

FIGURE: Gesture Annotation

Annotation Schema of the *Frankfurt Image GestURES*

Andy Lücking, Dennis Kurfürst, Désirée Walther, Marcel Mauri, Alexander Mehler

Abstract

This report documents the annotation format used in order to prepare the *FIGURE* corpus. For a description of data gathering, research rationale, agreement studies and first results please see the respective [LREC publication](#) (Lücking, Mehler, Walther, Mauri, and Kurfürst 2016).

Contents

| | | |
|-----|------------------------------------|---|
| 1 | Introduction | 1 |
| 2 | Gesture Terms | 1 |
| 3 | Phases | 2 |
| 4 | Gesture Types | 2 |
| 5 | Hands | 2 |
| 5.1 | Handshape | 2 |
| 5.2 | Hand Orientation | 2 |
| 5.3 | Hand Location | 4 |
| 5.4 | Hand Distance | 4 |
| 5.5 | Handshape Changes | 4 |
| 6 | Trajectories | 4 |
| 6.1 | Paths | 4 |
| 6.2 | Direction | 4 |
| 6.3 | Repetition | 5 |
| 7 | Relations Between Hands | 5 |
| 7.1 | Hand and Hand Contact | 5 |
| 7.2 | Hand and Arm Contact | 5 |
| 7.3 | Temporal relation | 5 |
| 8 | Sequences | 5 |
| 9 | Summary of Annotation Tiers | 5 |
| 10 | Availability | 6 |

1. Introduction

The **Frankfurt Image GestURE** (FIGURE) corpus aims at finding recurrent kinetic features of gestures associated with directional, evaluative and above all abstract stimulus terms. In total, 260 gestures have been elicited in a controlled experimental setting (Lücking, Mehler, Walther,

Mauri, and Kurfürst 2016). In order to get at the kinetic features of these gestures, they have been annotated according to the annotation schema described in this report.

The gestures from FIGURE usually exhibit the common threefold structure consisting of the three phases *preparation*, *stroke*, and *retraction* (Kendon 1980; McNeill 1992). Additionally, before the start of a gesture, there is a *presentation* phase where the actual term is displayed and processed by the subject. These phases are explicated in the phase-annotation of FIGURE. Note that this annotation focuses on the temporal properties of the target gestures and abstracts over eventual hold phases.

Each gesture is annotated as a feature vector describing a starting position, a trajectory and an end position (Gibbon et al. 2003). Gestures that performed by one hand and gestures that are performed by two hands are explicitly distinguished. In case of two-handed gestures the relations (e.g., timing) between both hands are captured. Drawing on the notion of a gesture sequence, gestures that consist of consecutive movements are accounted for. The annotation levels as well as the permitted annotation values for the FIGURE annotation are described below. The annotation has been carried out using *Elan*¹, a multimodal annotation tool that facilitates time-aligned video annotation.

2. Gesture Terms

The main annotation tier of FIGURE is the so-called **gesture tier**. On this tier, the temporal stretch of a gesture is demarcated. The content of the respective annotation element is the current stimulus term. In *Elan*, this tier is

¹<https://tla.mpi.nl/tools/tla-tools/elan>.

time-aligned to the video signal. It is used as a reference tier for other annotations.

3. Phases

A phase annotation is a sequence of the following elements:

1. *present*: the time span from looking at the monitor to starting of movement. Any movement that is interrupted or repaired is also put here.
2. *prep* (preparation): the time span from the start of a movement to the start of the gesture proper.
3. *stroke*: the gesture proper. This annotation element is aligned to the *Geste*-element from the FIGURE annotation (and via *symbolic association* to all its child elements).
4. *retr* (retraction): arm-and-hand movement that dissolve stroke position.

Since phases detail the temporal pattern of gestures, they subdivide the *gesture* tier described above in Section 2.

4. Gesture Types

At first, each gesture is classified according to the interaction between both hands. One-handed gestures are assigned the type “simplex” and are further divided in left-handed and right-handed gestures. “Complex”, i.e., two-handed gestures, are classified with regard to the status of their cooperation: if both hands work together in order to produce a single, common depiction, they are labeled as “complex-sym”. If both hands act individually and produce two (usually related) depictions, they are assigned the value “complex-ind”. In sum, the following gesture types are distinguished:

| (1) | value | description |
|-----|-------------|--|
| | simplex-rh | gesture with right hand only |
| | simplex-lh | gesture with left hand only |
| | complex-sym | gesture with both hands in symmetry |
| | complex-ind | gesture with both hands in individuality |

5. Hands

The basic building block *hand* is annotated with respect to

- its location of within gesture space (McNeill 1992) – see Figure 1;
- the handshape in terms of the *American Sign Language* (ASL) alphabet – see Figure 2;
- and the orientations of the back of hand and the palm according to the speaker-centered spatial reference system.

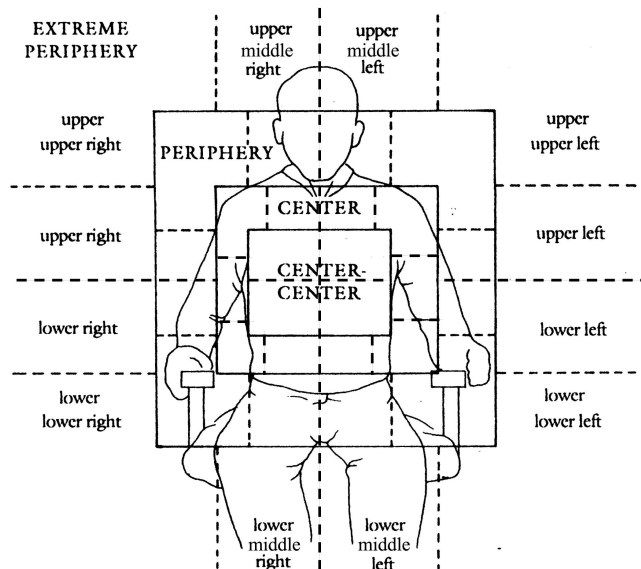
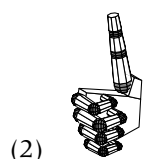


Figure 1. Transcription of gesture space according to Pedelty and McNeill (taken from McNeill (1992, p. 378)).

5.1 Handshape

The handshape is described at the start and the end of every gesture. The respective annotation values are taken from ASL and are exemplified in Figure 2. One handshape is added to the ASL handshapes, namely “thumbs-up” (illustrated in 2), which gets the name “10”.



(2)

5.2 Hand Orientation

The annotation of the orientation of the hand follows the orthogonal specification of the back of the hand (BoH) and the palm as used in Lücking, Bergmann, Hahn, Kopp, and Rieser (2010). The following basic values are permitted:

| (3) | value | description |
|-----|-------|---|
| | TL | Palm/Back of Hand pointing left |
| | TR | Palm/Back of Hand pointing right |
| | UP | Palm/Back of Hand pointing upwards |
| | DN | Palm/Back of Hand pointing downwards |
| | TB | Palm/Back of Hand pointing towards body |
| | AB | Palm/Back of Hand pointing away from body |

In order to capture intermediate values, orientation values can be combined by means of the ‘/’ operator. For instance, a left hand facing half down, half to the right has the palm value “DN/TR”.

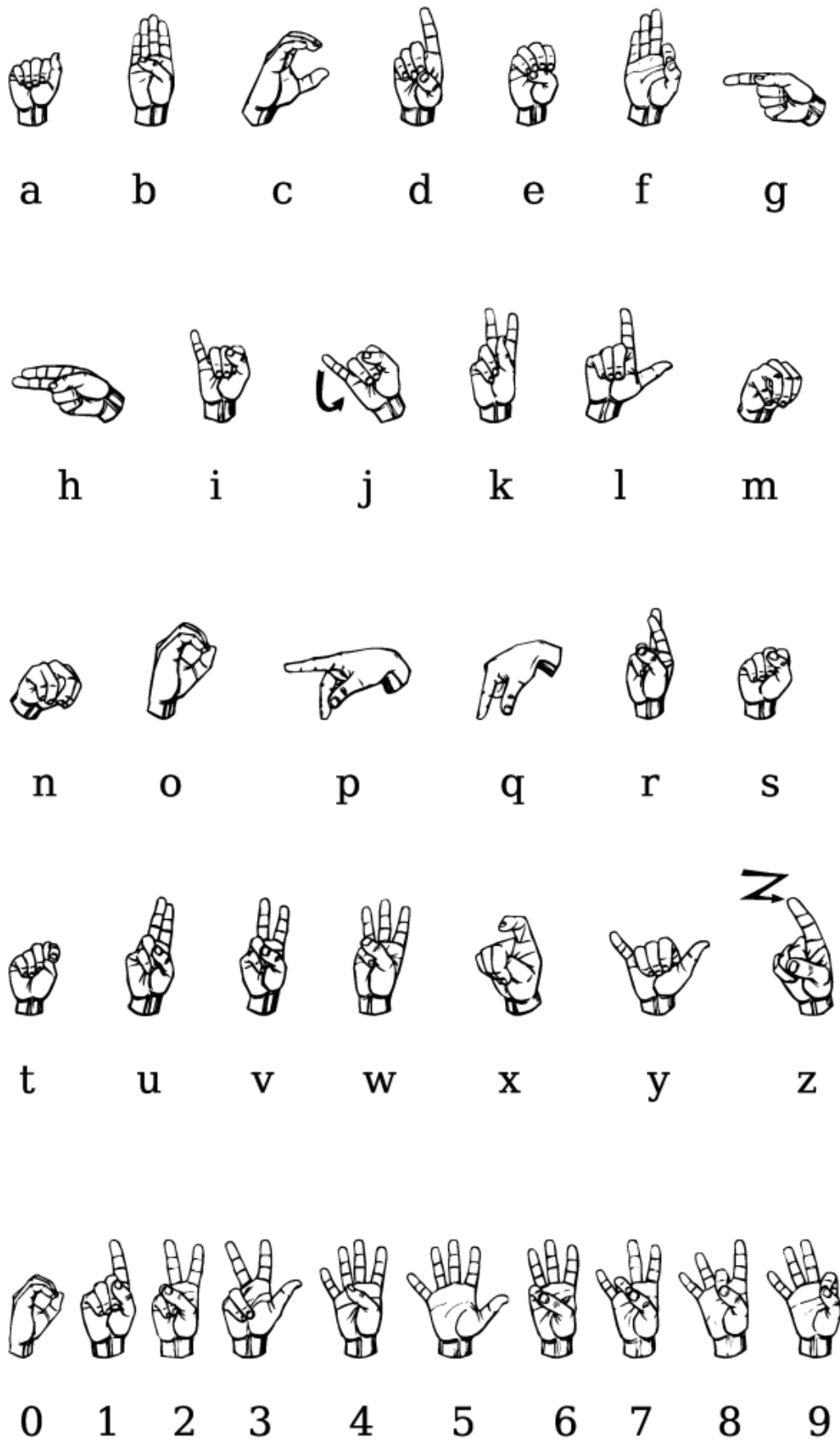


Figure 2. ASL handshapes (public domain).

5.3 Hand Location

Gesture space is divided according to Figure 1. The respective annotation values are taken from the SaGA schema (Lücking, Bergmann, Hahn, Kopp, and Rieser 2010) and given in (4).

| (4) | value | description |
|-----|--------|--------------------------------------|
| | CC | center center (@chest) |
| | C-UP | center-upper (@neck) |
| | C-UR | center-upper-right (@R-shldr) |
| | C-UL | center-upper-left (@L-shldr) |
| | C-RT | center-right (@R-arm) |
| | C-LT | center-left (@L-arm) |
| | C-LW | center-lower (@stomach) |
| | C-LR | center-lower-right |
| | C-LL | center-lower-left |
| | P-UP | periphery upper (@face) |
| | P-UR | periphery upper right (@abv R-shldr) |
| | P-UL | periphery upper left (@abv L-shldr) |
| | P-RT | periphery right |
| | P-LT | periphery left |
| | P-LW | periphery lower (@lap) |
| | P-LR | periphery lower right |
| | P-LL | periphery lower left |
| | (EP-UP | extreme periphery upper) |
| | EP-UUR | extreme periphery upper upper right |
| | EP-UUL | extreme periphery upper upper left |
| | EP-UMR | extreme periphery upper middle right |
| | EP-UML | extreme periphery upper middle left |
| | EP-UR | extreme periphery upper right |
| | EP-UL | extreme periphery upper left |
| | EP-RT | extreme periphery right |
| | EP-LT | extreme periphery left |
| | (EP-LW | extreme periphery lower) |
| | EP-LLR | extreme periphery lower lower right |
| | EP-LLL | extreme periphery lower lower left |
| | EP-LMR | extreme periphery lower middle right |
| | EP-LML | extreme periphery lower middle left |
| | EP-LR | extreme periphery right |
| | EP-LL | extreme periphery left |

5.4 Hand Distance

In order to capture three-dimensional gesture space information, distance is annotated separately (Lücking, Bergmann, Hahn, Kopp, and Rieser 2010), by means of the following annotation values:

| (5) | value | description |
|-----|-------|---|
| | D-C | Hand in contact with body |
| | D-CE | Hand between body and elbow's length away |
| | D-EK | Between elbow and knee |
| | D-KO | Between knee and length of outstretched arm in front away |
| | D-O | Length of outstretched arm in front away |

5.5 Handshape Changes

In case there is a change in a gesture's handshape between its start and end position, the path of the change is captured in terms of a "Handmove". Furthermore, "wiggling" or bent handshapes are also marked here. Possible values are given in (6):

| (6) | value | description |
|-----|--------|---------------------------------------|
| | static | no hand-movement |
| | line | straight trajectory between to points |
| | wiggle | hand-movement by shaking hand |
| | bent | flex hand |
| | ... | |

6. Trajectories

6.1 Paths

Movement trajectories are delimited by the start and end positions of the hand(s). The value of the feature path is used to label the shape of the performed movement (Lausberg and Sloetjes 2009). If required, the following list may be extended:

| (7) | value | value | value |
|-----|----------|-------------|------------|
| | • line | • 8 | • pentagon |
| | • zigzag | • U | |
| | • edge | • rectangle | • pointing |
| | • arc | • triangle | |
| | • curve | • star | • ... |

6.2 Direction

Since movement patterns be oriented in various ways within gesture space, their orientation is specified as a value of direction. To this end, the orientation values from BoH and palm of the hand can be recycled. In case of complex movement, several of such values are necessary, hence direction is a list-valued feature. List concatenation is done by combining values by means of '>' (cf. Lücking, Bergmann, Hahn, Kopp, and Rieser 2010). For instance, an edge that is first going into the line of the speaker's gaze and then upwards is coded as "AB>UP".

The path directions acknowledged in FIGURE are given in (8):

(8)

| value | description |
|-------|----------------|
| LT | left |
| RT | right |
| UP | up |
| DN | down |
| AB | away from body |
| TB | towards body |
| UR | upper right |
| UL | upper left |
| LR | lower right |
| LL | lower left |

6.3 Repetition

The value of repetition gives the number of iterations of a movement, ranging from 0 (no repetition) to an in principle unlimited number of repetitions.

7. Relations Between Hands

Due to complex gestures, the relation between both hands becomes a means of expression that has to be accounted for in annotation.

7.1 Hand and Hand Contact

Is there a contact between both hands? If this is so, the exact touching point is specified.

(9)

| value | description |
|----------------|---|
| Yes | contact without specification |
| No | no contact |
| Palm | contact between Palm of each hand |
| Palm with hole | contact between hands like drinking water out of them |
| BoH | contact with Back of Hand (BoH) |
| thumbs | contact between thumbs |
| first finger | contact between first fingers |
| second finger | contact between second fingers |
| third finger | contact between third fingers |
| fourth finger | contact between fourth fingers |
| all fingers | contact between all fingers |
| four fingers | contact between four fingers only |
| ... | |

Contact is annotated for both the start and the end of a gesture movement.

7.2 Hand and Arm Contact

In addition to contact, which is restricted to fingers and hands only, the contact between hand(s) and arm(s) has to be handled in a special way. FIGURE accounts for the relations given in (10):

(10)

| value | description |
|-------|-------------------------------------|
| UAR | hand in contact to right upper arm |
| LAR | hand in contact to left forearm |
| UAL | hand in contact to left upper arm |
| LAL | hand in contact to left forearm |
| UA | both hands in contact to upper arms |
| LA | both hands in contact to forearm |
| ... | |

Hand and arm contact is annotated for both the start and the end of a gesture movement.

7.3 Temporal relation

The temporal relations between the movements of both hands are captured in terms of the following values:

(11)

| value | description |
|----------|--|
| = | left hand is temporally equal to right hand |
| rh >= lh | right hand starts before left hand |
| lh >= rh | left hand starts before right hand |
| rh > lh | right hand is finished before left hand starts |
| lh > rh | left hand is finished before right hand starts |

8. Sequences

Different gestures can be a part of a sequence of gestures. If a gesture sequence is detected is classified into one if the following sequences types:

(12)

| value | description |
|-------------|-------------------------------|
| x | no sequence, only one gesture |
| seq-rh | sequence with only right hand |
| seq-lh | sequence with only left hand |
| seq-complex | sequence with both hands |

9. Summary of Annotation Tiers

The annotation is implemented in Elan in terms of the following annotation tier, where the **gesture** is the “mother” tier, which is time aligned to the video signal, and **Rechte Hand** respectively **Linke Hand** cover the time stretch of gestural movements for each hand separately. The complete list of annotation tiers is given in (13). Note that for the sake of data management also questionnaire information (ID of participant, mother tongue, handedness, age and gender) has been included into the annotation

files. (Possibly German labels of annotation tier has been translated into English.)

(13)

- Gesture (*Geste*)
- phase
- Right hand (*Rechte Hand*)
- start Position r.H.
- start Handshape r.H.
- start Palm Orth. r.H.
- start BoH r.H.
- Path r.H (*Pfad r.H.*)
- Hand Move r.H.
- path orientation r.H.
- path orientation l.H.
- end Position r.H.
- end Handshape r.H.
- end Palm Orth. r.H.
- end BoH r.H.
- start dist r.H.
- end dist r.H.
- path repeat r.H.
- Left hand (*Linke Hand*)
- start Position l.H.
- start Handshape l.H.
- start Palm Orth. l.H.
- start BoH l.H.
- Path l.H. (*Pfad l.H.*)
- Hand Move l.H.
- bothHand-sym.
- Type of gesture (*Typ Geste*)
- end Position l.H.
- end Handshape l.H.
- end Palm Orth. l.H.
- end BoH l.H.
- end dist l.H.
- start dist l.H.
- path repeat l.H.
- bh. start contact
- bh. end contact
- bh. temp. rel.
- Start contact arm
- End contact arm
- Sequence (*Sequenz*)
- Participant (*Versuchsperson*)
- Gender (*Geschlecht*)
- Handedness (*Händigkeit*)
- Age (*Alter*)
- First language (*Muttersprache*)

10. Availability

The consistency of the FIGURE annotation has been examined by means of inter-rater agreement studies on most of the levels collected above. The respective kappa values range from 0.42 to 0.98, with a mean value of 0.84 (see Lücking, Mehler, Walther, Mauri, and Kurfürst 2016). The data set is freely available under a ‘CC BY-SA 4.0’ (© ⓘ ⓘ) license at <https://hucompute.org/applications/corpora>.

References

- Gibbon, Dafydd et al. (2003). “A Computational Model of Arm Gestures in Conversation.” In: *Proceedings of the 8th European Conference on Speech Communication and Technology*. EUROSPEECH 2003. Geneva, Switzerland, pp. 813–816.
- Kendon, Adam (1980). “Gesticulation and Speech: Two Aspects of the Process of Utterance.” In: *The Relationship of Verbal and Nonverbal Communication*. Ed. by Mary Ritchie Key. Vol. 25. Contributions to the Sociology of Language. The Hague: Mouton, pp. 207–227.
- Lausberg, Hedda and Han Sloetjes (2009). “Coding gestural behavior with the NEUROGES-ELAN system.” In: *Behavior Research Methods* 41.3, pp. 841–849.
- Lücking, Andy, Kirsten Bergmann, Florian Hahn, Stefan Kopp, and Hannes Rieser (2010). “The Bielefeld Speech and Gesture Alignment Corpus (SaGA).” In: *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*. LREC 2010. 7th International Conference for Language Resources and Evaluation. Malta, pp. 92–98. doi: 10.13140/2.1.4216.1922.
- Lücking, Andy, Alexander Mehler, Désirée Walther, Marcel Mauri, and Dennis Kurfürst (2016). “Finding Recurrent Features of Image Schema Gestures: the FIGURE corpus.” In: *Proceedings of the 10th International Conference on Language Resources and Evaluation*. LREC 2016. Portorož (Slovenia).
- McNeill, David (1992). *Hand and Mind—What Gestures Reveal about Thought*. Chicago: Chicago University Press.