

Kinect-ing the Dots: Using Motion-Capture Technology to Distinguish Sign Language Linguistic From Gestural Expressions

Language and Speech

1–22

© The Author(s) 2023

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/00238309231169502

journals.sagepub.com/home/las**Rose Stamp** 

Department of English Literature and Linguistics, Bar-Ilan University, Israel

David Cohen**Hagit Hel-Or**

Department of Computer Science, University of Haifa, Israel

Wendy Sandler

Sign Language Research Lab, University of Haifa, Israel

Abstract

Just as vocalization proceeds in a continuous stream in speech, so too do movements of the hands, face, and body in sign languages. Here, we use motion-capture technology to distinguish lexical signs in sign language from other common types of expression in the signing stream. One type of expression is *constructed action*, the enactment of (aspects of) referents and events by (parts of) the body. Another is *classifier constructions*, the manual representation of analogue and gradient motions and locations simultaneously with specified referent morphemes. The term *signing* is commonly used for all of these, but we show that not all visual signals in sign languages are of the same type. In this study of Israeli Sign Language, we use motion capture to show that the motion of lexical signs differs significantly along several kinematic parameters from that of the two other modes of expression: constructed action and the classifier forms. In so doing, we show how motion-capture technology can help to define the universal linguistic category “word,” and to distinguish it from the expressive gestural elements that are commonly found across sign languages.

Keywords

Sign language, classifier constructions, gesture, constructed action, linguistic categories, motion capture, movement

Corresponding author:

Rose Stamp, Department of English Literature and Linguistics, Bar-Ilan University, Building 404, Ramat Gan 5290002, Israel.

Email: rose.stamp@biu.ac.il

Introduction

Until about 60 years ago, most people thought that sign languages consisted entirely of gestural hand motions and had little in common with spoken languages (e.g., Bloomfield, 1933). William Stokoe (1960) showed that this impression was mistaken. He demonstrated that signs in American Sign Language consist of contrastive sub-lexical units of handshapes, locations, and movements at a phonological level, just as spoken languages consist of meaningless phonemes—opening the gates to the study of sign languages as part of our natural linguistic endowment.

More recent work builds on Stokoe's model, but many researchers now acknowledge that sign languages incorporate gestural elements, and that signed words (lexical signs) and sentences are interspersed with gestural elements. Lexical signs are at one pole of the signing stream, with very constrained contrastive phonological elements (see, for example, Brentari, 1998; Sandler, 1989, 2012b). At the other pole is constructed action, in which the body or parts of it mimic real-word behavior of referents. Classifier constructions present a special challenge; they fall somewhere in between, as we demonstrate below. There is no precise spoken language equivalent of classifier constructions to guide us (Schembri, 2003), and their place in the typology of elements in the signing stream is still debated.

1.1 Signed word (*lexical signs*)

There is only one movement in the canonical sign language word (Coulter, 1982; Liddell & Johnson, 1989; Sandler, 1999; Wilbur, 1999). The movement can be simple or complex. The canonical phonological structure of sign language lexical words consists of a single handshape which moves from one location to another. Alternatively, the single movement can be produced hand-internally, by a change in finger positions or in orientations of the hand. These movements—either in a path from one location to another (Figure 1(a)) or hand internal movement from one position to another—can be combined. But if they are, this combination must be simultaneous (as in Figure 1(b)), always resulting in a single syllable nucleus, simple (path only or internal only), or complex (a simultaneous combination; see Sandler, 1993).¹ Importantly for this study, the path movement within a sign can be characterized by one of a very small number of features, typically (but not exclusively) [straight] or [curved] (Sandler, 1996), shown in Figure 1(c).

Lexical signs are described as highly stable, standardized in form and meaning, and frequently used in the language. In other words, the handshape, location, and movement of each lexical sign are conventionalised within each sign language community (Sandler, 2012b; Stokoe, 1960).² In spontaneous signing, mouth movements can accompany signs, but they are typically not mimetic

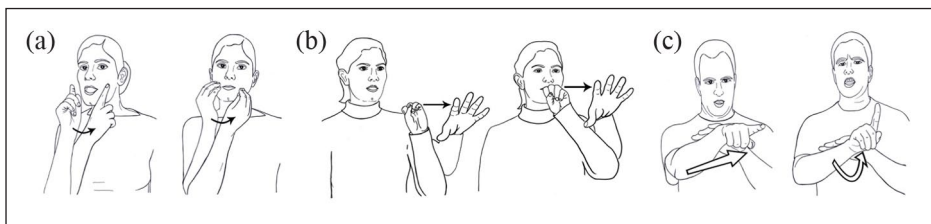


Figure 1. Minimal pairs of words in Israeli Sign Language: (a) MOTHER, NOON, distinguished by handshape features; (b) SEND, TATTLE, distinguished by location features; and (c) ESCAPE, BETRAY, distinguished by movement features.

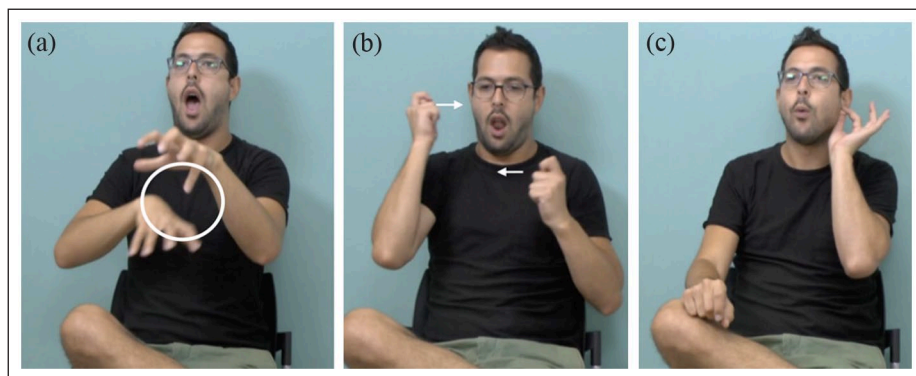


Figure 2. Examples of three lexical signs from ISL retellings of a Charlie Chaplin video: (a) LION, (b) SHOCK, and (c) FEMALE-PERSON. ISL: Israeli Sign Language.

of any referent in a discourse. Rather, they either result from mouthing the corresponding Hebrew words or represent conventionalized modifiers like adjectives or adverbs.³ Examples of mouthing Hebrew words are shown in Figure 2(a). The sign LION is produced in the neutral signing space in front of the body with both hands in a clawed shape, each moving forward in an alternating manner. This LION⁴ sign is conventionalised across signers of Israeli Sign Language (ISL). The images in Figure 2 show mouth shapes corresponding to part of the corresponding Hebrew word (indicated here with boldface and underlining): LION: [arye], SHOCK (helem), and FEMALE-PERSON ([baXura]). In sum, lexical signs are unarguably taken to be linguistic units, specified in the sign language lexicon.

1.2 Constructed action

At the other pole lies overt Constructed Action, in which the signer enacts the behavior of a referent, so that each body part—head, face, hand/s, and torso—represents the corresponding body part of the referent (see Figure 3).

Constructed action corresponds to certain kinds of gestural phenomena in spoken languages, in which speakers sometimes mimic the facial expression or action of a referent in a narrative (Lillo-Martin, 2012; Quinto-Pozos & Parrill, 2015; Schembri et al., 2005). The main difference is this: While spoken languages are primarily conveyed auditorily and gestural phenomena are mostly visual, so that separation of the linguistic and the gestural is more tractable, in sign languages, both types of communicative expression are conveyed visually by the body, and this has led to confusion in the literature on sign language and on gesture (Sandler, 2022). Because there is a steady stream of visual information in sign languages, it becomes difficult to distinguish the linguistic from the gestural—but it is also imperative to do so, to understand the system. Here we show that motion capture provides an objective measure in this regard.

1.3 Classifier constructions

The sign language stream is enriched by yet a third type of expression—classifier constructions.⁵ These constructions bear some similarity to lexical items, in that the hand is configured in one of a small set of lexically specified shapes. These shapes classify referents according to semantic class

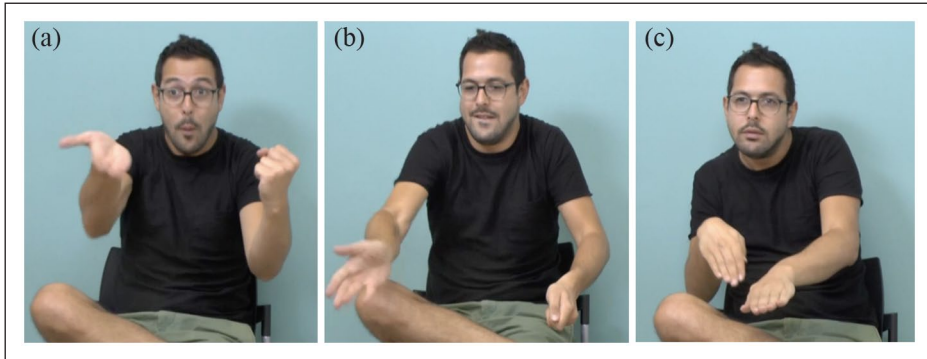


Figure 3. Examples of overt constructed action: (a) “woman beckons the man while holding the cage door open”; (b) “man splashes water (from a container)”; and (c) “lion walks.”

(e.g., human, small animal, and vehicle), size and shape (flat object; cylindrical object; and small, round object), or how the object is handled (handling a thick object; see McDonald, 1982; Schick, 1987; Supalla, 1982; Zwitserlood, 2012). It is because of this classificatory function that these structures are called classifier constructions. Entity and part-entity classifiers are handshapes that represent or partly represent an entity (see Figure 4(a)). For handling classifiers, the configuration of the hand represents the manipulation of the object itself. Size and shape specifiers describe the shape of an object. In other words, the classification is based on three main categories: handshapes which represent objects, handshapes which represent handling of objects, and handshapes which represent visual-geometric characteristics of objects. In all types, the hand, configured as a classifier, can move in different ways, denoting the motion of the referent. As we see in Figure 4, the body can sometimes enact the body of the referent directly, manifesting what Cormier et al. (2015) call reduced or subtle constructed action. It is generally accepted that the handshape classifiers are linguistic entities, listed in the mental lexicon, but the nature and extent of the motions they combine with still require precise definition. We wish to find an objective method for determining whether linguistic handshapes are produced with movements that are explicitly gestural in the sign language modality. It is the nature and extent of these movements that we track instrumentally in this study.

Indeed, some spoken languages such as Cayuga (Mithun, 1986) have classifiers that are similar to the semantic classifiers of sign languages, and still others, like Digueño (Langdon, 1970), have size and shape classifiers. But the mere existence of such morphemes is where the similarity ends. Spoken language classifiers, in languages that have them, are typically affixes bound to words. Sign language classifiers are also bound morphemes that cannot occur alone, but, unlike spoken language classifiers, they combine with locations and movements that are not listed in the lexicon. The motion and location elements to which classifiers are bound cannot be listed in a lexicon because they are not conventionalized, but rather they are gradient and analogous to real-world motion.⁶

The handshapes themselves, then, are discrete and conventionalized, and they differ from sign language to sign language, so that they qualify as linguistic elements. However, the classifier handshapes are produced with a range of movements and locations which represent real-world situations. These motions are not mimicry, like the real-world enactment performed in constructed action. Classifier constructions rather convey motion and location analogically in a scaled-down representation that corresponds to the size of the lexical classifier referents.



Figure 4. Examples of classifier constructions, in which the hand or hands are classifiers for types of entities or objects in the world: (a) “man (cockily) walks up to lion” [hand=classifier for legs of a person]; (b) “lion roars” [hand=classifier for mouth of an animal, here, the lion]; and (c) “man lands with his legs crossed (looking pleased with himself)” [hands=classifiers for individual legs of a person].

A sequence of classifier constructions in American Sign Language (ASL) from Supalla (1978) is shown in Figure 5(a) to (c). Each picture represents a classifier construction, in which each hand is configured as a classifier, but the single VEHICLE handshape classifier persists across propositions, to represent, “a car drives up a winding road, (the car) crosses over a bridge, and (the car) drives past a tree.” In the examples in Figure 5, the body of the signer is neutral, unlike the examples in Figure 4, in which the body manifests partial constructed action. Both Figures 4 and 5 are examples of classifier constructions. Chapters in Emmorey (2013) describe classifier constructions in different sign languages, and Zwitserlood (2012) provides an overview.

In his seminal book about gesture, McNeill (1992) distinguished linguistic from gestural elements in spoken language. According to McNeill (1992), linguistic elements are discrete, conventionalized, and compositional with respect to other linguistic elements, while gestures are gradient, idiosyncratic, and do not enter into compositional expressions with one another. According to these criteria, signed words (lexical) are clearly linguistic, while constructed action is gestural. But what about classifier constructions, whose conventionalized handshapes are listed in the lexicon and differ from sign language to sign language? Figure 6 shows the handshape classifier for VEHICLE in American Sign Language (ASL) and in British Sign Language (BSL) (also a vehicle classifier in Israeli Sign Language).

Studies show that handshapes in classifier constructions constitute a linguistic category while motion and location might reflect influences from gesture (Schembri et al., 2005; Singleton et al., 1993). This was further supported by perception and neuroimaging studies (Emmorey et al., 2003, 2013). Sandler and Lillo-Martin (2006) adopt a similar view, pointing out that classifier constructions with a single handshape classifier can span several intonational phrases, each corresponding to a proposition, as shown in Figure 5. They propose that the lexically specified handshapes are bound morphemes, which only combine with motions and locations post-lexically—outside of the lexicon—to form propositions. However, their account does not deal with the nature of the locations and movements.

Some researchers have claimed that gestural productions are less constrained in terms of size and use of space, compared with lexical productions (Cormier et al., 2012). Classifier constructions and constructed action have been described in a similar way, as “highly variable” (Cormier et al., 2015), “more gestural,” (Cormier et al., 2012), and “less constrained in their movements”

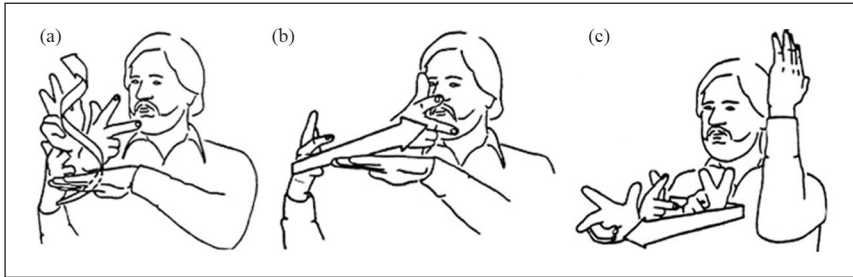


Figure 5. Sequence of classifier constructions reprinted from Supalla (1978) with permission. A car (ASL VEHICLE classifier) drives up a winding road, crosses over a bridge (FLAT OBJECT classifier), drives past a tree (TALL OBJECT classifier).

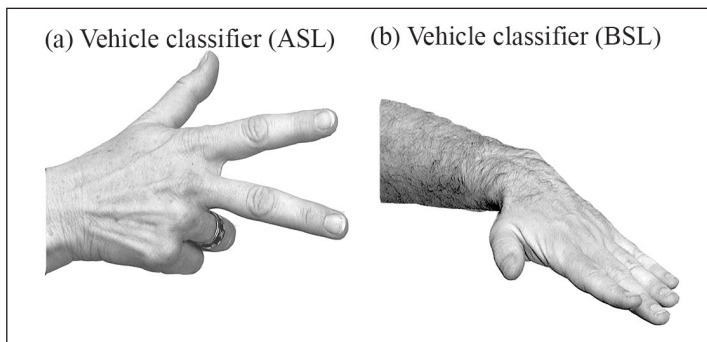


Figure 6. Example of a vehicle classifier in two sign languages: (a) ASL and (b) BSL.

(Sandler & Lillo-Martin, 2006, p. 196; Zwitserlood, 2012). In contrast, when compared with gestures by hearing individuals, lexical signs are smaller and more restricted in handshape (Schembri et al., 2005). They have been described through observation as “highly stable” (Brennan, 1992) with “fairly simple” movements (Jones, 2013). Yet, no study to date has provided any objective and quantifiable measures to support these claims.

How then can we determine whether the movements in classifier constructions are discrete linguistic elements, as Supalla proposed, or belong to a gestural component of language, as other researchers have surmised? In other words, we wish to determine where the gestural leaves off and the linguistic begins in sign language.

Using the objective measure of motion capture, we ask, “How does the motion of classifier constructions compare with the motion of lexical signs?” and “How does the motion of constructed action compare with each of these?” We know that classifier constructions have lexically specified handshapes, while constructed action has no lexical specification whatsoever. But we do not know whether the movement of the hands and body in the two more gestural modes of expression is similar or different to each other, or whether each can be distinguished from the movement in lexical signs. Since all three structures are commonly used in sign languages, objectively determining their kinematic properties is an important step toward understanding their form and distribution in these languages. In this way, we arrive at a better understanding of their behavior in the visual stream of motion that we see in sign languages and provide a more precise baseline for comparison of spoken and signed languages.

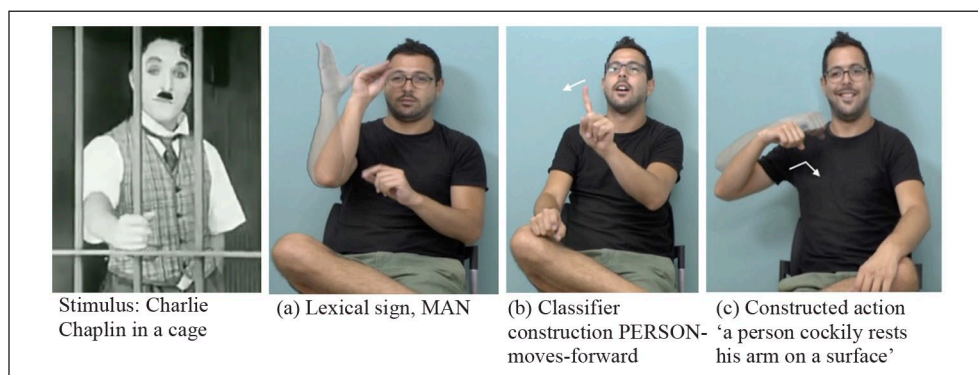


Figure 7. An example from our narrative data of the representation of Charlie Chaplin through the use of a lexical sign, a classifier construction, and constructed action. Stimulus: Charlie Chaplin in a cage; (a) lexical sign, MAN; (b) classifier construction PERSON-moves-forward; and (c) constructed action “a person cockily rests his arm on a surface.”

We see all three kinds of expression in Figure 7, a representation of Charlie Chaplin with (a) lexical sign, (b) a classifier construction, and (c) constructed action.

Other studies focus primarily on the hands, neglecting another crucial aspect of classifier constructions: the motion of the rest of the body and the face. Since the body is instrumental in conveying both linguistic and non-linguistic structure (Sandler, 2012, 2018), it is essential to include the rest of the body and the face in any analysis.

In the conclusion, we discuss how our findings contribute to our understanding of linguistic and gestural properties of sign language, an important distinction for our understanding of the language capacity (Goldin-Meadow, 2003; Goldin-Meadow et al., 2012; Sandler, 2022).

2 The current study

We use motion-capture technology to track the motions of signers’ bodies⁷ as they produce the three different types of expression during a semi-spontaneous language task. We compare the motion of classifier constructions with that of lexical signs, which represent the fully linguistic words of sign languages. Furthermore, we compare the motion of classifier constructions with that of constructed action—arguably the most gestural elements in sign languages (e.g., Quinto-Pozos & Parrill, 2012), in the sense of McNeill (1992). Finally, we will discuss how our findings contribute to the understanding of linguistic and gestural properties of sign language, an important distinction for defining the language capacity (Goldin-Meadow, 2003; Goldin-Meadow et al., 2012; Sandler, 2022).

2.1 Research questions and hypotheses

Until relatively recently, the ability to measure the motion of signs and gestures accurately and objectively has been restricted by limitations in technology. Measuring parameters such as sign size, for example, has relied on subjective measures or other methods (such as projecting a grid onto a signed video output, for example, Hill et al., 2009). In a recent review paper about gesture and sign language, Goldin-Meadow and Brentari (2017) state, “We have good methods for

classifying and measuring handshape, but the techniques currently available for capturing motion are less well developed” (p. 16).

The introduction of motion-capture technology into sign language research was initially greeted with skepticism. Most of the early technologies involved wearing motion detector sensors or body markers such as gloves or sphere markers. This form of invasive measurement was in conflict with the classic methodologies in linguistics that try to reduce the signers’ anxieties in front of the camera. Non-invasive technologies at the time used two-dimensional (2D) computer vision-based tracking via pattern matching and flow algorithms requiring stereo or other structure from motion algorithms to compute three-dimensional (3D) position. These methods were error-prone, tending to lose tracking and required careful calibration and synchronization. The launch of advanced technologies such as Kinect in 2010 (V1) and 2014 (V2), a non-invasive motion tracking device which uses infra-red to detect the body without the use of body markers, was a welcome relief for researchers (e.g., Namboodiripad et al., 2016). Many researchers have since adapted this technology, originally used as a gaming device, for the measure of sign language (Sato et al., 2022) and gesture data (Namboodiripad et al., 2016). In this study, we draw on this technology to compare the kinematic parameters of the three types of signs that are of interest—lexical signs, classifier constructions, and constructed action.

3 Method

3.1 Grammar of the body

Several earlier studies point to the importance of attending specifically to parts of the body in understanding sign language structure (Dachkovsky et al., 2018; Meir et al., 2010; Sandler, 2012a, 2018; Sandler et al., 2011), investigating what Sandler (2018) calls the Grammar of the Body. Here, we look specifically at kinematic parameters and the contribution of different articulators in the production of different sign types.

3.2 Participants and task

Fifteen participants were recruited for this study. All participants are deaf native signers whose preferred language is Israeli Sign Language (ISL). Participants performed a series of tasks, one of which is analyzed as part of this study: the narrative retelling task. Participants were asked to watch two silent movie clips and to retell the stories to their interlocutor, another deaf ISL signer matched for age. The aim of the first retelling was to acquaint signers with the task procedure and to allow them to relax in front of the cameras. Only the narrative elicited from the second retelling was analyzed. The clip stimulus was the *Lion’s Cage* (3 min 23 s), excerpted from a black-and-white Charlie Chaplin movie, in which Chaplin finds himself trapped in a lion’s cage with a real lion. In his attempt to free himself, Chaplin interacts with several different characters and engages in a number of actions. To elicit a detailed narrative, signers were informed that their interlocutors would complete a comprehension task after their retelling which involved ordering five movie stills in chronological order of the events as they were described. To avoid issues related to memory, signers were first shown the whole clip from start to finish and then they were shown the clip in parts and asked to retell the story after they watched each part. The story was split into five parts (each clip segment lasted 37, 52, 45, 31, and 30 s respectively). After all the parts were retold and participants had fully internalized the story, they were asked to retell the whole story from start to finish. This final retelling was analyzed as part of this study. The motions of each signer were tracked using a Kinect motion tracking camera, described in Section 3.3. Two participants were

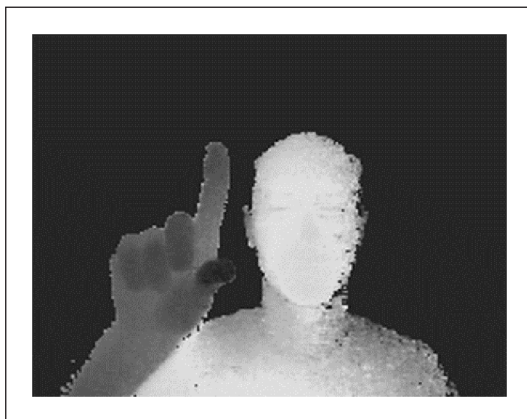


Figure 8. A three-dimensional point cloud (depth image) captured by the Kinect V2.

excluded from the analysis as the motion-capture data were unanalysable. Of the 13 participants remaining, there were seven females and six males ($M_{\text{age}} = 47.5$ years). Narrative retellings varied from 1 min 23 s to 4 min 27 s ($M = 2$ min 42 s). A total of 36 min 17 s were analyzed.

3.3 Kinect motion tracking technology

To track motions of the signers, a Kinect Version 2 (V2) motion tracking system was implemented. The Kinect V2 (Microsoft, 2018) is a portable 3D camera capable of recording depth information using Time of Flight technology (Foix et al., 2011; Hansard et al., 2013). The camera captures a video stream of 3D point clouds (often referred to as 2.5D images or depth images) in which every pixel value represents the distance from the camera. Figure 8 shows an example of one such captured frame in which the brighter the pixels, the greater the distance from the camera.

In addition, when a participant is recorded, the system supplies a skeleton representation of the participant computed per frame from the point cloud (Shotton et al., 2011). The skeleton data contains 25 major skeleton joints of the human body, connected by line segments. Figure 9 shows the layout of the skeleton with the joints labeled. For every frame, the system outputs the 3D location of each of the 25 skeleton joints (a triplet x, y, z in meters, given in the camera's frame of reference). In this study, the signers were positioned 2 m from the Kinect camera and were recorded while signing in a sitting position. Both an RGB image and the skeleton data were captured per frame. Figure 10 shows an example.⁸

We developed a specialized code to collect and analyze the acquired skeletons. The Kinect's sampling rate is 30 frames per second. This sampling rate was achieved by using a strong laptop which supports all Kinect requirements. In addition, we used a recording platform developed in our lab which buffered the Kinect-recorded frames to allow for any delays in writing the recorded files onto the computer. Therefore, frames were never skipped due to writing or memory lags. Kinect resolution for the depth image is 512×424 pixels with field of view $60^\circ \times 70^\circ$. Depth range is between 0.5 and 4.5 m when the camera is placed 2 m from the subject well within range. To overcome noise and jitter in the tracked skeleton joints, we considered both the confidence and the joint's movement between frames. If the confidence of a point was below threshold or a large difference in interframe joint position was detected, then these points in these frames were removed (these missing points were not taken into account when computing the extracted features).

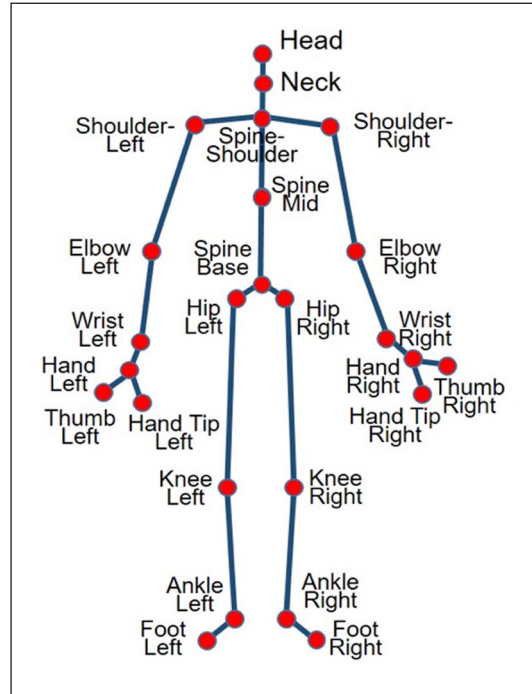


Figure 9. Skeleton of human subject consisting of 25 points.

From the skeleton data, we extracted spatio-temporal features associated with the signing in our analysis. The full recording session was dissected into segments, consisting of a single sign or action. For each segment to be analyzed, a sequence of skeletons (frames) was captured. Various measurements were computed per skeleton frame and then combined to produce a set of measurements representing the whole segment. Prior to computing the measurements, the skeletons were normalized to a standard size using the method in Weibel et al. (2016) to eliminate size effects. In the following, we briefly describe the collected measurements. For a skeleton joint p , we track its 3D location per frame i : $p_i = (x_p, y_p, z_i)$ and compute the following measures across the N frames of the segment:

- *Duration* (D)⁹—the time elapsed in seconds between the first and last frame in the sequence.
- *Distance covered* ($DW(p)$)—the distance in meters, traversed by the skeleton joint p during the sequence. It is computed by summing the distance between positions of the joint in consecutive frames.
- *Average speed* ($S(p)$)—computed as the Distance covered by p divided by the duration: $S(p) = (DW(p)) / D$
- *Variance* (std squared— $\sigma^2(p)$) of Location—the variance of the 3D position of the joint p across the sequence. The variance is computed by calculating the mean squared distance between the position in each frame to the average 3D position, across the sequence.
- *Volume* ($V(p)$)—the 3D volume of the space traversed by joint p during the sequence. It is computed by determining the volume of the convex bounding polygon of all 3D locations p_i of the joint throughout the sequence.

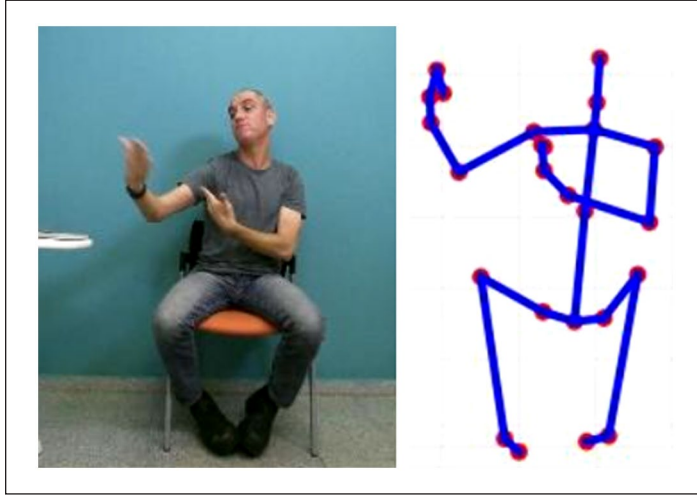


Figure 10. RGB image of signer (left) and the corresponding 3D skeleton (right).

- *Mean distance from body plane* ($DB(p)$)—the distance of the skeleton joint from the body plane. The body plane is computed as the 2D plane spanned by the three skeleton joints, namely, shoulder-left, shoulder-right and spine-base (see Figure 9); the plane normal (Pl) is then given by

$$Pl = \frac{(p^6 - p^0) \times (p^8 - p^0)}{\|p^6 - p^0\|_2 \|p^8 - p^0\|_2}.$$

The distance from joint p to the body plane is measured along the line perpendicular to the plane and passing through the joint (see Figure 11):

$$DB(p) = \left| (p - p^1) \circ Pl \right|$$

Results were analyzed using these measures collected for three joints of the skeleton: Head movement was analyzed using the head point; the hands were analyzed using the signer's dominant hand (right or left); and the torso was analyzed using the spine-shoulder point at the top of the spine (see Figure 11).

3.4 Data coding

A total of 2,430 individual sign productions were parsed from the narratives and categorized into one of three sign types: lexical signs, classifier constructions, and constructed action, according to accepted criteria as described in Section 1. The first frame of each sign was determined based on the moment at which the target handshape and orientation occurred. In turn, the end of the sign was the last frame displaying the target sign handshape and orientation. The first author, a non-native hearing signer, together with a deaf native ISL signer, extracted all of the signs from the data. Fingerspellings,¹⁰ pointing signs, and any signs comprising less than two frames were removed ($n=98$). To further validate the findings, we removed any results which were 2 std above or below the average for any parameter and articulator ($n=390$).

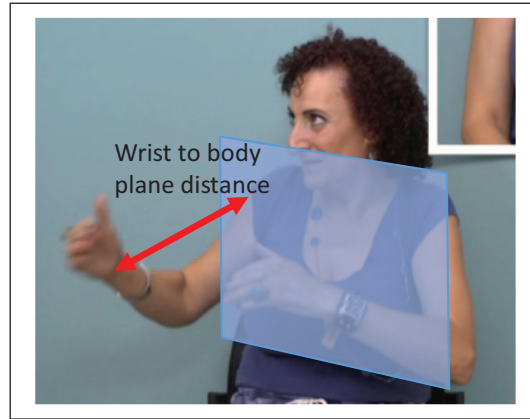


Figure 11. Distance from skeleton joint to body plane, for example: wrist joint to body plane.

All recorded sequences were processed by our algorithm to calculate a set of motion-related measures. Specifically, the recorded 3D skeleton data collected by Kinect was analyzed, and the following kinematic parameters were computed: duration, distance covered, speed, position variance, volume, and mean distance from body plane (see Section 3.3).

3.5 Data analysis

We conducted six different linear mixed models for each dependent parameter separately to determine the discriminability between the groups. Separate analyses were also conducted for each articulator—the hand, the head, and the torso—and therefore in total 18 linear mixed models were performed. For each, group (LS, CC, and CA) was included as the fixed effect, with participants as random intercepts. One of the assumptions of the model is that the residual error is normally distributed. Since the model did not meet this assumption for some parameters, these were transformed into a log-normal distribution. Post hoc tests, adjusted for multiple comparisons with Bonferroni corrections, were run to explain the differences between the groups.

4 Results

As part of our study, we were interested in the relationship between each kinematic parameter and group (i.e., lexical signs, classifier constructions, and constructed action). In other words, are lexical signs smaller in volume? Are examples of constructed action shorter in duration?

4.1 Statistical analyses

Table 1 shows the results of the statistical analyses of 18 separate linear mixed models. The F and p values indicate how well the model can discriminate across the groups for each parameter separately.

4.1.1 Hand. We performed post hoc analyses to explain the differences, using Bonferroni adjustment (due to multiple comparisons). The following results were found for the different motion features associated with the hand joint:

Table 1. Group Effect Results.

	Hand	Head	Torso
Duration	$F(2, 2353) = 170.63$, $p < .0001$	$F(2, 2411) = 273.05$, $p < .0001$	$F(2, 2411) = 273.05$, $p < .0001$
Distance covered	$F(2, 2353) = 137.19$, $p < .0001$	$F(2, 2411) = 249.84$, $p < .001$	$F(2, 2411) = 211.73$, $p < .0001$
Speed	$F(2, 2353) = 3.57$, $p = .0284$	$F(2, 2411) = 46.22$, $p < .0001$	$F(2, 2411) = 35.85$, $p < .0001$
Variance	$F(2, 2353) = 28.82$, $p < .0001$	$F(2, 2411) = 227.57$, $p < .0001$	$F(2, 2411) = 205.89$, $p < .0001$
Volume	$F(2, 2171) = 96.47$, $p < .0001$	$F(2, 2220) = 208.05$, $p < .0001$	$F(2, 2220) = 196.68$, $p < .0001$
Mean distance from body plane	$F(2, 2353) = 39.91$, $p < .0001$	$F(2, 2411) = 8.9$, $p = .0001$	$F(2, 2411) = 8.94$, $p < .0001$

- *Duration*: Lexical signs are significantly shorter in duration than CL, $t(2356) = 12.48$, $p < .0001$, and CA, $t(2356) = 16.19$, $p < .0001$; however, CL and CA are not significantly different from one another.
- *Distance covered*: Lexical signs are significantly shorter in distance covered than CL, $t(2356) = 12.18$, $p < .0001$, and CA, $t(2356) = 13.80$, $p < .0001$; however, CL and CA are not significantly different from one another.
- *Speed*: Lexical signs are significantly faster than CL, $t(2356) = 12.18$, $p < .0001$; however, there was no significant difference between lexical signs and CA, and CL and CA.
- *Variance*: Lexical signs are significantly smaller than CL, $t(2356) = 5.4$, $p < .0001$; lexical signs are significantly distinct from CA, $t(2353) = -3.9$, $p = .0003$, in which lexical signs are larger than CA, and CL is distinct from CA, in which CL is larger than CA, $t(2356) = -7.58$, $p < .0001$. CA > lexical > CL.
- *Volume*: Lexical signs are significantly smaller than CL, $t(2171) = 9.88$, $p < .0001$, and CA, $t(2171) = 11.94$, $p < .0001$; however, CL and CA are not significantly different from one another.
- *Mean distance from body plane*: Lexical signs are significantly closer than CL to the body, $t(2353) = 6.23$, $p < .0001$, and CA, $t(2353) = 7.7$, $p < .0001$; however, CL and CA are not significantly different from one another.

4.1.2 Head. The following results were found for the different motion features associated with the head joint:

- *Duration*: Lexical signs are significantly shorter than CL, $t(2411) = 15.21$, $p < .0001$, and CA, $t(2411) = 20.89$, $p < .0001$; however, CL and CA are not significantly different from one another.
- *Distance covered*: Lexical signs are significantly shorter than CL, $t(2411) = 16.09$, $p < .0001$, and CA, $t(2411) = 18.94$, $p < .0001$; however, CL and CA are not significantly different from one another.
- *Speed*: Lexical signs are significantly faster than CL, $t(2411) = 8.19$, $p < .0001$, and CA, $t(2411) = 6.86$, $p < .0001$; however, CL and CA are not significantly different from one another.

- *Variance*: Lexical signs are significantly smaller than CL, $t(2411)=14.69, p < .0001$, and CA, $t(2411)=18.56, p < .0001$; however, CL and CA are not significantly different from one another.
- *Volume*: Lexical signs are significantly smaller than CL, $t(2220)=14.79, p < .0001$, and CA, $t(2220)=17.40, p < .0001$; however, CL and CA are not significantly different from one another.
- *Mean distance from body plane*: Lexical signs are significantly closer than CL to the body, $t(2411)=2.43, p = .0461$, and CA, $t(2411)=3.94, p = .0003$; however, CL and CA are not significantly different from one another.

4.1.3 *Torso*. The following results were found for the different motion features associated with the torso joint:

- *Duration*: Lexical signs are significantly shorter than CL, $t(2411)=15.21, p < .0001$, and CA, $t(2411)=20.89, p < .0001$, and CL and CA are significantly different from one another, $t(2411)=2.89, p = .0118$.
- *Distance covered*: Lexical signs are significantly shorter than CL, $t(2411)=15.94=09, p < .0001$, and CA, $t(2411)=17.21, p < .0001$; however, CL and CA are not significantly different from one another.
- *Speed*: Lexical signs are significantly faster than CL, $t(2411)=7.52, p < .0001$, and CA, $t(2411)=5.59, p < .0001$; however, CL and CA are not significantly different from one another.
- *Variance*: Lexical signs are significantly smaller than CL, $t(2411)=14.22, p < .0001$, and CA, $t(2411)=17.48, p < .0001$; however, CL and CA are not significantly different from one another.
- *Volume*: Lexical signs are significantly smaller than CL, $t(2220)=14.55, p < .0001$, and CA, $t(2220)=16.78, p < .0001$; however, CL and CA are not significantly different from one another.
- *Mean distance from body plane*: Lexical signs are significantly closer than CL to the body, $t(2411)=2.63, p = .0254$, and CA, $t(2411)=3.84, p = .0004$; however, CL and CA are not significantly different from one another.

Overall, it was found that lexical signs for all articulators are shorter in duration and distance covered, faster in speed, smaller in variance and volume, and closer to the body, than classifier constructions and constructed action. From comparing the F values (see Table 1), we can see that duration and distance covered are two of the most important parameters for predicting group effects. Speed and mean distance from the body plane were the least important parameters. A summary of the results is presented in Table 2.

5 Discussion

Our findings reveal kinematic distinctions among all three categories: lexical signs, classifier constructions, and constructed action. First, lexical signs are significantly shorter in duration and distance covered, faster in speed, smaller in variance and volume, and closer to the body, compared with both classifier constructions and constructed action. Second, classifier constructions and constructed action pattern the same on these measures. In other words, though they differ in linguistic structure (the hands represent discrete morphemes in classifier constructions only), the classifier

Table 2. Summary of Results Showing All Significant and Non-Significant Results (NS = Non-Significant) and the Difference Across the Three Categories: Lexical Signs, Classifier Constructions, and Constructed Action.

	Articulator	Lexical sign	Classifier construction	Constructed action
Duration	Hand	Shorter		Longer
	Head	Shorter		Longer
	Torso	Shorter		Longer
Distance covered	Hand	Shorter		Longer
	Head	Shorter		Longer
	Torso	Shorter		Longer
Speed	Hand	Faster	Slower	NS
	Head	Fastest		Slowest
	Torso	Fastest		Slowest
Variance	Hand	Smaller	Largest	Smallest
	Head	Smallest		Largest
	Torso	Smallest		Largest
Volume	Hand	Smallest		Larger
	Head	Smaller		Larger
	Torso	Smaller		Larger
Mean distance from body plane	Hand	Closest		Furthest
	Head	Closest		Furthest
	Torso	Closest		Furthest

constructions in our data are *kinematically* similar to constructed action. Differences are in the degree, so that constructed action is more extreme in our measures compared with classifier constructions. Finally, the hands are not the only reliable articulator for distinguishing lexical signs from both classifier constructions and constructed action; head and body show the same distinctions. This finding has interesting implications for sign language structure and motion capture, as elaborated in the conclusion.

Several researchers have stressed the difficulties in defining different kinds of signing (e.g., Cormier et al., 2012, 2015; Lillo-Martin, 2012). This study provides some initial kinematic criteria which may be used in distinguishing lexical signs, classifier constructions, and constructed action: Six kinematic parameters differed consistently across sign types: duration, distance covered, speed, variance, volume, and mean distance from the body plane. Future studies may draw on these parameters to distinguish lexicalized signs from examples of classifier constructions and constructed action.

We found that the motions of classifier constructions pattern in a similar way to those of constructed action, and that these patterns reflect properties described as gestural in earlier literature, such as larger variance and volume (Cormier et al., 2012, 2015; Sandler & Lillo-Martin, 2006). In other words, classifier constructions and constructed action are longer in distance covered and duration, slower in speed, more variable and larger in size, and further from the body when compared with lexical signs in our data set. Lexical signs, in contrast, are shorter in distance covered and duration, faster in speed, less variable and smaller in size, and closer to the body, providing objective support for claims in the literature that were based mainly on observation. On these measures, as noted, constructed action is more extreme than classifier constructions in our data, though not statistically distinct. We do not seek kinematic differences in handshape in the forms we

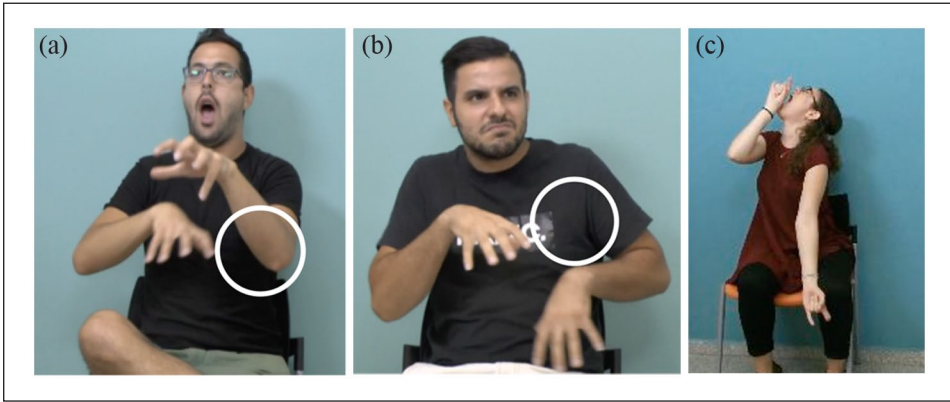


Figure 12. Example of the lexical sign for “lion” shown in (a) along with two examples of constructed action in (b) and (c): (a) Lexical sign, LION; (b) constructed action, “lion walks”; and (c) classifier construction, “lion opens his jaw wide.”

studied. Linguistic analyses distinguish them, showing that classifier handshapes are discrete and lexically listed, as opposed to handshapes in constructed action, which vary analogically according to the action conveyed. This supports previous claims that the movement parameter in classifier constructions should be analyzed as gestural in nature (Schembri et al., 2005).

The data set combined multiple different examples of lexical signs, classifier constructions, and constructed action; however, we see similar patterns when we compare the lexical sign LION with the classifier construction depicting the lion’s jaw to the constructed action rendition of a lion walking (see Figure 12), the closest equivalent in terms of conceptual similarity. Figure 12(a) shows the lexical sign for “lion” which has evolved from an iconic representation of the lion’s paws, Figure 12(b) is the constructed action of the lion walking, in which the signer’s hands represent the lion’s paws, and in Figure 12(c) the signer uses a classifier to depict the jaws of the lion. In the constructed action and the classifier construction, the signer is fully enacting the lion, displacing the head and torso, in different ways (see Figure 12(b) and (c)), and adopting mimetic facial expressions. In contrast, we see that the lexical sign (Figure 12(a)) is produced in the neutral signing space with a neutral facial expression and relaxed and upright head and body position. The shape of the hands may be iconically motivated—the paws of the lion—but they are lexically specified together with the location and movement to mean “lion.” As aforementioned, the mouth shape in 12a is not mimetic, as it is in (b) and (c), but rather shows the mouthing for the Hebrew word meaning lion, [arye], accompanying the lexical sign. The mouth shapes in (b) and (c) represent the mouth of the lion, and the rest of the body articulators are similarly recruited to enact the lion walking. A descriptive analysis of these three examples (including 87 lexical signs meaning “lion,” 30 examples of constructed action of a lion, and 22 examples depicting a lion using a classifier construction) reveals the same pattern kinematically—the lexical sign for “lion” is systematically shorter, smaller, and faster than the examples of constructed action and classifier constructions.

Examples shown above in Figure 4 and repeated for convenience in Figure 13¹¹ indicate that classifier constructions are often simultaneously accompanied by partial constructed action: mimetic action of the face and head. This finding suggests that other forms of constructed action (see footnote 15), combine simultaneously with classifier constructions on the hands. This simultaneous use of handshape classifiers and partial constructed action on other parts of the body is a common phenomenon in sign languages (Dudis, 2004; Quinto-Pozos & Parrill, 2012).

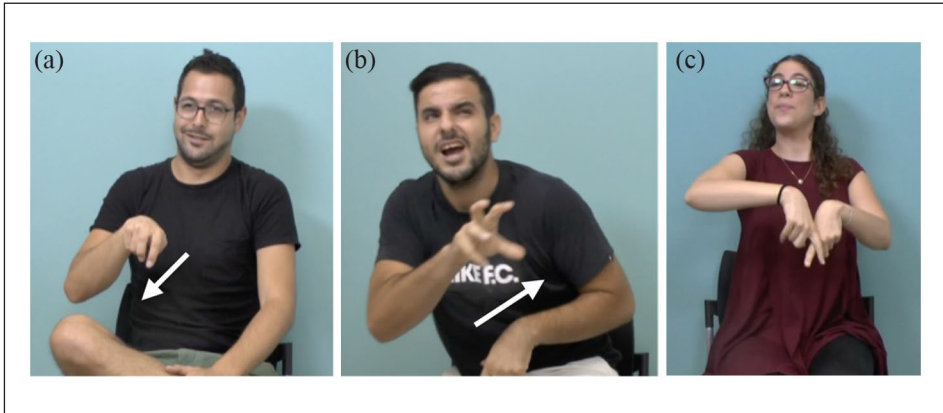


Figure 13. Three examples of a classifier construction on the hands accompanied by constructed action with the face, head, and/or torso (same as Figure 4): (a) “man (cockily) walks up to lion” [hands = classifiers for legs of a person]; (b) “lion roars” [hand = classifier for mouth of an animal, here, the lion]; and (c) “man lands with his legs crossed (looking pleased with himself)” [hands = classifiers for individual legs of a person].

Three subtypes of constructed action (overt, reduced, and subtle) were compared by Jantunen and colleagues (2020) using motion-capture technology. Overt constructed action, in which the signer is fully in character role and their articulators reflect the actions of the character in a direct way, is comparable with the examples of overt constructed action in this study (see Figure 12(b)). In reduced constructed action, the signer takes on the role of character and narrator, and in contrast, in subtle constructed action, there is minimal character perspective and fewer articulators involved in enacting. Using the Qualisys Motion Tracking system, they found that the head and the torso move faster and over a larger area in reduced and overt constructed action than in subtle constructed action or regular narration. Assuming that regular narration in their study is comparable with lexical signs in this study, the results are very similar. Because their findings show distinct differences only between the extremes (between regular narration and reduced/overt constructed action, and subtle constructed action and overt constructed action), they concluded that a three-way distinction for constructed action is not supported (Jantunen et al., 2020).

Although the findings reveal some interesting and quantifiable differences between sign types, there are a number of limitations to this study. The use of the Kinect motion capture provides objective and quantitative data and facilitates acquisition of quantitative measures; however, it is a specialized depth camera and not as ubiquitous as the standard RGB video camera. New technology can now provide 3D tracking information from 2D videos (e.g., OpenPose, Cao et al., 2019; Mediapipe, Lugaresi et al., 2019), thus implying that tracking technology may revert back to 2D video captures. Although tracking is found to be very good in the image plane, depth information and accuracy is still lacking using these technologies. It is questionable, and open to further study, whether depth information is indeed necessary for sign language analysis or whether 2D tracking (with loss of depth) can suffice.

6 Conclusion

To the naïve observer, signing seems to be a constant stream of movement. From a purely linguistic point of view, some of these movements are part of the lexical description of the signed words, and some are conventionalized prosodic signals. But it is now widely acknowledged that there are also

gestural elements in signing (Liddell, 2003; Müller, 2018; Perniss et al., 2015), and our study uses a novel approach that allows for gestural influence in distinguishing and characterizing the types of movement that we observe. With a special focus on constructions that are only partially linguistic, classifier constructions, we ask whether their movements pattern like those of linguistic sign words or similar to the fully gestural movements of pantomime-like constructed action.

By using motion-capture technology to compare the motion of classifier constructions with that of lexical signs and with constructed action, we find the answer: The motion of classifier constructions patterns similar to overt examples of constructed action, and unlike the motion of lexical signs. Lexical signs are shorter in duration and distance, faster, smaller in variance and volume, and closer to the body compared with examples of classifier constructions and constructed actions. This is the case even though lexical signs participate in sentences that involve a fair amount of conventionalized bodily motion (Sandler, 2018), but that conventionalized motion is found here to be much more rule-governed and constrained.

The findings from this study provide a useful foundation for future studies incorporating motion-capture technology into research on sign language and gesture. They also have significant implications for sign language linguistics. We adopt the noncontroversial insight that overt constructed action, also referred to as enactment and pantomime, is gestural rather than linguistically structured. The fact that the motion of classifier constructions patterns similarly shows that this aspect of the construction is also gestural in nature, providing objective and quantified instrumental support for observational claims that appear in the literature. In this way, we confirm that the word is identifiable as a universal construct in all languages, regardless of the physical channel in which they are conveyed. This finding is the first to provide instrumental support for other studies which make the claim that classifier constructions are gestural in nature (e.g., Schembri et al., 2005). The identification of classifier constructions as gestural has important ramifications for other areas of sign language linguistics, including understanding the mental lexicon, the gradience of classifier construction production, and acquisition of classifier constructions.

Our findings about hybrid classifier constructions underscore a property of sign languages that we consider fascinating and unique to these languages: The visual and bodily nature of expression in sign languages unifies linguistic and gestural elements in a way that spoken languages cannot. They reveal, as only sign languages can, the flexibility and richness of our capacity for language.

Acknowledgements

Thanks to Shai Davidi for video and other technical assistance, to Shiri Barnhart for her administrative help, and to Debi Menashe for her ISL expertise. Special thanks go to all of deaf participants who were involved in our study.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme, Grant Agreement No. 340140.

ORCID iD

Rose Stamp  <https://orcid.org/0000-0002-1993-559X>

Notes

1. For detailed phonological models, see Brentari (1998), Liddell and Johnson (1989), Perlmutter (1992), and Sandler (1989, 2010).

2. Battison (1978) demonstrated that orientation of the hand can be contrastive, adding orientation as parameter. However, its status as a separate major category has been challenged by evidence that orientation is dependent on other major category features (Crasborn & van der kooji, 1997; Sandler, 1987, 1989; Sandler & Lillo-Martin, 2006).
3. The mouth can also manifest iconic gestures for concepts like “narrow” and “round” (Sandler, 2009).
4. Lexical signs are presented with English glosses shown in capital letters. The gloss represents the nearest equivalent English word for the sign meaning. For example, the gloss LION refers to a specific sign which represents the concept of a “lion.”
5. The term “classifier” was first introduced into the description of sign languages by Frishberg (1975), but these forms have gone by a number of different names in the sign language literature: classifier constructions, verbs of motion and location, verbal predicates, lexical verbs, noun incorporation, classifier predicates, and depicting verbs (Liddell, 2003; Sandler & Lillo-Martin, 2006; Schick, 1987; Supalla, 1982; Zwitserlood, 2012).
6. In his original treatment of classifiers in American Sign Language (Supalla, 1982, 1986), Ted Supalla argued that locations and motions are indeed linguistic elements. Most subsequent work following from his seminal analysis departs from this view.
7. In this study, we do not track facial expression, which, though certainly important in distinguishing the types of structures we are interested in (Baker-Shenk, 1983; Dachkovsky, 2018), requires different technology.
8. Sometimes, Kinect experiences jitter (noise) at skeleton joints measured as std is between 0.5 and 2 cm (see, for example, Tölgyessy et al., 2021); however, the higher jitter is at the lower limbs (knees and ankles) and at larger distances from the camera. In this study, our signers sat at about 2 m from the camera and only the joints of the upper part of the body were considered and, therefore, the jitter was in the low range of the noise spectrum.
9. Duration and distance covered were both found to be important indicators of sign type. There is a relationship between these parameters: When the duration is longer, there is a higher chance that the distance covered will be longer. The same is true for sign variance and volume. Despite the strong relationship between these two motion measures, it is possible for one to be independent of the other.
10. Fingerspelling is a non-native part of the sign language lexicon which is borrowed from the surrounding spoken language. Fingerspellings are handshapes which represent each individual letter of a spelled word in the ambient spoken language. For example, the name “Yael” is fingerspelled in ISL with each individual letter in the Hebrew alphabet represented by a different handshape.
11. Since we coded overt constructed action only if *all* articulators were re-enacting the actions of the referent, none of the examples shown in Figure 13 are interpreted as overt constructed action in our study.

References

- Baker-Shenk, C. (1983). *A microanalysis of the nonmanual components of questions in American Sign Language*. University of California. <https://www.semanticscholar.org/paper/A-Microanalysis-of-the-Nonmanual-Components-of-in-Baker-Shenk/3dc31646436bf43f7c35ee84c51b860028e8949d>
- Battison, R. (1978). *Lexical borrowing in American Sign Language*. Linstok Press.
- Bloomfield, L. (1933). *Language*. Holt, Rinehart & Winston.
- Brennan, M. (1992). An introduction to the visual world of BSL. In D. Brien (Ed.), *Dictionary of British Sign Language/English* (pp. 1–133). Faber & Faber.
- Brentari, D. (1998). *A prosodic model of sign language phonology*. MIT Press.
- Cao, Z., Martinez, G. H., Simon, T., Wei, S., & Sheikh, Y. A. (2019). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43, 172–186.
- Cormier, K., Quinto-Pozos, D., Sevcikova, Z., & Schembri, A. (2012). Lexicalisation and de-lexicalisation processes in sign languages: Comparing depicting constructions and viewpoint gestures. *Language & Communication*, 32(4), 329–348. <https://doi.org/10.1016/j.langcom.2012.09.004>
- Cormier, K., Smith, S., & Sevcikova-Sehyr, Z. (2015). Rethinking constructed action. *Sign Language & Linguistics*, 18(2), 167–204.

- Coulter, G. (1982, January 8–11). *On the nature of ASL as a monosyllabic language* [Paper presentation]. Paper Presented at the Annual Meeting of the Linguistic Society of America, San Diego, CA, United States.
- Crasborn, O., & van der kooji, E. (1997). Relative orientation in sign language phonology. *Linguistics in the Netherlands*, 4(1), 37–48.
- Dachkovsky, S. (2018). *Grammaticalization of intonation in Israeli Sign Language: From information structure to relative clause relations* [Doctoral thesis, University of Haifa].
- Dachkovsky, S., Stamp, R., & Sandler, W. (2018). Constructing complexity in a young sign language. *Frontiers in Psychology*, 9, Article 2202. <https://doi.org/10.3389/fpsyg.2018.02202>
- Dudis, P. G. (2004). Body partitioning and real-space blends. *Cognitive Linguistics*, 15(2), 223–238.
- Emmorey, K. (Ed.). (2013). *Perspectives on classifier constructions in sign languages* (1st ed.). Psychology Press.
- Emmorey, K., McCullough, S., & Brentari, D. (2003). Categorical perception in American Sign Language. *Language and Cognitive Processes*, 18(1), 21–45. <https://doi.org/10.1080/01690960143000416>
- Emmorey, K., McCullough, S., Mehta, S., Ponto, L. L. B., & Grabowski, T. J. (2013). The biology of linguistic expression impacts neural correlates for spatial language. *Journal of Cognitive Neuroscience*, 25(4), 517–533. https://doi.org/10.1162/jocn_a_00339
- Foix, S., Alenya, G., & Torras, C. (2011). Lock-in Time-of-Flight (ToF) cameras: A survey. *IEEE Sensors Journal*, 11(9), 1917–1926.
- Frishberg, N. (1975). Arbitrariness and iconicity: Historical change in American Sign Language. *Language*, 51, 696–719.
- Goldin-Meadow, S. (2003). *The resilience of language: What gesture creation in deaf children can tell us about how all children learning language*. Psychology Press.
- Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *The Behavioral and Brain Sciences*, 40, Article e46. <https://doi.org/10.1017/S0140525X15001247>
- Goldin-Meadow, S., Shield, A., Lenzen, D., Herzig, M., & Padden, C. (2012). The gestures ASL signers use tell us when they are ready to learn math. *Cognition*, 123, 448–453.
- Hansard, M., Lee, S., Choi, O., & Horaud, R. P. (2013). *Time-of-Flight cameras: Principles, methods and applications*. Springer. <https://doi.org/10.1007/978-1-4471-4658-2>
- Hill, J., McCaskill, C., Lucas, C., & Bayley, R. (2009). *Signing outside the box: The size of signing space in Black ASL* [Paper presentation]. 38th New Ways in Analysing Variation (NWAV38). Ottawa, Ontario, Canada.
- Jantunen, T., De Weerd, D., Burger, B., & Puupponen, A. (2020). The more you move, the more action you construct: A motion capture study on head and upper-torso movements in constructed action in Finnish Sign Language narratives. *Gesture*, 19(1), 72–96. <https://doi.org/10.1075/gest.19042.jan>
- Jones, S. P. (2013). *Classifier constructions as procedural signs in American Sign Language* [Master's thesis, University of North Dakota].
- Langdon, M. (1970). *A grammar of Digueño. The Mesa Grande dialect*. University of California Press.
- Liddell, S. K. (2003). *Grammar, gesture and meaning in American Sign Language*. Cambridge University Press.
- Liddell, S. K., & Johnson, R. E. (1989). American Sign Language: The phonological base. *Sign Language Studies*, 64, 195–278.
- Lillo-Martin, D. (2012). Utterance reports and constructed action. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign language: An international handbook* (pp. 365–388). De Gruyter.
- Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M., Lee, J., Chang, W.-T., Hua, W., Georg, M., & Grundmann, M. (2019, June 17). *MediaPipe: A framework for perceiving and processing reality* [Paper presentation]. Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, United States. https://mixedreality.cs.cornell.edu/s/NewTitle_May1_MediaPipe_CVPR_CV4ARVR_Workshop_2019.pdf
- McDonald, B. H. (1982). *Aspects of the American sign language predicate system*. PhD Dissertation, University of Buffalo.

- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago Press.
- Meir, I., Sandler, W., Padden, C. A., Aronoff, M., Marschark, M., & Spencer, P. E. (2010). Emerging sign languages. In M. Marschark & P. E. Spencer (Eds.), *Oxford handbook of deaf studies, language, and education* (Vol. 2). Oxford University Press.
- Microsoft. (2018). *Kinect for xbox one* [Computer software]. <https://www.xbox.com/en-us/>
- Mithun, M. (1986). The convergence of noun classification systems. In C. Craig (Ed.), *Noun classes and categorization* (pp. 379–397). Benjamins.
- Müller, C. (2018). How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture, 16*(2), 277–304.
- NambooDiripad, S., Lenzen, D., Lepic, R., & Verhoef, T. (2016). Measuring conventionalization in the manual modality. *Journal of Language Evolution, 1*(2), 109–118. <https://doi.org/10.1093/jole/lzw005>
- Perlmutter, D. M. (1992). Sonority and syllable structure in American Sign Language. *Linguistic Inquiry, 23*(3), 407–442.
- Perniss, P., Özyürek, A., & Morgan, G. (2015). The influence of the visual modality on language structure and conventionalization: Insights from sign language and gesture. *Topics in Cognitive Science, 7*(1), 2–11. <https://doi.org/10.1111/tops.12127>
- Quinto-Pozos, D., & Parrill, F. (2012, July 24–27). *Comparing viewpoint strategies used by co-speech gesturers and signers* [Paper presentation]. 5th Conference of the International Society for Gesture Studies, Lund, Sweden.
- Quinto-Pozos, D., & Parrill, F. (2015). Signers and co-speech gesturers adopt similar strategies for portraying viewpoint in narratives. *Topics in Cognitive Science, 7*(1), 12–35.
- Sandler, W. (1987). Assimilation and feature hierarchy in ASL. In A. Bosch, B. Need, & E. Schiller (Eds.), *Chicago linguistics society parasession on autosegmental phonology* (pp. 266–278). Chicago Linguistic Society.
- Sandler, W. (1989). *Phonological representation of the sign*. Foris.
- Sandler, W. (1993). A sonority cycle in American Sign Language. *Phonology, 10*(2), 243–279.
- Sandler, W. (1996). Phonological features and feature classes: The case of movements in sign language. *Lingua. International Review of General Linguistics. Revue Internationale de Linguistique Generale, 98*(1–3), 197–220. https://www.academia.edu/64463666/Phonological_features_and_feature_classes_The_case_of_movements_in_sign_language
- Sandler, W. (1999). Cliticization and prosodic words in a sign language. In T. Hall & U. Kleinhenz (Eds.), *Current studies in linguistic theory: Studies on the phonological word* (pp. 223–255). John Benjamins.
- Sandler, W. (2009). Symbiotic symbolization by hand and mouth in sign language. *Semiotica, 174*(1/4), 241–275.
- Sandler, W. (2010). Prosody and syntax in sign languages. *Transactions of the Philological Society, 108*(3), 298–328. <https://doi.org/10.1111/j.1467-968X.2010.01242.x>
- Sandler, W. (2012a). Dedicated gestures in the emergence of sign language. *Gesture, 12*(3), 265–307.
- Sandler, W. (2012b). The phonological organization of sign languages. *Language and Linguistics Compass, 6*(3), 162–182. <https://doi.org/10.1002/lnc3.326>
- Sandler, W. (2018). The body as evidence for the nature of language. *Frontiers in Psychology, 9*, Article 1782. <https://doi.org/10.3389/fpsyg.2018.01782>
- Sandler, W. (2022). Redefining multimodality. *Frontiers in Communication, 6*, 758993. <https://doi.org/10.3389/fcomm.2021.758993>
- Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge University Press.
- Sandler, W., Meir, I., Dachkovsky, S., Padden, C., & Aronoff, M. (2011). The emergence of complexity in prosody and syntax. *Lingua. International Review of General Linguistics. Revue Internationale de Linguistique Generale, 121*(13), 2014–2033. <https://doi.org/10.1016/j.lingua.2011.05.007>
- Sato, A., Kirby, S., & Flaherty, M. (2022). Language emergence can take multiple paths: Using motion capture to track axis use in Nicaraguan Sign Language. *Glossa: A Journal of General Linguistics, 7*(1), Article 1. <https://doi.org/10.16995/glossa.6151>
- Schembri, A. (2003). Rethinking “classifiers” in signed languages. In K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 3–34). Lawrence Erlbaum.

- 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. *The Journal of Deaf Studies and Deaf Education*, 10(3), 272–290. <https://doi.org/10.1093/deafed/eni029>
- Schick, B. S. (1987). *The acquisition of classifier predicates in American Sign Language* [Doctoral dissertation]. <https://docs.lib.purdue.edu/dissertations/AAI8807670/>
- Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., & Blake, A. (2011, June 20–25). Real-time human pose recognition in parts from single depth images. In *IEEE Computer Society conference on computer vision and pattern recognition* (pp. 1297–1304). <https://ieeexplore.ieee.org/document/5995316>
- Singleton, J., Morford, J., & Goldin-Meadow, S. (1993). Once is not enough: Standards of well-formedness in manual communication created over three different timespans. *Language*, 69, 683–715.
- Stokoe, W. C., Jr. (1960). Sign language structure: An outline of the visual communication systems of the American deaf. *The Journal of Deaf Studies and Deaf Education*, 10(1), 3–37. <https://doi.org/10.1093/deafed/eni001>
- Supalla, T. (1978). Morphology of verbs of motion and location in American Sign Language. In F. Caccamise & D. Hicks (Eds.), *American Sign Language in a bilingual, bicultural context* (pp. 27–45). National Association of the Deaf.
- Supalla, T. (1982). *Structure and acquisition of verbs of motion and location in American Sign Language* [Unpublished doctoral dissertation]. University of California at San Diego.
- Supalla, T. (1986). The classifier system in American Sign Language. In C. Craig (Ed.), *Noun classes and categorization* (pp. 180–214). John Benjamins.
- Tölgyessy, M., Dekan, M., & Chovanec, L. (2021). Skeleton tracking accuracy and precision evaluation of Kinect V1, Kinect V2, and the Azure Kinect. *Applied Sciences*, 11(12), Article 12. <https://doi.org/10.3390/app11125756>
- Weibel, N., Hwang, S.-O., Rick, S., Sayyari, E., Lenzen, D., & Hollan, J. (2016, January 5–8). Hands that speak: An integrated approach to studying complex human communicative body movements. In *2016 49th Hawaii international conference on system sciences (HICSS)* (pp. 610–619). <https://doi.org/10.1109/HICSS.2016.82>
- Wilbur, R. B. (1999). Stress in ASL: Empirical evidence and linguistic issues. *Language and Speech*, 42(2–3), 229–250. <https://doi.org/10.1177/00238309990420020501>
- Zwitserslood, I. (2012). Classifiers. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign language. An international handbook* (pp. 158–186). De Gruyter.