 2019, *inter alia*), apparently irrespective of its relationship to speech: Gestures are imagistic depictions of events that lack internal structure. For McNeill in particular, if a gesture appears to have discernible elements of meaning, it is that the entire gesture is first understood

from which pieces of its form may be consequently identified. For instance, McNeill (1992) describes an utterance where a participant represents a character dropping a bowling ball using co-speech gesture. The speech conveys the proposition compositionally, while a singular gesture globally represents the scene. It is from the co-occurring speech that we understand that the round handshape of the gesture refers to the bowling ball and that its downwards movement means *drop*. However, McNeill's claims have not been experimentally verified. Another possibility is that from the form of the gesture alone one can recover that (a) there is an object, and it is round; (b) there is an agent; and (c) the agent released the round object. That is, the gesture may be composed of meaningful elements that may be processed bottom-up (Wilbur & Malaia, 2008). Co-occurring speech may disambiguate the gesture, but may not be required for its transitive interpretation, the domain of interest in this study, or for the identification of other grammatical information (e.g., telicity: Strickland et al., 2015; distributivity: Marshall & Morgan, 2015; and phi features: Schlenker & Chemla, 2018). Furthermore, while the holistic view has been argued in depth for co-speech gestures, relatively little attention has been given to the potential structure of silent gesture, where such a view may not hold.

The holistic view predicts that gestures conveying a particular referent (e.g., an object or event) will vary considerably both within and between gesturers: There is no standard of form as one sees in signed and spoken languages, whereby a unit of meaning is consistently paired with a unit of form. For example, studies on event representation in silent gesture and sign have noted that there is greater variability in handshape production among non-signing gesturers than among signers of natural sign languages (Goldin-Meadow, et al., 1996; Schembri et al., 2005; Brentari et al., 2012), where handshape is indeed morphological (Brentari, 1998; Benedicto & Brentari, 2004). Despite noted variability in handshape production among gesturers, more recent work on silent gesture has demonstrated consistencies in the gross- and fine-level strategies that gesturers use to convey specific meanings (actions *vs.* objects: Padden et al., 2015; objects of different semantic classes: Ortega & Özyürek, 2016, 2020, Hwang et al. 2014, van Nispen et al., 2017; path and manner: Özçalışkan et al., 2016). On the other hand, there are strategies that gesturers *do not consider* (e.g., cupping both hands to order a drink at a bar). Further, gesturers have intuitions on what form a gesture may take in order to effect a particular meaning (Tieu et al., 2018; Schlenker & Chemla, 2018), and there are clinically relevant standards for silent gesture production in the assessment of brain lesioned individuals (van Nispen et al., 2017; see Dovern et al., 2012 for a review of such assessments): the neuropsychological literature on deficits in silent gesture production and perception implicitly argues for limits on the form that a gesture may take, such that stroke patients are differentiated from normal controls in their ability to accurately produce or interpret silent gesture. Such systematicity implies that there are rules for how gestures must look, or that there is some tentative link between form and meaning, which may stabilize over time in language development (Senghas et al., 1997; Senghas & Coppola 2001; Brentari & Coppola, 2013; Motamedi et al., 2019).

A critical exploration of the potential subunit structure of silent gesture has been suggested by Goldin-Meadow et al. (1996), but only later adopted in earnest by Brentari et al. (2012, 2017), who evaluated whether silent gesture has internal structure by way of argument structure.[1] The authors elicited silent gestures that represented the *falling, location* (intransitive), and *putting* (transitive) of two objects (*toy airplanes* and *lollipops*) and found that gesturers distinguish between intransitive and transitive events by manipulating two manual features, *Finger* and *Joint complexity* (explained in Section 2.4).

[1] The authors' main focus was the morphophonological properties of handshapes used by hearing and deaf subjects. However, the handshapes included (entity and handling) correspond to intransitive and transitive predicates (Benedicto & Brentari). We focus on this aspect of their studies.

These features were additionally found to be active in the coding of this distinction in signing populations (i.e., home signers, and signers of young and established sign languages), albeit using a distinct pattern. That is, their studies demonstrated that the *same* manual resources are recruited for the same linguistic (or communicative) function irrespective of population. Their studies, then, strongly suggest that manual features already in use by non-signers are reorganized for linguistic contrasts.[2] However, the generalizability of their findings is limited, in that only three events involving two objects were analyzed for only two manual features. In the present silent gesture elicitation experiment we expand upon Brentari and colleague's work by considering a wider array of events (72) involving a multitude of different objects (43). Moreover, we include six visual and manual features in our analysis, most of which are independently linked to transitivity distinctions in sign language or gesture research (Authors, *submitted*).

Finally, a full explanation of the decomposition[3] of gesture includes discussion of where the subunits come from. For instance, research on constituent ordering in silent gesture has demonstrated a clear agent-object-action preference even among gesturers whose native language displays a different dominant constituent ordering, suggesting that this facet of language may have domain general origins (e.g., Özçalışkan et al., 2016; Hall et al., 2014; Padden et al., 2015; Meir et al., 2017). Further evidence demonstrates that the semantics of an event affects how participants represent them in silent gesture (Schouwstra & de Swart, 2014; Christensen, Fusaroli, & Tylén, 2016; Napoli et al., 2017). Such biases suggest a cognitive foundation on which linguistic devices are constructed and from which the same may drift on the way to Language *tout court* (e.g., Gell-Mann & Ruhlen, 2011; Kemmerer, 2012).
Beyond constituent ordering, this connection between grammar, meaning, representation, and cognition has been demonstrated across a number of diverse linguistic phenomena in sign language with explicit extensions to gesture (boundedness: Wilbur, 2008, Malaia & Wilbur, 2012, Strickland et al., 2015, Kuhn et al., 2020; scalar structure: Aristodemo & Geraci, 2018; motion events: Schembri et al., 2005; discourse reference: Schlenker & Chemla, 2018), suggesting that these mappings are more pervasive and explanatory than previously thought. To that end, we ground our results in the multidisciplinary literature on manual action, and suggest that constructs pre-evolved for manual action production and perception constitute the subunits in a complex handshape, and thus constrain the form a silent gesture may have.

2 Method

Based on previous findings (Brentari et al., 2012, 2017), we hypothesized that non-signers produce the transitive-intransitive contrast in their silent gestures. Further, we hypothesized that a constrained set of visual and manual features combine to form the basis of this contrast. The systematic use of handshape features to express transitivity across a wide array of events provides strong evidence to the overall hypothesis that gestures have compositional structure. To test our hypotheses, we conducted a silent gesture production study and handshape modeling analysis, wherein participants viewed video clips of

[2] The authors stress that signs are not mere elaborations of silent gesture. The point here is only that non-signers and signers alike choose to represent transitivity contrasts using handshape, where it is also possible that one group would use handshape and the other, say, eye-brow position.
[3] To note, we distinguish between *compositionality* and *combinatoriality*, where the former applies to the combination of meaningful forms (e.g., morphemes, words) into meaningful structures, while the latter applies generally to anything that recombines irrespective of meaning. As discussed, silent gestures do combine at the syntactic level, where 'word'-like gestures are combined into gestured utterances, but have been demonstrated not to have phonological structure (Brentari et al., 2012). Our proposal, instead, is that silent gestures exhibit compositionality at the morphological level.

transitive and intransitive actions (action videos) and communicated them using only their hands. We describe how the action videos were created in 2.1, before describing our participants in 2.2 and the task in 2.3. Finally, we describe the six handshape features included in the study (2.4) and how they were coded in the participants' productions (2.5).

2.1 Action video production

Eighty live action videos were produced, 40 of which depicted the movement or change of shape of an object (intransitives), and 40 of which depicted the manipulation of an object (transitives). The experimenter was present in each video. In intransitive videos, the action unfolded in front of the experimenter (e.g., *The race car drove into the box*), but the experimenter did not watch the event. This was to avoid participants describing a *watching* event. In some intransitive videos, the experimenter was the performer of the event (e.g., *Someone walked backwards*). In transitive videos, the experimenter either initiated the action (e.g., *Someone hit the water bottle with the ball*), or watched as the action unfolded (e.g., *The ball hit the water bottle*). Objects occurring in the videos were all small enough to be manipulable by one or both hands. This was to ensure that gesturers used a consistent perspective in their productions, given the reported overlap between transitivity and viewpoint perspective in sign languages (Perniss, 2007).

To assess the degree to which the actions were 'intransitive' or 'transitive,' we collected 27–30 sentence descriptions of each video from annotators on Amazon Mechanical Turk (AMT). Sentences were annotated for containing a transitive or intransitive verb. In cases where sentences contained more than one verb, the verb most relevant to the video was retained and the remaining verbs discarded.[4] Verbs were coded as transitive if they could take an object, irrespective of whether they occurred with an object in the sentence. For example, *The linguist ate* would be coded as transitive since *eat* may take an object. Ditransitive verbs were also coded as transitive. All other verbs were annotated as intransitive. Action videos labeled with >80% consistency were retained for use in gesture elicitation. Seventy-two videos (36 intransitive, 36 transitive) met this criterion.

2.2 Participants

We elicited silent gestures from six participants (F = 3, M = 3; 27—35 years old, M_{age} = 31.66). Participants were either graduate students at an American university, or members of the surrounding community. No participant had significant experience with a sign language beyond knowledge of the manual alphabet or a few basic signs. No participant had significant acting experience.

2.3 Silent gesture elicitation

Two action lists were created. List A was sorted alphabetically, and List B reverse alphabetically. Three participants saw List A, and the other three List B. As such, transitive and intransitive events were quasi-randomly dispersed in each list: The maximum number of contiguous in/transitive events was four; the minimum was zero.

Participants each sat in front of a blue screen with a laptop situated at their left. They were instructed to view the action videos one by one by pressing an appropriate key, and then produce a manual

[4] In many cases, participants wrote phrases like, "I see a man X-ing," "It looks like he is X-ing" or "The man watches something X." In all cases, the verb represented by X was selected.

representation of the action without speech. There was minimally a one-second delay between the end of the video and the start of a production. Participants thus were not directly copying the actions depicted in the videos.

To ensure full verb-like gestures, participants were instructed only to produce the action in the videos, and to refrain from independently representing event participants. Participants were also instructed to use just their hands and to keep their productions within a reasonable window. This was to avoid eliciting full-body mimetic forms, akin to constructed action (*aka* role shift) in sign languages, which may have different properties than silent gesture (Cormier, Smith, & Zwets, 2013). The result was (6 × 72 =) 432 silent gestures for annotation.

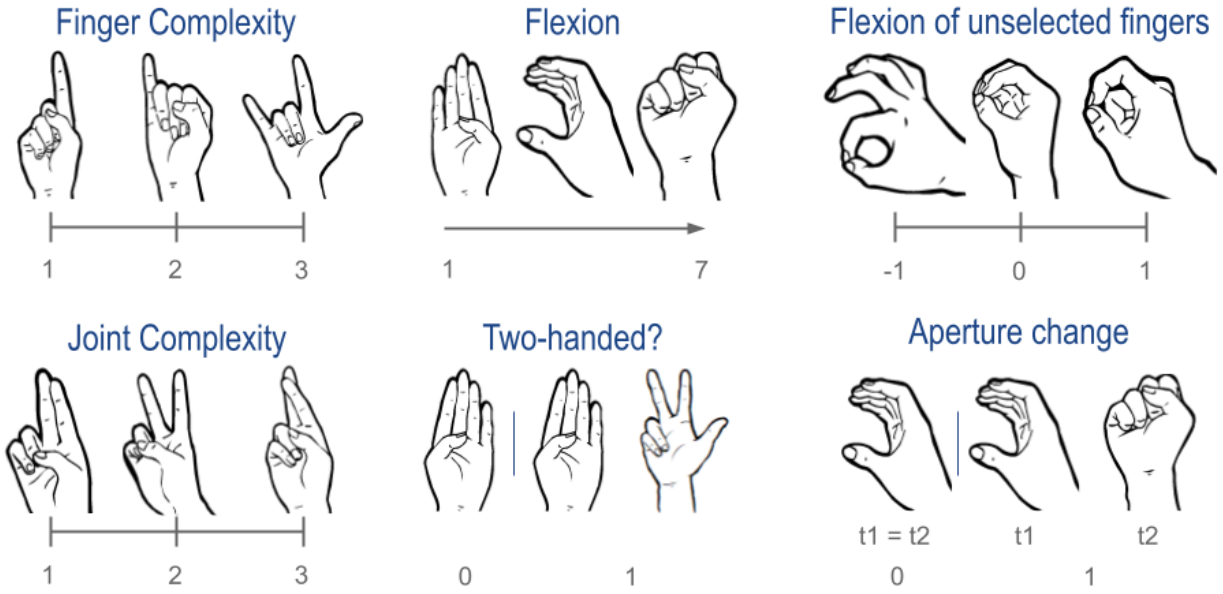

Figure 1: Illustrations of handshape parameters included in the analysis. Handshape images were generated from the sign language handshape font created by CSLDS at CUHK.

2.4 Manual/visual features under analysis

We identified six manual and visual features that are demonstrably related to transitivity coding in sign languages and/or gesture: *Finger complexity*, *Joint complexity*, *Flexion, Flexion of the unselected fingers* (*NSF flexion*), *Aperture change* and *Number of hands*. *Finger complexity* is related to the number and kind of finger groups exhibited by a handshape. *Joint complexity* is related to the number and kinds (1st, 2nd and 3rd knuckle) of joints that are involved in a handshape. Silent gestures depicting transitive events (*transitive gestures*) were found to exhibit higher finger and joint complexity in Brentari et al., 2012. However, in a subsequent analysis, Brentari et al., 2017 found that transitive gestures only exhibited higher *joint* complexity over intransitive gestures. Nevertheless both parameters were included in our study. Both are scored on a scale from 1 (least complex) to 3 (most complex) in accordance with Brentari and colleagues. See Fig. 1 for illustrations.

Flexion and *NSF flexion* code the degree of closure of the selected and unselected fingers, respectively, where *selected fingers* are the salient fingers in a handshape (e.g., the fingers that give the 'middle finger' and 'thumbs up' gestures their names), and *unselected fingers* are the backgrounded

fingers. While no study we found explicitly links *Flexion* to transitivity coding in gesture, we hypothesized that higher degrees of closure are consistent with *grasping*, hence *transitive*, handshapes. Further, the examples in Benedicto & Brentari (2004) that illustrate the differences between transitive and intransitive classifier constructions (iconic constructions in sign languages that depict the movement and manipulation of objects) can be captured by a difference in the degree of flexion exhibited by their handshapes. With respect to *NSF flexion*, Hassemer & Winter (2016, 2018) found that extended unselected fingers result in more intransitive parses, whereas curled unselected fingers result in more transitive parses, when the selected fingers are held constant. In our study, *Flexion* was scored on an ordinal scale from 1 ('extended') to 7 ('closed'), while *NSF flexion* was scored on a scale from -1 to 1, where, '-1' indicates that the unselected fingers are extended, '1' that they are closed, and '0' if there are no unselected fingers (i.e., all fingers are selected).

Next, *Aperture change* refers to a change in the closure of selected fingers of the dominant hand from open to closed, or *vice versa*. We hypothesized that a change in aperture is consistent with the grasping (e.g., *grab*) or releasing (*drop*) of an object. Finally, *Number of hands* refers to whether the production involved one or two hands. In sign languages, each hand may represent an entity, with the relative movement between the hands encoding the nature of their relationship. For example, in American Sign Language (ASL), the verb HIT[5] is articulated such that the dominant hand, representing an agent, strikes a stationary non-dominant hand, the patient; Zwitserlood, 2003; Lepic et al. 2016, Börstell et al., 2016). Both features were scored categorically for presence (1) or absence (0).

2.5 Coding & Preprocessing

If participants produced multi-gesture strings, verb-like gestures were first identified. All other gestures were excluded. Verb-like gestures were hand annotated for handshape using the coding system in Eccarius & Brentari (2008). Coding was performed by the author and an undergraduate researcher in tandem, such that each handshape code was agreed upon by both parties. From the handshape codes, individual handshape parameters (*Finger complexity, Joint complexity, Flexion*, and *NSF flexion*) could be extracted. The other parameters, *Aperture change* and *Number of hands*, were deduced from the coding of two (or more) handshapes exhibited by the dominant hand (*Aperture change*), or the presence of a handshape code for the non-dominant hand (*Number of hands*). All such were extracted via a simple Python find-replace routine.

3 Analysis

Work on transitivity encoding in silent gesture to date is not generalizable as it concerns (a) relatively few event types (e.g., *falling, putting*), (b) just two handshape features (*Finger* and *Joint complexity*), and (c) whether the proportion of these features is higher in one class of predicate or the other (Brentari et al., 2012, 2017). We argue that this simultaneously underestimates the total information available in the signal while overestimating the importance of a select few. To analyze our data, then, we use a machine-learning paradigm. This type of analysis learns to associate predictors (here, *handshape parameters*) with labels (here, *transitive* or *intransitive*) using a subset of the total data, and then makes item-by-item transitivity class decisions based on these predictors. This more faithfully emulates a sender's task in communicating

[5] By convention, signs are denoted in all capital letters.

transitivity information. We describe the specific parameters of the model below (3.1) and then report the results in 3.2.

3.1 Model parameters

The analysis was coded in Python 3.0, using Scikit-learn's (Pedregosa et al., 2011) suite of machine-learning packages. We assess the degree to which handshape parameters predict transitivity class, *transitive* or *intransitive*, using a six-fold leave-one-out cross-validation paradigm, where the dataset is split into six partitions. A classifier is trained on five of the partitions (the training set), and makes predictions on the remaining partition (the test set), returning an accuracy score (i.e., what percentage of the test set items were correctly labeled by the classifier). This is done six times (each iteration is called a *fold*), such that each partition serves as the test set once. This provides an estimate of how generalizable the model is.

Each partition contained the same number of transitive and intransitive gestures. Gestures were otherwise randomly assigned to a partition. For each fold, numerical data (*Finger complexity, Joint complexity, Flexion,* and *NSF flexion*) were scaled separately in training and testing sets to have a zero mean and unit variance, using the *StandardScaler* function. Categorical data (*Aperture change, Number of hands*) were unscaled. We then used a *select-k-best* heuristic to prune irrelevant features from the model, where 'best' was defined on the basis of one-way ANOVAs. The *k* best features are then retained for use in the model. We chose k = 4 as we expected that *Aperture change* and *Joint complexity* were irrelevant to transitivity coding based on a pilot analysis (see fn. 6 for details). Scaling and feature selection were performed on the training and test sets separately for each fold to prevent information leakage into the test set.

Finally, we used a Support Vector Classifier (SVC) with a linear kernel as our classifier. We chose a regularization parameter (C) value of 100. All other model parameters were set to their default values.[6] A new classifier was trained for each fold to prevent information from the entire dataset influencing classifier performance on just a piece of it (i.e., training on the test set), as this results in loss of generalization.

3.2 Results

3.2.1 Classifier Performance

We computed both per-fold accuracy, or the proportion of correct identifications in each test set, and the models' overall average accuracy, the sum of the number of correct per-fold identifications divided by the total number of items. Significance testing was performed by computing the cumulative mass function of the binomial distribution (as implemented in Scipy's *Stats* module; Virtanen et al., 2020) against chance ($p = 0.5$). To assess the quality of the classification, we calculated the Matthew's Correlation Coefficient (MCC). The MCC ranges from -1 (a perfect dissociation) to 1 (a perfect association), with 0 indicating no relation between the actual transitivity of the items and the classifier's predictions. Detailed results are presented in Table 1 and illustrated in Fig. 2.

[6] Model configuration was performed on pre-study data. This dataset consists of 350 pantomimes from five other non-signers (70 items per participant), with 79% of the pre-study events different from those in the main analysis.

Table:1: Per-fold and mean classifier accuracies. 'Baseline' refers to performance of the classifiers should they only choose the most frequent class. Overall quality of the analysis is reported in the final column (MCC). Accuracies in bold are significant at $p < 0.05$.

Sorted by	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Mean	Std	Baseline[a]	MCC
Random	**0.6111**	**0.7083**	**0.6240**	0.5972	**0.6389**	**0.6667**	**0.6412**	0.0371	0.50	0.2828
Participants	**0.6389**	**0.6806**	0.5000	**0.625**	**0.6667**	**0.6111**	**0.6204**	0.0587	0.50	0.2416
Events	**0.6250**	0.5972	**0.6806**	0.5972	0.5278	**0.8611**	**0.6481**	0.1054	0.50	0.2973

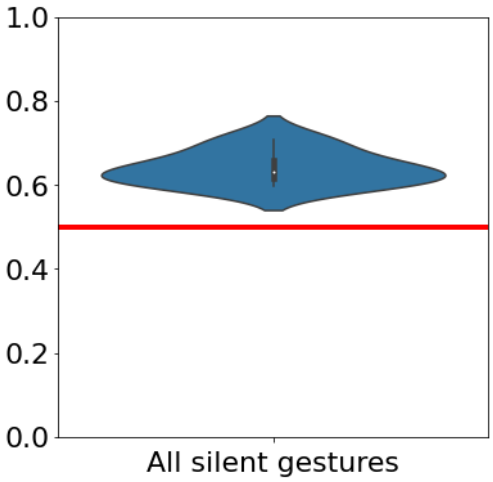

Figure 2: Violin plot of classifier accuracies across all events. Accuracy plotted on the y-axis, where 0 = 0% accuracy and 1.0 = 100% accuracy. Red line indicates chance.

Classifiers were significantly predictive with a mean accuracy of 64.12% (281/432 correct identifications; std = 0.0371; $p < 0.001$). Five of the six classifiers additionally identified their targets significantly above chance. The MCC is fair, at 0.2828, indicating that there is a true distinction between transitive and intransitive events and classifiers were able to learn it to a limited extent. This is unexpected if assuming that silent gestures are holistic: In such a case, we would expect that classifier performance would be at or below chance, as there would be no correlation between the predictors (i.e., the subunits) and transitivity class.

To verify this interpretation, we entertained a few alternative hypotheses. First, we assumed the silent gestures from one or a few participants may have been the most influential in setting model weights. Thus, we sorted the dataset by participant, such that classifiers were trained on data from five of the six participants and tested on data from the sixth. Mean classifier accuracy was 62.04% (cf. 64.12%). Performance was lower compared to where data was pseudo-randomly percolated into each fold, but not appreciably worse. Second, we tested the contribution of individual events on classifier accuracy. We sorted the dataset by events, such that the training set and testing set contained disjoint sets of events. Mean classifier accuracy was 64.81%. Here, mean classifier accuracy was identical. We thus concluded that the effect is participant neutral and that the transitivity of the events is apparent in silent gesture generally, and not specific to individual events.

At the same time, from the distribution of per-fold accuracies (Table 1), there appear to be subsets of the data where transitivity coding is either more or less predictable than in others. In particular, the

analysis where the data were sorted by events showed wide variability in classifier performance (min: 50%, max: 86.11%). Since 72 unique event types were included in the model, it is possible that some event types are less reliably encoded than others. To test this hypothesis, we deployed several follow-up analyses, using semantically-defined partitions of the dataset, which we describe in Section 4. We turn first to the analysis of the model predictors, the handshape features.

3.2.2 Analysis of model predictors

To assess which visual features may underlie the coding of transitivity distinctions in silent gesture, we analyzed the model weights output by the classifiers. To determine which handshape parameters were most informative for the classifiers' decision, we extracted the model weights from each of the six classifiers and averaged them (Table 2).

Table 2: Average model weights across six folds. Positive weights correspond with transitive items, negative weights with intransitive items. † out of 5 folds; ‡ crosses 0

Sorted by	No. of hands	Finger Complexity	Flexion	NSF flexion
Random	0.9306	0.4631	0.7862	0.0474‡
Participant	1.1298	0.3405	0.6547	0.1302‡
Event	0.9629	0.4122	0.7847†	0.2064‡

The same three predictors were identified across all six folds, namely *Number of hands, Finger complexity,* and *NSF flexion*. In five of the six folds, *Flexion* was additionally identified. Each feature on average characterized *transitive* gestures, though *NSF flexion* weakly characterized intransitive gestures in some folds in each analysis. We discuss possible interpretations for the importance of these features for transitivity coding in Sections 5.3.1 and 5.3.2.

On the other hand, in no fold were *Joint complexity* and *Aperture change* selected, as we also observed in our pre-study (fn. 6). However, Brentari and colleagues (2012, 2017) found that *Joint complexity* was consistently higher in transitive silent gestures over intransitive ones, which we failed to replicate. This may be due to the wider range of objects we included in our study. With respect to *Aperture change*, we hypothesized that a change in aperture would also be relevant to transitivity coding, as grasping/releasing an object entails such. However, this option was rarely used by our participants (45/432 productions), and was only consistently used in the events *Someone dropped a ball* (6/6 participants) and *The light turned on* (4/6 participants). The change in aperture may be more related to aspectual characteristics of these specific events: *Aperture change* has also been shown to be relevant to the coding of *telicity* in sign languages (e.g., the signs DROP and ILLUMINATE in ASL both involve a change in aperture; Wilbur, 2008).

4 Follow up analyses

The above results suggest that, while transitivity distinctions manifest across all 72 event types, subsets of events may be more strongly predictable: Classifier accuracy was appreciably worse in one or two folds in each analysis. Given previous literature on silent gesture, we hypothesized that there may be different transitivity coding strategies to represent actions with different event semantics.

4.1 Identification of subclasses

The third group (*Tool-use* and *Manner*) differs from the second group (*Manipulation* and *Movement*) in having additional entailments, which may be mapped onto additional visual cues.Each gesture was identified as belonging to one of five semantic classes, which we then sorted into three groups for follow-up analyses. Productions were labeled as being *Alternating* (n = 144) or showcasing *Manipulation* (n = 138), *Movement* (n = 120), *Tool-use* (n = 42), or *Manner* (n = 72). The *Alternating* group contained both transitive and intransitive gestures. The second group contained events of *Manipulation* (transitive) and *Movement* (intransitive), and the final group contained events of *Tool-use* (transitive class) and *Manner* (intransitive class).

We chose these semantic classes based on previous research on silent gesture and sign language. The *Alternating* category is the subject of Brentari and colleagues work (e.g., 2012, 2017) on transitivity expression in silent gesture and sign and thus expected a transitivity contrast to be present in this group. As for movement and manner, such information has been shown to manifest distinctly and consistently in the gestures of (English-speaking) non-signers (Schembri et al., 2005; Özyürek et al., 2005; Özçalışkan et al., 2016). Finally, recent research has shown that gesturers faithfully reflect object and tool shape, size, and use information in production (Padden et al., 2015; Ortega & Özyürek, 2016). We predicted that non-signers in our study would thus represent path, manner, and manipulation using distinct visual features.

We defined *alternating* events as those events that are near-identical in meaning except in the presence of an agent. For instance, while the verbs *put* and *fall* are not alternates in English, the events *Someone put the book on its side* and *The book fell on its side* have nearly the same truth conditions: a book starts upright and ends on its side. The crucial difference is the presence of an initiator. See Fig. 3 for an illustration of this point. To note, we do not imply that silent gestures have *transitive* or *unaccusative* syntax.

Table 3: Counts of the syntactic and semantic categories used in this study. *Tool-use* and *Manner* events form a proper subset of *Manipulation* and *Movement* events. These categories are disjoint from the set of *Alternating* events. Five events (out of 72) did not fall into any subcategory (*Unclassified*).

Analysis	Transitive	Intransitive	Total (per class)	Grand total (Total × 6 participants)
All events	36	36	72	432
Alternating	12	12	24	144
Manip./Mvmt.	23	20	43	258
Tool/Manner	7	12	19	114
Unclassified	0	5	5	30

The other categories were determined via the lexical entailments of the most frequent gloss for each item from the AMT action labeling survey (see Section 2.1). A production was marked as *Manipulation* if the event entails that the hand(s) hold an object (e.g., *Someone dropped a ball*, but not *Someone slapped a balloon*). A production was marked as *Movement* if it is entailed that the subject of an event is displaced (e.g., *Someone walked backwards*, but not *Someone bent over*). *Tool-use* describes events wherein a tool is used to affect some object (e.g., *Someone hammered a nail*, but not *Someone*

broke a stick [with their hands]), and *Manner* describes events where manner information is given along with path information (e.g., *The toy crawled*, but not *Someone approached a coat rack*).

A given production could thus share multiple labels. All *Manner* events were also *Movement* events, and all *Tool-use* events were also *Manipulation* events. Events identified as *Alternating* also depicted manipulation and movement (e.g., *Someone dropped a ball* and *The ball dropped*). However, in the analyses of *Movement/Manipulation* and *Tool-use/Manner* events, *Alternating* events were excluded. See Table 3. A full list of events and their semantic and transitivity designations can be found in the supplementary materials.

(a) (b)

Figure 3: The alternating events *Someone put the book on its side* (a) and *The book fell over* (b). The crucial difference between (a) and (b) is the presence of an agent in (a).

4.2 Model parameters

The same parameters from Analysis 1 were used in Analysis 2, with the following exceptions: In each subanalysis, silent gestures from one class outnumbered those from the other. As classifiers may use this frequency information to bias their guesses (e.g., they simply guess the most prevalent class), we used Scikit-learn's stratified k-fold splitter (*StratifiedKFold*). The splitter partitions the dataset such that the proportion of transitive to intransitive items is preserved across each partition. Data were otherwise randomly percolated into each partition. Further, during classification, each class was assigned a weight inversely proportional to their frequency (using the *class_weight='balanced'* flag).

4.3 Results

4.2.1 Classifier performance

Because intransitive items outnumbered transitive items, the proportion of intransitive to transitive items was used as a blind baseline, or, the accuracy the classifiers would achieve if simply guessing the most prevalent class. The baselines of the *Alternate*, *Manipulation vs. Movement*, and *Tool-use vs. Manner* analyses were $p = 0.52$, $p = 0.56$, and $p = 0.63$, respectively.

Classifiers trained on *Alternating* events were not individually accurate above chance, nor were they as a group: Mean accuracy was just 51.39% (74/144; std = 0.0786; $p > 0.05$). The MCC was near zero (0.0285), indicating very little correspondence between classifier predictions and the actual transitivity of the silent gestures in this set. On the other hand, performance improved as the semantic domain of the events was restricted: Classifiers trained on *Manipulation/Movement* events were significantly predictive at 67.44% (174/258; std= 0.0672; $p < 0.001$). The MCC is additionally higher, at

0.3486. Further, classifiers trained on *Tool-use/Manner* events were significantly predictive at 83.33% (93/114; std = 0.0828; $p < 0.001$), with the highest MCC value (0.6482), indicating that classifiers were successful in learning to distinguish intransitive from transitive events.[7] The results are summarized in Table 4 and visualized in Fig. 4.

Table 4: Per-fold and mean classifier accuracies. 'Baseline' refers to performance of the classifiers should they only choose the most frequent class. Overall quality of the analysis is reported in the final column (MCC). Accuracies in bold are significant at $p < 0.05$.

Analysis	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Mean	Std	Baseline	MCC
All	**0.6111**	**0.7083**	**0.6240**	0.5972	**0.6389**	**0.6667**	**0.6412**	0.0371	0.50	0.2828
Alternating	0.5000	0.4583	0.5000	0.6667	0.4167	0.5417	0.5139	0.0786	0.52	0.0285
Manip./Mvmt.	**0.6977**	0.5814	**0.7442**	**0.7209**	**0.7209**	0.5814	**0.6744**	0.0520	0.56	0.3486
Tool-use/Manner	**0.9474**	**0.8421**	0.7895	**0.8947**	0.6842	**0.8421**	**0.8333**	0.0828	0.63	0.6481

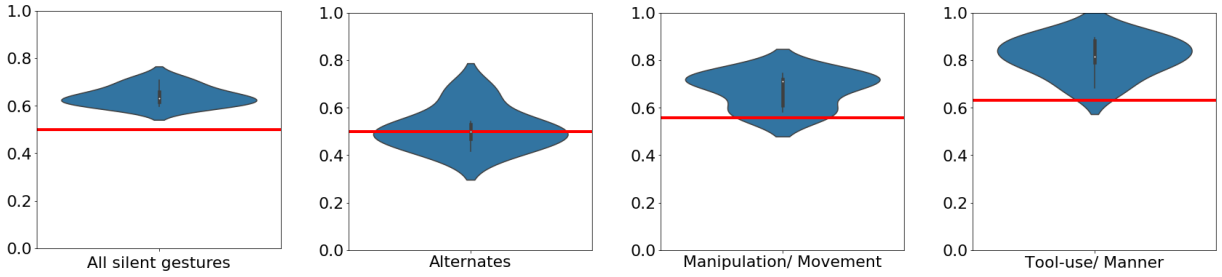

Figure 4:: Violin plots of classifier accuracies across all events (a) and different subsets of events (b–d). Accuracy plotted on the y-axis, where 0 = 0% accuracy and 1.0 = 100% accuracy. Red line indicates the blind baseline used to compute significance.

[7] As before, we ran two corollary analyses testing to see whether some or all of the results could be explained due to particular participants or events (i.e. our results are not general across either). In the first analysis, then, we sorted the dataset by participant, such that classifiers were trained on data from five of the six participants and tested on data from the sixth. Mean classifier accuracy was 57.64% (cf. 51.39%) for *Alternating* events, 67.05% (cf. 67.44%) for the analysis of *Manipulation/Movement* events, and 78.95% (cf. 83.33%) for the analysis of *Tool-use/Manner* events. Performance was generally lower in this suite of analyses, *vis a vis* those where data was pseudo-randomly percolated into each fold. But, because performance was not appreciably worse with this splitting schema and mostly the same predictors were the most influential in all cases, we believe that transitivity coding is *participant-neutral*.

In the second analysis, we tested the contribution of individual events on classifier accuracy. We sorted the dataset by events, such that the training set and testing set contained disjoint sets of events. Mean classifier accuracy was 49.31% for *Alternating* events, 65.89% for *Manipulation/Movement* events, and 83.33% for *Tool-use/Manner* events. Again, as mean classifier accuracy was not appreciably different, we thus conclude that transitivity of the events is relevant generally, and not specific to individual events.

4.2.2. Analysis of model predictors

To assess which visual features were informative for making transitivity distinctions, we analyzed the model weights output by the classifiers. Again, we extracted the six sets of model weights (one set per fold) and averaged them. We report these in Table 5.

Table 5:. Average model weights of each analysis across six folds. Positive weights correspond with transitive items, negative weights with intransitive items. Weights are reported for only those features that were selected in at least five out of the six folds. † Out of five folds. *analysis not significant.

Analysis	No. of hands	Finger Complexity	Flexion	NSF flexion
All events	0.9306	0.4631	0.7862	0.0474
Alternating*	1.6328	-	0.0808	-0.1662†
Manip./Mvmt.	1.4777	0.2974	0.7883	0.0711
Tool/Manner	0.6981	1.2626	1.2583	0.0871

Similar to the analysis of *All events*, repeated in the first row of Table 5, the same four out of six model predictors were consistently identified across folds as being most relevant to coding the transitivity distinctions within the three subclasses of events, namely: *Number of hands, Finger complexity, Flexion, and NSF flexion*. That these predictors should correlate with transitive gestures irrespective of semantic class demonstrates a consistent form-meaning correspondence indicative of subunit structure.

On the other hand, *Joint complexity* and *Aperture change* were again near-zero across all folds across all analyses, indicating that these features do not typify either syntactic class (transitive or intransitive) irrespective of the semantic class. We also note that *NSF flexion* was a weak predictor in all analyses, though its inclusion ultimately led to higher classifier accuracy. This may be due to the preponderance of four-finger handshapes in the dataset, where all fingers are selected: In the entire dataset, 50% of handshapes exhibited no unselected fingers (coded as 0), while 16% had extended unselected fingers and 34% had closed unselected fingers. As such, *NSF flexion* may not have been a generally informative feature, but relevant where present.

For the *Manipulation/Movement* and *All events* analyses, the same relative weighting of predictors was observed, with the handshape features *Finger complexity* and *Flexion* subordinate to the feature *Number of hands*. However, handshape features generally became more important than the number of hands in the analysis of *Tool-use/Manner* events. This shift may reflect a difference between how transitive events belonging to different semantic classes are distinguished from each other. Finally, every predictor was near-zero in the analysis of *Alternating* events, except for *Number of hands* and *NSF flexion*, explaining the low classifier accuracy observed above. We note, though, that *NSF flexion* was only informative in 5 of the 6 folds in this analysis, indicating that handshape was not strongly informative for *Alternating* events. We discuss why this may be so in Section 5.2.

5 Discussion

5.1 Events, transitivity and semantics

Our analyses demonstrated that handshape features predict transitivity class across 72 distinct event types, thus illustrating that silent gestures have subunit structure: Handshape features predict transitivity irrespective of event. Although transitivity distinctions are robust across a range of different events and across gesturers, silent gestures do not form a monolithic class, but vary predictably in their form-meaning correspondences: Certain, narrowly defined semantic categories enjoy a reliable coding preference (e.g., *Tool-use* events are more reliably coded than those simply entailing object manipulation). Further, differences in the weighting of handshape and visual features may distinguish between events of different semantic categories, in addition to distinguishing transitive from intransitive events: While the handshape categories (*Flexion* and *Finger complexity*) were less important than *Number of hands* in *Manipulation* events, they are more important in *Tool-use* events.

The differences in classifier performance between events of different classes is consistent with research on sign languages that demonstrates that verbs of particular semantic types will surface with one of several morphological strategies for argument realization (Meir, 2002; Oomen, 2017; Börstell, 2017). That is, the underlying semantics of verbs across sign languages affects how arguments are marked in a probabilistic way. For example, verbs that denote transfer (e.g., GIVE) are more likely to use path movement to indicate subject and (indirect) object than those that do not: The verb GIVE in ASL and Israeli Sign Language (ISL) is articulated such that the hand begins at the locus of the *giver* and moves to the locus of the *receiver*. Other verbs, like HATE (a psych verb), are less likely to use path movement, but may use other strategies (here, palm orientation). Verbs of perception and cognition, on the other hand, are often anchored at implicated organs (SEE is articulated at the eyes) or metaphorically linked body parts (THINK is articulated at the head), blocking morphological argument marking strategies altogether (Oomen, 2017; Lourenço, 2018). The diversity of argument-marking strategies in sign languages thus suggests *why* different semantic classes of events are more or less reliably coded in silent gesture by the features we included in our model, and offers additional considerations for the selection of stimuli for future studies.

5.2 Alternating events

Our analysis indicates that handshape features are not relevant to the coding of transitivity contrasts among gestures depicting *Alternating events*, and thus, we do not find evidence of combinatorial structure in this subset of the data. This is surprising given that alternating events were the subject of Brentari et al.'s (2012, 2017) experiments, where the authors did find a difference between *Finger* and *Joint complexity* in the coding of transitive over intransitive alternating events in gesture. However, it would be surprising that only a specific class of events in our own dataset are encoded holistically. We suggest two possibilities that may explain the failure of the classifiers in our study to learn transitivity distinctions among these gestures.

First, Brentari and colleagues' stimuli (the placement, location and movement of toy airplanes and lollipops) were chosen to elicit high complexity handshapes, a restriction we did not adopt. In our case, we believe that the low separability of gestures denoting *Alternating events* is related to the polysemy of certain handshapes, where the shape of the hand is consistent with both the shape/size of an object and how that object is manipulated. For instance, our gesturers mostly used handshapes consistent

with showing volume in conveying the related concepts *The boxed moved* and *Someone moved the box* (see Fig. 5). No handshape feature could thus differentiate between these events. We note that this handshape polysemy is also attested in sign language classifier constructions (Zwitserlood, 2003), where transitive and intransitive events are differentiated instead by movement.

However, other alternating events exhibited clear differences between transitive and intransitive pairs. For instance, the events *Someone put the book on its side* and *The book fell on its side* were visually distinctive in several gesturers' productions (Fig. 3a,b). Comparing (a) and (b), we see that the transitive alternate is two-handed, and involves greater flexion of the dominant hand, two strong predictors of transitive events. The intransitive alternate, by contrast, is one-handed with a completely extended handshape, hallmarks of intransitive productions. We suspect, then, that some visually distinct transitive-intransitive pairs, like those in (a) and (b), may have been misclassified as a consequence of model weights being tuned to polysemous pairs or pairs that were otherwise confounding, as in productions like Fig. 5a,b.

(a) *Someone moved a box*

(b) *The box moved*

Figure 5: Example alternating pair, *Someone moved a box* (a) and *The box moved* (b). Handshape does not differentiate between the transitive and intransitive variant.

5.3 Subunits independently grounded in cognition

Our study demonstrates that silent gestures may be decomposed into distinct features and that gesturers manipulate these features to distinguish between transitive and intransitive events. We propose that the features identified here have their roots in the cognitive systems implicated in manual action perception and production. We then explain the observation that two-handed gestures in our dataset are more likely to be transitive with recourse to the distribution of two-handed signs in sign language corpora: The subject and object of the event are mapped to each hand. This addresses the question of *what combines* in a decomposable gesture.

5.3.1 Handshape

Of the four significant predictors in each analysis, three (*Finger complexity, Flexion,* and *NSF flexion*) were related to handshape. The shape of the hand is related to the grasping of objects, based on object shape, size, or function. For example, in our dataset each representation of the placement of a book contained a handshape that is consistent with actually moving a book. That is, a gesturer may rely on an

already present repository of manual actions to inform what handshape to use to represent an action (Hostetter & Alibali, 2008, 2019).

Stores of gestural knowledge that form the basis of gesture production and perception ('praxicons' in some models) are independently motivated in the literature surrounding manual action, in particular the literature on ideomotor apraxia, an impaired ability to produce or comprehend silent gestures (see Rumiati et al., 2010 for a review). Both literatures have postulated the existence of stored action-object representations (though the details between them differ considerably). These representations range from those involved in understanding and executing low-level actions, like simple grasping or novel hand movements, to those subserving more complex actions, like tool use and other familiar actions (Fagg & Arbib, 1998; Arbib, 2005; Rothi et al., 1997; Cubelli et al., 2000; Buxbaum, 2017).

These representations may be componential: Different handshapes are required for interactions with different objects (*grabbing a pen vs. grabbing a ball*), or different interactions with the same object (*grabbing vs. using a toothbrush*). The gestured use of tools in particular has specific motor requirements beyond handshape, including the orientation of the palm and wrists, and characteristic movements. For example, gesturing a hammer driving a nail requires a particular handshape (i.e., a power grip) with a particular iterative movement of the forelimb. These features are invariant, with the specific orientation of the movement dependent on context: Hammering gestures oriented towards a wall or towards a horizontal surface are still both hammering gestures. Changes in invariant parameters serve to differentiate one gesture from another, where for example the inappropriate substitution of one handshape for another is a diagnostic of impairment (Buxbaum, 2001; Buxbaum, Kyle & Menon, 2005).

With respect to handshape, Hassemer & Winter (2018) found that the independent manipulation of the flexion of the selected and unselected fingers influences the perception of the handshape as being congruent with representing an object (intransitive parse) or holding/demonstrating an object (transitive parse) in a probabilistic way. We replicate this finding, having shown that gestures can be differentiated from each other by dint of their *Finger complexity*, *Flexion*, and *NSF flexion*. Again, a difference in *Flexion* (of the dominant hand) differentiates the gestures for holding a book (Fig. 3b) and a book falling (Fig. 3b).

Further behavioral and neuropsychological studies have demonstrated a difference between gestures used to depict the transport of manipulable objects, akin to *Manipulation* events in the present study, and gestures used to show the function of manipulable objects, akin to *Tool-use* events (Bub et al., 2008; Jax & Buxbaum, 2010; Garcea & Buxbaum, 2019). In particular, handshapes for the functional use of objects have been argued to be dependent on learned action-object representations stored in semantic memory (Rothi et al., 1997; Cubelli et al., 2000; Buxbaum, 2017). The implication of semantic memory in gesturing tool use suggests that such gesturing is a communicative act rather than a replication of the motor actions implicated in actual tool use (Goldenberg, 2017). This point is congruent with our data, in that no handshape produced to the event *Someone is hammering a nail* was consistent with actually holding a hammer: Instead, a clenched fist was produced. In all, this may explain why classifier performance on *Tool-use* events was better than on those just denoting manipulation: Participants may have used more conventionalized, consistent representations when representing events involving tool use than when representing other event types.

Finally, we are not the first to make the argument that the roots of grammatical information in sign languages may be in domains outside of language proper, and that non-signers are sensitive to this information as a result. For instance, constructs pre-evolved for vision—such as the detection of deceleration, jerk, and duration of movement—have been co-opted by the linguistic system in unrelated

sign languages (Malaia & Wilbur, 2012; Malaia et al., 2013). One feature in particular, the rapid deceleration of the hand(s) towards a point in space, guides non-signer judgments in the segmentation of naturalistic (Zacks et al., 2001, 2009), communicative, and linguistic actions (Strickland et al., 2015; Kuhn et al., 2020), though data are still missing from gesture production. The recruitment of constructs pre-evolved in other cognitive domains for the coding of grammatical phenomena is thus independently motivated.

On the other hand, action-object representations seem only appropriate to explain transitive events. But what about *intransitive* events? We found that *any* intransitive event (*Movement* or *Manner*) contrasts with *Manipulation* and *Tool-use* events, but that intransitive events do not contrast with each other.[8] As such, there does not seem to be an *intransitive*-denoting strategy detectable among the handshape features included in the model. Instead, intransitives contrasted with transitives by dint of involving *less* flexion, *less* complexity and so on, irrespective of subclass.

We suggest that lack of a consistent strategy in the representation of intransitive events is experiential. While gesturers may call upon stored representations for how to manipulate or use objects when gesturing about the objects they see, they may have less experience in representing an object with the hand (Brentari et al., 2012), leading to underspecification (Schembri et al., 2005).[9] This may surface as a tendency to represent an object by showing how it is used or by tracing it, while avoiding mapping perceptual properties of the object to the hand (Padden et al. 2015, Ortega & Özyürek, 2016). However, we did not see the use of handling strategies to represent intransitive events. Instead, in many cases, the hand was simply unspecified: a lax hand, an index finger (for tracing of path movement), or similar. The variation we see in the handshapes used to represent objects does not suggest an underlying cognitive strategy.

In sum, while others have noted that handshape production in silent gesture tasks is variable within and between subjects even when representing the same events (Goldin-Meadow et al., 1996; Schembri et al., 2005 Brentari et al., 2012, 2017), we have demonstrated that gesturers nevertheless choose from a constrained set of handshapes. These handshapes are used to affect particular meanings and can be distinguished by their internal structure (e.g., *Flexion*, *Finger complexity*). We argue that these form-meaning mappings can be explained with recourse to stores of action knowledge. Finally, it is important to stress that lexicon and praxicon are different stores of distinct information, and we are not equating the representations stored in each or the processes that combine these representations. Emmorey et al. (2011) demonstrate the differences in neural responses between tool-use silent gestures and tool-use classifier constructions from sign languages. While these silent gestures and classifiers constructions shared many superficial visual similarities, and had both been rated by non-signing participants as being 'meaningful,' the two dissociate with respect to the nature of the information they carry and the neural processing of this information. Thus the transition from a compositional gestural system to language also involves neural reorganization.

5.3.2 Number of hands

The feature *Number of hands* was also consistently identified as predictive of transitivity across all analyses, with two-handed signs being more likely to be transitive than intransitive: 73% (315/432) of

[8] This was determined via two ancillary analyses where classifiers with (a) *Tool-use* events were contrasted with *Movement*, non-*Manner* events and (b) *Manner* events were contrasted with *Movement* events.

[9] We exclude learned, culturally relevant gestures like *cutting-with-scissors* or *talk-on-telephone*, where the hands do represent scissors and telephones and not the holding of scissors or telephones.

silent gestures were two-handed, with 57% (181/315) of those being transitive (cf. 30% of one-handed silent gestures were transitive). We found that for many productions each hand may be analyzed as mapping to a unique event participant.[10] In this way, transitive events, which always involve two participants, are more likely to be two-handed than intransitive productions, which mostly involve one event participant.

For transitive events, we analyze the hands to represent an agent and theme. In the event *Someone cut bread with a knife*, all six participants used their second hand to represent (holding) the bread (Fig. 6c). For intransitive events, like *The ball dropped*, the single event participant was mapped to a single hand (Fig. 6a). However, in other intransitive events, the dominant hand represented a figure and while the non-dominant hand represented a ground: In the event *The toy car passed the block tower* (Fig. 6b), four participants used their dominant hand to convey the car, and their second hand to convey the tower (the other two productions were one-handed). That is, transitive events necessarily involve two participants while intransitive events only sometimes do, and themes are more likely to be represented by the non-dominant hand than grounds. Combined, these facts may explain the strong association of the presence of the non-dominant hand and transitive gestures.

(a) The ball bounced *(b) The toy car passed the block tower* *(c) Someone cut bread with a knife*

Figure 6: Silent gestures demonstrating a participant-to-hand mapping strategy.

Participants' mapping of event participants to each hand exhibited is consistent with strategies employed by sign languages across classifier constructions (Zwitserlood, 2003; Aronoff et al., 2005; Eccarius & Brentari, 2008) and lexical signs (Lepic et al., 2016; Börstell et al., 2016; Östling et al., 2018), although it has not been demonstrated yet whether the use of this strategy predicts the transitivity of a construction or sign. Rather, the interaction of the hands in two-handed classifier construction and signs is potentially more relevant: In classifier constructions, the relationship between the hands specifies the event. Movement of one hand towards the other may encode events of *approaching, passing* or *hitting* depending on whether one hand is in plane with the other or makes contact with it, which we saw

[10] In other cases, the second hand mirrors the setting and movement of the dominant hand in a symmetrical way, and can be understood to mean the physical extent of the object. We discuss these cases, which generally involve *Alternating* events, in Section 5.2.

mirrored in non-signer productions of the events *The man approached a coat rack, The toy car passed the block tower,* and *The ball hit a water bottle.*

The same observations have been made for two-handed lexical verbs. For instance, the signs COMMUNICATE and COMPETE are both two-handed signs in American Sign Language (ASL), Swedish Sign Language (SSL), and Israeli Sign Language (ISL), wherein the dominant hand interacts with (e.g., comes into contact with, is directed towards) the non-dominant hand. Lepic et al. claim that this iconically represents the interaction of two event participants. Similarly, signs like PARTICIPATE (ASL/SSL) and ESCAPE (ASL[11]/ISL) are two-handed, where one hand represents a figure and the other the ground.[12] Further, non-signers are sensitive to this strategy when ascribing meaning to ASL signs they see for the first time (Authors., *submitted*), and are more likely to assign a transitive interpretation to two-handed signs asymmetrical movement (i.e., one hand moves towards a stationary hand).
The tendency to use two-handed signs for concepts denoting two-participant events (i.e., the interaction of two participants or the movement of a figure relative to a ground) across geographically and genetically unrelated sign languages (see especially Östling et al., 2018) may indicate a conceptual, rather than historical or idiosyncratic, origin to how signs and classifier constructions look. In addition, the creation of new signs (or silent gestures) has been analyzed as a structure-mapping operation, whereby elements of experience or perception are mapped onto the body (Emmorey, 2014; Taub, 2001) in a componential way. We argue, then, that the use of the second hand in representing event participants in silent gesture is a reflection of this cognitively-based structure-mapping operation.

6 Conclusion

Gesture is not widely considered to be a part of language proper (see Goldin-Meadow & Brentari, 2017 for a review), although many argue for its special relationship and interaction with language when it co-occurs with speech (e.g., Özyürek et al., 2005, 2008), complements speech (Schlenker & Chemla, 2018; Schlenker, 2019), or takes on the full communicative load (Goldin-Meadow, McNeill, & Singleton, 1996; Wilbur & Malaia, 2008). As such, many have argued that individual silent gesture units (‘word’-like gestures) are holistic representations of their referents (McNeill, 2005). However, recent research has suggested that the basis of semiotic elements in language is consistently represented in the gestures of non-signing individuals, such as signaling path and manner in motion events (Özçalışkan et al., 2016), objects and their functions (e.g., Padden et al., 2015; Ortega & Özyürek, 2020), and event participants and their syntactic functions (e.g., Goldin-Meadow et al., 2008; Hall et al., 2014; Meir et al., 2017). The consistency observed in gesturing parts of a referent (e.g., path contours, object shape) already provides some counter-evidence to the claim that silent gestures are holistic: Aspects of the form of the gesture are predictable from properties of the referent they describe. Nevertheless, few studies have explicitly addressed the putative roots of grammatical information in gesture or the specific visual form of these roots. Further, while there seems to be a growing assumption that silent gestures may have subunit structure (see commentary to Goldin-Meadow & Brentari, 2017), no study to our knowledge has explicitly tested this possibility.

---

[11] Lepic et al. do not list ESCAPE as representing a figure/ground mapping strategy in ASL, although the sign is articulated such that the index finger of the dominant hand springs loose from the non-dominant hand.
[12] PARTICIPATE metaphorically represents *entering a container* where the *container* is a group.

To this end, we investigated whether transitivity contrasts are present in the form of silent gestures. The results corroborate previous studies on transitivity representations in silent gesture (Brentari et al., 2012, 2017), and extend them to novel event types. We also examined additional handshape parameters and the use of the non-dominant hand, most of which have been individually linked to transitivity coding or perception in separate studies on sign languages or gesture. We found that transitivity distinctions are reliably coded throughout a corpus of 432 silent gestures, demonstrating that gesturers consistently manipulate handshape and vary the number of hands they use in a production to distinguish between transitive and intransitive events. We argue that silent gestures demonstrate internal structure, though we remain agnostic about whether that structure is concatenative, hierarchical or something else. We also do not commit to any characterization of the process that built that structure (e.g., recursive, concatenative, etc.).

We further showed that transitivity distinctions were more robust among *Manipulation* events (e.g., *Someone dropped a ball*) and even more so among *Tool-use* events (e.g., *Someone hammered the nail*). However, transitivity contrasts were absent in *Alternating* events (e.g., *The box moved* vs. *Someone moved the box*). As such, silent gestures do not seem to form a homogenous class, but vary predictably in their form-meaning correspondences. This suggests that certain semantic categories may enjoy reliable coding preferences. We find this consistent with research on sign languages, which demonstrates that distinct semantic verb classes exhibit different argument marking strategies (e.g., Meir, 2002; Oomen, 2017; Börstell, 2017).

Analysis of the model predictors identified four visual characteristics that were consistently recruited for distinguishing transitive from intransitive events (*Finger complexity, Flexion, NSF flexion*, and *Number of hands*), with transitives more marked than intransitives across all four. We emphasize, then, that the consistency we observed is not in the use of a particular handshape (potentially evincing phonological structure; Brentari et al., 2012), but in a particular *set* of handshapes that are characterized by their phonetic descriptions. We suggest that silent gestures may already be composed of features, and that these features are borrowed from cognitive systems related to manual action-object representations. Specifically, we suggest that independently motivated action-object representations subserving the execution and perception of manual actions, i.e. those related to handshape, not only explain the robust transitivity distinction observed across the dataset, but also why this distinction was the most apparent in the *Tool-use* subset. With respect to two-handed silent gestures, we observed a tendency to express event participants on each hand, which may be explained via structure mapping (Emmorey, 2014; Taub, 2001). The mapping of event participants to each hand is similar to strategies employed in the classifier system and lexica of genetically and geographically distinct sign languages, evincing a cognitive, rather than historical or idiosyncratic, underpinning to this strategy.

Our account thus adds to a growing literature that the source of grammatical information may stem from or be shared with other cognitive abilities: Specifically, we argue that the consistency in transitivity coding we observed in silent gesture is foundational to the eventual linguistic repurposing of visual features for transitivity marking in sign languages, as argued for telicity marking in sign languages (Malaia & Wilbur, 2012) or constituent ordering across silent gesture, and spoken and sign language (Hall et al., 2014; Napoli et al., 2017; Gell-Mann & Ruhlen, 2011; Kemmerer, 2012).

Declarations of Interest

None

Acknowledgments

Handshape images in Figure 1 were generated from the sign language handshape font created by CSLDS at CUHK.

References

Arbib M. A. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *The Behavioral and brain sciences*, 28(2), 105–167. https://doi.org/10.1017/s0140525x05000038

Arbib, M. A. (2010) Holophrasis and the protolanguage spectrum. In Arbib, M. A., & Bickerton, D. (Eds.) *The emergence of protolanguage: Holophrasis vs compositionality* (Vol. 24; pp 154–168). John Benjamins Publishing. https://doi.org/10.1075/is.9.1.11arb

Aristodemo, V., & Geraci, C. (2018). Visible degrees in Italian sign language. *Natural Language & Linguistic Theory*, *36*(3), 685–699.

Aronoff, M., Meir, I., & Sandler, W. (2005). The paradox of sign language morphology. Language, 81(2), 301–344. https://doi.org/10.1353/lan.2005.0043

Authors. (*submitted*). Visual form of ASL verb signs predicts non-signer judgment of transitivity.

Benedicto, E., & Brentari, D. (2004). Where did all the arguments go?: Argument-changing properties of classifiers in ASL. *Natural Language & Linguistic Theory*, *22*(4), 743–810.

Börstell, C. (2017). *Object marking in the signed modality: Verbal and nominal strategies in Swedish Sign Language and other sign languages.* Ph.D. dissertation, Stockholm University, Sotchholm, Sweden.

Börstell, C., Lepic, R., & Belsitzman, G. (2016). Articulatory plurality is a property of lexical plurals in sign language. *Lingvisticæ Investigationes*, 39(2), 391–407.

Brentari, D., & Coppola, M. (2013). What sign language creation teaches us about language. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(2), 201–211.

Brentari, D., Coppola, M., Cho, P. W., & Senghas, A. (2017). Handshape complexity as a precursor to phonology: variation, emergence, and acquisition. *Language Acquisition*, 24 (4), 283–306.

Brentari, D., Coppola, M., Mazzoni, L., & Goldin-Meadow, S. (2012). When does a system become phonological? Handshape production in gesturers, signers, and homesigners. *Natural Language & Linguistic Theory*, 30 (1), 1–31.

Bub, D. N., Masson, M. E., & Cree, G. S. (2008). Evocation of functional and volumetric gestural knowledge by objects and words. *Cognition*, *106*(1), 27–58.

Buxbaum, L. J. (2001). Ideomotor apraxia: a call to action. *Neurocase*, 7(6), 445-458.

Buxbaum L. J. (2017). Learning, remembering, and predicting how to use tools: Distributed neurocognitive mechanisms: Comment on Osiurak and Badets (2016). *Psychological review*, 124(3), 346–360. https://doi.org/10.1037/rev0000051

Buxbaum, L. J., Kyle, K. M., & Menon, R. (2005). On beyond mirror neurons: internal representations subserving imitation and recognition of skilled object-related actions in humans. *Cognitive Brain Research*, 25(1), 226-239.

Christensen, P., Fusaroli, R., & Tylén, K. (2016). Environmental constraints shaping constituent order in emerging communication systems: Structural iconicity, interactive alignment and conventionalization. *Cognition*, 146, 67–80. DOI: https://doi.org/10.1016/j.cognition.2015.09.004

Cormier, K, Smith, S, & Zwets, M. (2013). Framing constructed action in British Sign Language narratives. *Journal of Pragmatics*, 55, 119–139.

Cubelli, R., Marchetti, C., Boscolo, G., & Della Sala, S. (2000). Cognition in action: Testing a model of limb apraxia. *Brain and Cognition*, 44(2), 144–165. https://doi.org/10.1006/brcg.2000.1226

Dovern, A., Fink, G. R., & Weiss, P. H. (2012). Diagnosis and treatment of upper limb apraxia. *Journal of neurology*, 259(7), 1269–1283.

Duncan, S. (2005) Gesture in signing: A case study from Taiwan sign language. *Language and Linguistics* 6(2):279–318.

Eccarius, P., & Brentari, D. (2008). Handshape coding made easier: A theoretically based notation for phonological transcription. *Sign Language & Linguistics*, 11 (1), 69–101.

Emmorey, K. (2014). Iconicity as structure mapping. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651), 20130301.

Emmorey, K., McCullough, S., Mehta, S., Ponto, L. L. B., & Grabowski, T. J. (2011). Sign language and pantomime production differentially engage frontal and parietal cortices, *Language and Cognitive Processes*, 26:7, 878–901, DOI: 10.1080/01690965.2010.492643

Fagg, A. H., & Arbib, M. A. (1998). Modeling parietal–premotor interactions in primate control of grasping. *Neural Networks*, 11(7-8), 1277-1303.

Garcea, FE, Buxbaum, LJ. (2019). Gesturing tool use and tool transport actions modulates inferior parietal functional connectivity with the dorsal and ventral object processing pathways. *Human Brain Mapping*, 40, 2867– 2883. https://doi.org/10.1002/hbm.24565

Gell-Mann, M., & Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences of the United States of America*, 108(42), 17290–17295. DOI: https://doi.org/10.1073/pnas.1113716108

Goldenberg, G. (2017). Facets of pantomime. *Journal of the International Neuropsychological Society*, 23(02), 121–127. doi:10.1017/s1355617716000989

Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences*, 40, E46. doi:10.1017/S0140525X15001247

Goldin-Meadow, S., McNeill, D., & Singleton, J. (1996). Silence is liberating: removing the handcuffs on grammatical expression in the manual modality. *Psychological Review*, *103*(1), 34–55. https://doi.org/10.1037/0033-295X.103.1.34

Goldin-Meadow, S., So, W.-C., Ozyurek, A., & Mylander, C. The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, 2008, 105(27), 9163-9168. doi:10.1073pnas.0710060105.

Hall, M. L., Ferreira, V. S., & Mayberry, R. I. (2014). Investigating constituent order change with elicited pantomime: A functional account of SVO emergence. *Cognitive Science*, *38*(5), 943–972.

Hassemer, J., & Winter, B. (2016). Producing and perceiving gestures conveying height or shape. *Gesture*, 15 (3), 404–424.

Hassemer, J., & Winter, B. (2018). Decoding gestural iconicity. *Cognitive Science*, 42(8), 3034–3049.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, 15, 495–514. https://doi.org/10.3758/PBR.15.3.495

Hostetter, A.B., Alibali, M.W. (2019). Gesture as simulated action: Revisiting the framework. *Psychonomic Bulletin & Review*, 26, 721–752. https://doi.org/10.3758/s13423-018-1548-0

Jax, S. A., & Buxbaum, L. J. (2010). Response interference between functional and structural actions linked to the same familiar object. *Cognition*, 115(2), 350-355.

Kemmerer, D. (2012). The cross‐linguistic prevalence of SOV and SVO word orders reflects the sequential and hierarchical representation of action in Broca's Area. *Language and Linguistics Compass*, *6*(1), 50–66.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press

Kuhn, J., Geraci, C., Schlenker, P., & Strickland, B. (2020). Boundaries in space and time: iconic biases across modalities. https://doi.org/10.31234/osf.io/mkwaz

Lepic, R., Börstell, C., Belsitzman, G., & Sandler, W. (2016). Taking meaning in hand: iconic motivations in two-handed signs. *Sign Language & Linguistics*, *19*(1), 37–81.

Lourenço, G. (2018). Verb agreement in Brazilian Sign Language: Morphophonology, syntax & semantics. Ph.D. dissertation, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

Malaia, E., & Wilbur, R. B. (2012). Kinematic signatures of telic and atelic events in ASL predicates. *Language and Speech*, 55(3), 407–421.

Malaia, E., Wilbur, R. B., & Milković, M. (2013). Kinematic parameters of signed verbs. *Journal of Speech, Language, and Hearing Research*, 56(5), 1677–1688. https://doi.org/10.1044/1092-4388(2013/12-0257)

Marshall, C. R., & Morgan, G. (2015). From gesture to sign language: Conventionalization of classifier

constructions by adult hearing learners of BSL. Topics in Cognitive Science, 7(1), 61–80.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* University of Chicago Press.

McNeill, D. (2005) *Gesture and thought*. University of Chicago Press, Chicago

McNeill, D., Duncan, S. D., Cole, J., Gallagher, S., & Bertenthal, B. (2010). Growth points from the very beginning. In Arbib, M. A., & Bickerton, D. (Eds.) *The emergence of protolanguage: Holophrasis vs compositionality* (Vol. 24; pp 117–132). John Benjamins Publishing. https://doi.org/10.1075/is.9.1.09mcn

Meir, I. (2002). A cross-modality perspective on verb agreement. *Natural Language & Linguistic Theory*, *20*(2), 413-450.

Meir, I., Aronoff, M., Börstell, C., Hwang, S. O., Ilkbasaran, D., Kastner, I., ... & Sandler, W. (2017). The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages. *Cognition*, *158*, 189–207.

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition, 192, 103964.* doi:10.1016/j.cognition.2019.05.001

Müller, C. (2017). How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture*, 16(2), 277-304.

Napoli, D. J., Spence, R. S., & de Quadros, R. M. (2017). Influence of predicate sense on word order in sign languages: Intensional and extensional verbs. *Language*, 93(3), 641-670.

van Nispen K., van de Sandt-Koenderman, WME, and Krahmer, E. (2017). Production and comprehension of pantomimes used to depict objects. *Frontiers in Psychology.* 8:1095. doi: 10.3389/fpsyg.2017.01095

Oomen, M. (2017). Iconicity in argument structure. *Sign language & linguistics*, 20 (1), 55–108.

Ortega, G., & Özyürek, A. (2016). Generalisable patterns of gesture distinguish semantic categories in communication without language. In A. Papafragou, D. Grodner, D. Mirman (Eds.), *Proceedings of the 38th Annual Meeting of the Cognitive Science Society (CogSci 2016)* (pp. 1182–1187)

Ortega, G., & Özyürek, A. (2020). Types of iconicity and combinatorial strategies distinguish semantic categories in silent gesture across cultures. *Language and Cognition*, *12*(1), 84–113.

Östling, R., Börstell, C., & Courtaux, S. (2018) Visual iconicity across sign languages: Large-scale automated video analysis of iconic articulators and locations. *Frontiers in Psychology.* 9:725. doi: 10.3389/fpsyg.2018.00725

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016). Does language shape silent gesture?. *Cognition*, *148*, 10–18. https://doi.org/10.1016/j.cognition.2015.12.001

Özyürek, A., Kita, S., Allen, S., Furman, R., & Brown, A. (2005). How does linguistic framing of events influence co-speech gestures? Insights from crosslinguistic variations and similarities. *Gesture*, 5(1-2), 219–240. https://doi.org/10.1075/gest.5.1.15ozy

Özyürek, A., Kita, S., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2008). Development of cross-linguistic variation in speech and gesture: Motion events in English and Turkish. *Developmental Psychology, 44*(4), 1040–1054. https://doi.org/10.1037/0012-1649.44.4.1040

Padden, C., Hwang, S. O., Lepic, R., & Seegers, S. (2015). Tools for language: Patterned iconicity in sign language nouns and verbs. *Topics in Cognitive Science*, *7*(1), 81–94.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

Perniss, P. (2007). Achieving spatial coherence in German Sign Language narratives: The use of classifiers and perspective. *Lingua*, 117(7), 1315–1338.

Rothi, L. J., Ochipa, C., & Heilman, K. M. (1997). A cognitive neuropsychological model of limb praxis and apraxia. In L. J. Rothi, & K. M. Heilman (Eds.), *Apraxia: The neuropsychology of action* (pp. 29–49). Hove: Psychology.

Schembri, A., Jones, C., & Burnham, D. (2005). Comparing action gestures and classifier verbs of motion: Evidence from Australian Sign Language, Taiwan Sign Language, and nonsigners' gestures without speech. *Journal of Deaf Studies and Deaf Education*, *10*(3), 272–290.

Schlenker, P. (2019). Gestural semantics. *Natural Language and Linguistic Theory*, 37, 735–784. https://doi.org/10.1007/s11049-018-9414-3

Schlenker, P., & Chemla, E. (2018). Gestural agreement. *Natural Language & Linguistic Theory*. 36, 587–625. https://doi.org/10.1007/s11049-017-9378-8

Schouwstra, M., & de Swart, H. (2014). The semantic origins of word order. *Cognition, 131*(3), 431–436. https://doi.org/10.1016/j.cognition.2014.03.004

Senghas, A., Newport, E. L., & Supalla, T. (1997). Argument structure in Nicaraguan Sign Language: The emergence of grammatical devices. In *E. Hughes & A. Greenhill (Eds.), Proceedings of the 21st annual Boston University Conference on Language Development*.

Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological science*, 12(4), 323–328.

Strickland, B., Geraci, C., Chemla, E., Schlenker, P., Kelepir, M., & Pfau, R. (2015). Event representations constrain the structure of language: Sign language as a window into universally accessible linguistic biases. *PNAS*, 112(19), 5968–5973.

Taub, S. F. (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge: Cambridge University Press.

Tieu, L., Pasternak, R., Schlenker, P., & Chemla, E. (2018). Co-speech gesture projection: Evidence from inferential judgments. *Glossa: A Journal of General Linguistics*, 3(1), 109. DOI: http://doi.org/10.5334/gjgl.580

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., … , & SciPy 1.0 Contributors. (2020) SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272.

Wilbur, R. B. (2008). Complex predicates involving events, time and aspect: Is this why sign languages look so similar? In J. Quer (Ed.), *Signs of the time: Selected papers from TISLR 2004*, 217–250. Hamburg: Signum Press

Wilbur, R. B., & Malaia, E. (2008). Contributions of sign language research to gesture understanding: What can multimodal computational systems learn from sign language research. *International Journal of Semantic Computing*, 2(1), 5–19.

Wilcox, S. (2004). Gesture and language: Cross-linguistic and historical data from signed languages. *Gesture*, *4*(1), 43-73.

Zacks, J.M., Braver, T.S., Sheridan, M.A., Donaldson, D.I., Snyder, A.Z., Ollinger, J.M., et al. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, 4(6), 651–655.

Zacks, J.M., Kumar, S., Abrams, R.A., & Mehta, R. (2009). Using movement and intentions to understand human activity. *Cognition*, 112(2), 201–216.

Zwitserlood, I. E. P. (2003). *Classifying hand configurations in Nederlandse Gebarentaal (Sign Language of the Netherlands)*. Ph.D. dissertation, Utrecht University, Utrecht, Netherlands.

Supplementary Materials

https://osf.io/8xvjt/?view_only=c5d8e902e18f496c946fe27752a77cbf