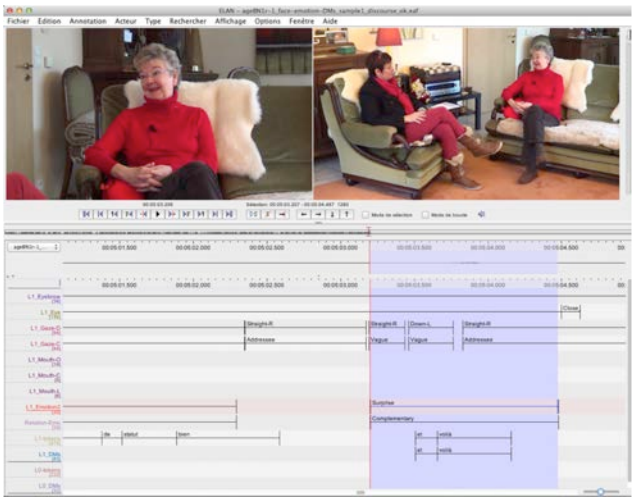
Annotation Manual – Version 1.8
This manual was last updated on 2016-05-02

CorpAGEst (2013-2015): "A corpus-based multimodal approach to the pragmatic competence of the elderly"
People Marie Curie Actions (PIEF-GA-2012-328282)

Multimodal annotation guidelines

I. Gestural annotation guidelines

(*ELAN software*)



Original Author: Catherine Bolly
With the collaboration of: Delphine Belin, Dominique Boutet, Julie Kairet, Alysson Lepeut, Anna Sáfár, Anaïs Thomas

Contact: catherine.bolly(at)uclouvain.be; cbolly(at)uni-koeln.de
Web: http://corpagest.org

Content

Introduction

☞ The CorpAGEst project

The CorpAGEst project ("A corpus-based multimodal approach to the pragmatic competence of the elderly") aims at establishing the gestural and verbal profile of very old people in aging, looking at their pragmatic competence from a naturalistic perspective. The CorpAGEst assumption is that multimodal (inter)subjective markers of stance are highly relevant cues for the measurement of communicative competence in later life. The project aims *in fine* at a better understanding of the way in which the verbal and gestural dimensions interact to make sense in real-world settings (thus going far beyond the specific scope of the present project). This project has received funding from the European Union Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n° [PIEF-GA-2012-328282].

The CorpAGEst corpus (Bolly & Boutet, submitted) is comprised of face-to-face conversations between an adult and a very old subject (75 y. old and more) living at home or in a residential home. The corpus data consist of semi-directed interviews, which have been audio and video recorded, transcribed and aligned to the sound signal. The corpus is two-fold, including transversal and longitudinal subcorpora. Contextual independent variables are part of the corpus design, namely environment (private *vs.* residential home), the social tie between the participants (familiar *vs.* unknown interviewer) and the task type (focusing on past events *vs.* present-day life). The corpus is part of the international CLARe initiative ("Corpora for Language and Aging Research"), which combines methods in linguistics and issues in aging, and advocates for more corpus-based "naturalistic" approaches in the field.

The multimodal data (text, sound and video) were aligned to the sound signal in partition mode, using the Praat program (Boersma & Weernink, 2014), the EasyAlign plugin (Goldman, 2011), and the ELAN software (Wittenburg *et al.*, 2006). The transcription standards adopted for the oral component were slightly adapted from those of the Valibel research center (Dister *et al.*, 2007 [2009]), as described in the part of the manual dedicated to speech.

For further detail, see the project website: http://corpagest.org.

☞ Multimodal annotation: from form-based to function-based analysis

The perspective adopted in the CorpAGEst project is a form-based one (see Müller *et al.*, 2013), extended and applied to facial expressions, gaze, hand gestures, and body gestures (*viz.* head, shoulders, torso, legs, feet). Notably, the annotation procedure lays on a triple principle, according to which a visible action is considered as a *potentially* meaningful gesture unit in the ongoing flow of interaction: (i) the "visibility" criterion: identification of all actions that are visible in the interaction flow, through the eyes of the camera recorder and through those of the analyst; (ii) the "meaning potential" criterion: from the semantic-pragmatic perspective, every visible action identified must potentially convey one semantic-pragmatic meaning in the particular context of its realization (thus also including beats, adaptors, deictic, and interactive gestures), from the point of view of the analyst; (iii) the "formal distinctiveness" criterion: to distinguish between consecutive moves in the interaction flow, there must be at least one change in formal/physiological parameters (e.g. shape for the hand, direction of the head, etc.), by comparison with the preceding and following gesture phase or move.

In the same manner, speech data are investigated first in terms of linguistic parameters (syntactic category and position, procedural/conceptual meaning, meaning in context, presence of a co-occurring DM, etc.), and then according to their function in use. The protocol for discourse markers identification and annotation follows the one developed within the MDMA project ("Model for Discourse Marker Annotation" – see Bolly *et al.*, 2015; Bolly *et al.*, forthc.). The methodology of MDMA starts from an independent selection of candidate discourse markers by several expert coders, which then undergo syntactic and semantic description through an operational annotation model. A specific section of the

manual (still in progress) is dedicated to speech transcription (via Praat), alignment (via EasyAlign), and annotation (in ELAN).

Starting with mono-modal analyses (gesture *vs.* speech) and focusing on one group of articulators at a time within each modality (*viz.* face, gaze, head, shoulders, torso, hands, legs, and feet – for the nonverbal mode), the annotation procedure next moves to a multimodal and functional perspective on pragmatic cues (*viz.* emotions and (non)verbal pragmatic markers. The model for the annotation of pragmatic functions is part of the MDMA project (see above) and is a collaborative work (see Bolly & Crible, Antwerp 2015), which has been developed to be transferrable to several modalities and languages (see Bolly *et al.,* Göttingen 2015).

Modality: nonverbal/gesture	Articulators
1. Facial displays	
	Eyebrows
	Eyes
	Gaze
	Mouth
2. Hand gestures	
	Hands
3. Body gestures	
	Head
	Shoulders
	Torso
	Legs
	Feet
Modality: verbal/speech	Levels of analysis
	Pragmatic markers
FUNCTION-BASED ANALYSIS - Multimodal annotation of emotions - Multimodal annotation of pragmatic functions	

Table 1. Form-based and function-based approach to corpus data in CorpAGEst

I. Gestural annotation guidelines (ELAN software)

Modality: nonverbal/gesture	Articulators
1. Facial displays	
	Eyebrows
	Eyes
	Gaze
	Mouth
2. Hand gestures	
	Hands
3. Body gestures	
	Head
	Shoulders
	Torso
	Legs
	Feet

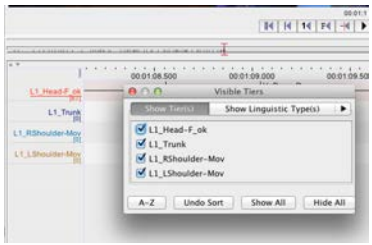
Table 2. Articulators taken into account for gestural annotation (in ELAN)

Annotation principles for the nonverbal mode

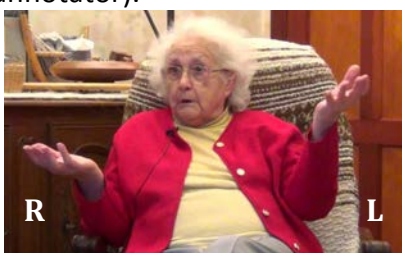
Annotation without any recourse to the sound signal (check that the sound is on "mute" in ELAN): as a reminder, the approach is form-based and uni-modal at the first stage of analysis; the multimodal and functional approach comes in a next step.

Focus on one articulator AND one Tier at a time while annotating (e.g. first segmenting all the gestures in the sample, then annotating every stroke in terms of "Shape", then according to their "Orientation", then "Position", etc.): only make the Tier(s) you want to work on visible on the screen, to avoid being influenced by the existing annotations on the other Tiers. (But see the exception for "Hand-Ph" and "Hand-Mov", where the decision between one single stroke with repeated moves or distinct moves is decisive).

🛠 Hide/Show function in ELAN: click right in the left column with Tiers' names



Annotation of laterality: what is "Right" and what is "Left"? There are two main Tiers for each hand/shoulder/arm. For instance, the "Right Shoulder" and "Left Shoulder" Tiers refer to the body parts of the study subject (from his/her perspective), and do not correspond to the "right" or "left" side of the screen (from the point of view of the annotator).



How to put gestures' boundaries?

An important principle applying to all annotations concerns the boundary placement. The right boundary of a move or phase has to be put on the frame where the move is already initiated (most of the time, this is noticeable by the blurred image at that precise position of the cursor). For instance, after a stroke, the return phase only begins when the image appears as blurred, not on the previous clear image. Thus, if the next frame following the end of the stroke movement is blurred, it means that the return phase has already started.

When the boundary between two moves is fuzzy, the end of the first move corresponds to the very beginning of the second move, even if the first is still ongoing (often fading out).

> This boundary principle is similarly applied to the beginning of any phase (e.g., stroke, closing eyes, etc.) that is the left side of the segment.

> This principle is valid for every annotation on every nonverbal articulator.

What can we do with saccadic movements?

A saccade is defined here as an abrupt movement that consists in a change of position that does not entail a change of any physiological parameter (e.g., shape, direction, amplitude, velocity). This is often characterized by the presence of a very quick pause within a continuous move. In such cases, the move is annotated as a single move and must not be divided into two distinct (repeated) moves (see the following examples: a single "Up" move of the head occurs with a visible pause in ageBM1 sample2, from 00:01:17.314 to 00:01:17.946; a single "Turn" move of the head occurs with a visible pause in ageBM1 sample2 from 00:02:12.211 to 00:02:12.935).

Perceptive saliency and visibility, from the point of view of the coder (not of the interlocutor), as a necessary (but not sufficient) criterion to take a move/action as a unit to be annotated…

The perceptive *saliency* of any potentially meaningful visible action (even if very quick or produced with a narrow amplitude) plays a crucial role in the decision to take this action into account for the annotation. Yet, when there is too much hesitation for one action to be annotated or not (that is, it is visible but quite imperceptible still), there is a methodological bias that must lead the annotator to the choice of NOT annotating this action (see the quasi imperceptible squeezing of Albertine in ageDA1r-1_sample3 from 00:03:11.374 to 00:03:15.875). This decision should be made in order to avoid any discrepancy between annotators and to improve the inter-coder agreement score.

Ambiguous cases and inter-coder agreement

The coder may have recourse to another coder judgment on cases that remain ambiguous and are not decidable on the basis of common annotation principles.

> **Ask a second person** (if you annotate on your own) **or a third one** (if there are two coders) in order to balance your choice.

1. Facial displays

1.1. Tier structure

Mainly inspired from the *MUMIN* project (Allwood *et al.*, 2005, 2007), the ELAN annotation scheme dedicated to the physiological description of facial expressions is comprised of 9 annotation Tiers: "Eyebrow", "Eye", "Gaze", "Gaze-D", "Gaze-C", "Mouth", "Mouth-O", "Mouth-C", and "Mouth-L".

Eyebrows' move [Eyebrow]	
Eyes' move [Eye]	
Gaze [Gaze]	
	Gaze direction [Gaze-D]
	Gaze contact [Gaze-C]
Mouth [Mouth]	
	Mouth Openness [Mouth-O]
	Lips' corners [Mouth-C]
	Lips' shape [Mouth-L]

Table 3. Tiers' structure and articulators for the annotation of face and gaze (in ELAN)

Annotation principles

What parameters?

Facial displays (including gaze) were identified according to their location in the face (eyebrows, eyes, gaze, mouth) and then annotated in terms of physiological features (e.g., closed-both for the eyes, corners up or retracted for the lips). Eye-gazing is included in the scheme, insofar as it plays a major role in providing feedback and establishing or sustaining the focus of shared attention, thus mirroring reciprocal arrangement during human-human interactions (cf. Jokinen *et al.*, 2009).

With or without recourse to sound?

The annotation is made independently from the sound signal to avoid any interpretive bias in the semiotics of gesture at this stage in the analysis (cf. Bressem, 2008). Note that there is an exception: for the "Openness", "Lips' corners" and "Lips' shape" only, the sound signal can be activated to distinguish between moves that accompany speech production (not taken into account) and those that are produced alone (tagged as such in the ELAN file).

How to put moves' boundaries?

Following the general principle for gestural annotation (see above), the left boundary of each annotation – that is, the beginning of the move – must be assigned to the first frame that corresponds to a visible change in the face, mostly on a blurred image (e.g., when the eyes begin to close, not when they are completely closed); in the same manner, the right boundary of facial expressions – that is, the end of the move – must be put on the frame corresponding to the absence of any visible change, mostly a fixed image (e.g., when the eyes are fully open again).

What can I do with fading effect at the end of moves?

In the case of moves that disappear very slowly, in a quite undetectable way (e.g. eyebrows going back to their neutral position), the right boundary must be put on the frame corresponding to the recovered neutral position. It may be sometimes of some help to look at simultaneous physiological features from another location in the face (e.g., wrinkling the forehead while eyebrow raising) to detect the very end of such fading moves.

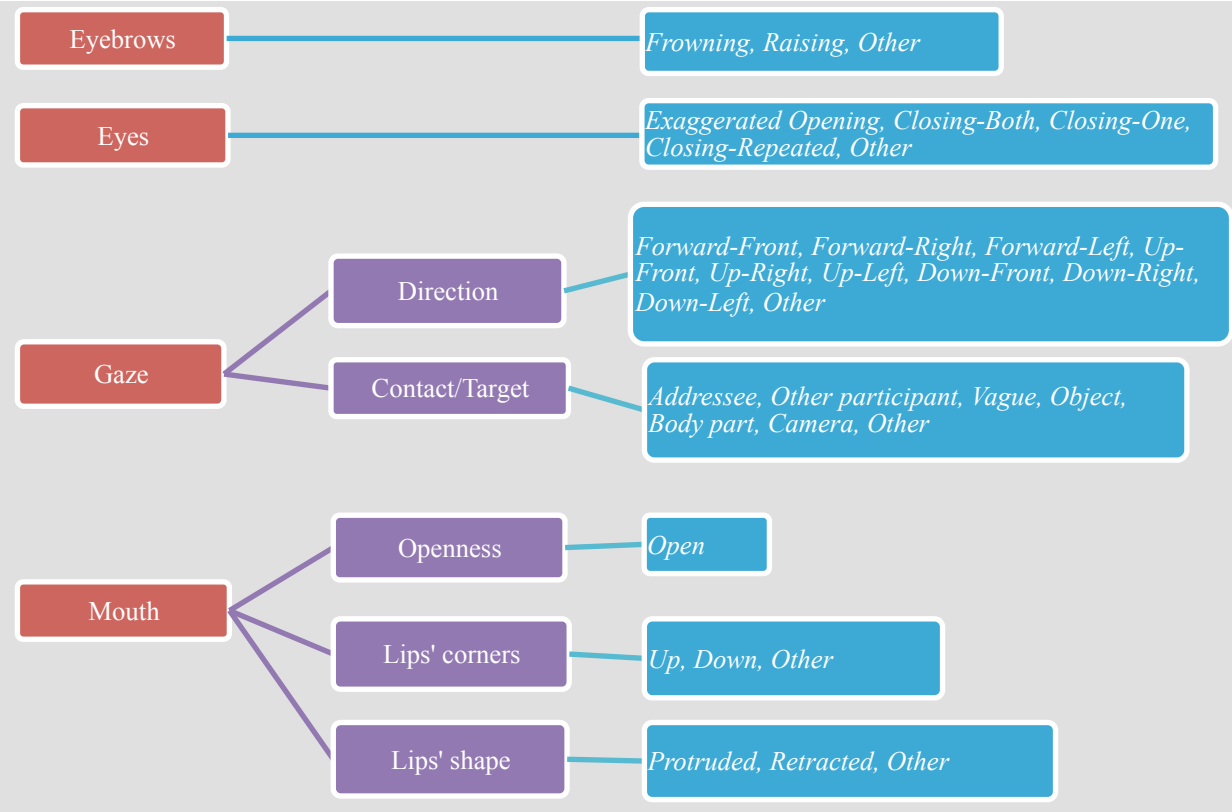


Figure 1. Tiers' structure and dependency between physiological parameters for facial displays (in ELAN)

1.2. Guidelines per Tier

1.2.1. Eyebrows' move [Eyebrow]

Eyebrows' moves are described using a Controlled Vocabulary, distinguishing between "Frowning" and "Raising" moves (see MUMIN project – Allwood *et al.*, 2005, 2007).

Frowning [Frown]	Frowning of one single or the two eyebrow(s).
Raising [Rais]	Raising of one single or the two eyebrow(s).
Other [Other]	Any other visible eyebrows' move, which is considered to be potentially meaningful, but cannot be classified as a "Frowning" or "Raising" move.

1.2.2. Eyes' move [Eye]

Eyes' moves are described using a Controlled Vocabulary, distinguishing between "Exaggerated opening" of the eyes and various types of "Closing" one or the two eye(s) (see MUMIN project – Allwood *et al.*, 2005, 2007).

Exaggerated Opening [X-Open]	Wide opening of the eyes.
Closing-Both [Close-BE]	Closing of the two eyes.
Closing-One [Close-B]	Closing of only one eye.
Closing-Repeated [Close-R]	Quick repeated closings of the eyes.
Other [Other]	Any other visible eyes' move, which is considered to be potentially meaningful, but cannot be classified according to the listed categories.

1.2.3. Gaze [Gaze]

It has been decided to annotate gaze all along the video data (with the obvious exception of closing eyes) rather than to delimit so-called "gaze-units", as it allows for a highlighting of transition in gaze direction and nature of the target. The 2 child Tiers (to which a Controlled Vocabulary has been assigned – see below) are associated with a pre-identified "Gaze" span (without any annotation tag) on the parent Tier.

Child Tiers of *Gaze*

On the child Tiers associated with the "Gaze" span (which has been previously identified), information is given on the "Direction" of the gaze (from the point of view of the communicating people) and on the type of "Contact" resulting (or not) from gaze.

Gaze direction [Gaze-D]

Forward-Front [Straight-F]	Gaze is forward-oriented, in front of the study subject.
Forward-Right [Straight-R]	Gaze is forward-oriented, towards the right side of the study subject.
Forward-Left [Straight-L]	Gaze is forward-oriented, towards the left side of the study subject.
Up-Front [Up-F]	Gaze is upward-oriented, in front of the study subject.
Up-Right [Up-R]	Gaze is upward-oriented, towards the right side of the study subject.
Up-Left [Up-L]	Gaze is upward -oriented, towards the left side of the study subject.
Down-Front [Down-F]	Gaze is downward-oriented, in front of the study subject.
Down-Right [Down-R]	Gaze is downward -oriented, towards the right side of the study subject.
Down-Left [Down-L]	Gaze is downward -oriented, towards the left side of the study subject.
Other [Other]	Any other perceivable gaze, which is considered to be potentially meaningful, but cannot be classified according to the listed categories.

Gaze contact/target [Gaze-C]

Addressee [Addressee]	Gaze is addressed to the main interlocutor.
Other Participant [Participant]	Gaze is addressed to another person than the main interlocutor.
Vague [Vague]	Gaze is lacking expression and don't have any precise target (often with pupil dilation).
Object [Object]	The target of the gaze is a concrete object, which is identifiable in the ongoing communication situation.
Body part [Body]	The target of the gaze is part of the body of one participant in the ongoing communication situation. The contact can be established with the study subject himself (self-contact) or with someone else (contact with others).
Camera [Cam]	The target of the gaze is one of the video cameras.
Other [Other]	The target of the gaze is undetermined or not clearly identifiable.

1.2.4. Mouth [Mouth]

Every move of the mouth, which does not seem to be either neutral or resulting from a vocalization process, is selected on the first Tier "Mouth" and then annotated according to the three physiological parameters below: mouth "Openness", move of the "Lips' corners" and "Lips' shape" (adapted from MUMIN project – Allwood *et al.*, 2005, 2007).

Child Tiers of *Mouth*

The 3 child Tiers (to which a Controlled Vocabulary has been assigned – see below) are associated with a pre-identified "Mouth" unit (without any annotation tag) on the parent Tier.

Openness [Mouth-O]

"Closed mouth" is considered the neutral position. As a consequence, only "Open mouth" moves – which do not result from speech production – are annotated: "we do not annotate when the mouth is open because the person is uttering an open vowel" (Allwood *et al.*, 2005, 2007).

Lips' corners [Mouth-C]

Corners up [Up-C]	Upward move of one single or the two lips' corner(s).
Corners down [Down-C]	Downward move of one single or the two lips' corner(s).
Other [Other]	Any other move, which is considered to be potentially meaningful, but cannot be classified according to the listed categories.

Lips' shape [Mouth-L]

Protruded [Protruded]	Lips are protruded.
Retracted [Retracted]	Lips are retracted.
Other [Other]	Any other noticeable shape, which is considered to be potentially meaningful, but cannot be classified according to the listed categories.

2. Hand gestures

2.1. Tier structure

Right Hand Phases [RHand-Ph]	
	Shape [RHand-Sh]
	Orientation [RHand-O]
	Position [RHand-Pos]
	Movement [RHand-Mov]
Right Hand Contact [RHand-Contact]	
	Target [RHand-Contact-T]
	Body/Object [RHand-Contact-BO]
	Activity [RHand-Contact-A]
Left Hand Phases [LHand-Ph]	
	Shape [LHand-Sh]
	Orientation [LHand-O]
	Position [LHand-Pos]
	Movement [LHand-Mov]
Left Hand Contact [LHand-Contact]	
	Target [LHand-Contact-T]
	Body/Object [LHand-Contact-BO]
	Activity [LHand-Contact-A]
Hand Symmetry [Hand_Symmetry]	
	Plane [Hand_Symm-Pl]
	Synchronization [Hand_Symm-S]

Table 4. Tiers' structure for the annotation of hand gestures (in ELAN)

Annotation principles

If there is a slight change from one move to another, how to decide between one single phase and two distinct phases? Where to put the boundaries?

The basic principle is that if only one change in one parameter (e.g., change in "Shape" but not in "Orientation", "Position" or "Movement" for the hand), then the move has to be considered as a whole gesture move/phase. But if there is a change in at least two parameters, then it has to be considered as two consecutive moves/phases.

How to annotate "Hands Contact"? Is it dependent on the "Stroke"? The answer to the latter question is "No"…

The second main group of Tiers, "Contact" and its child Tiers, serves to describe contact between each hand and the environment, including other body parts of the speaker, objects, and the partner. The presence of contact is marked by empty annotations on the "Contact" Tier, and is further characterized on the child Tiers "Target", "Body/Object" and "Activity". The above-mentioned Tiers are present for the right and the left hand, respectively.

What about symmetry between the two hands?

Finally, the Tier "Hand Symmetry" and its child Tiers mark symmetry between both hands – hence there is only one set of these per speaker.

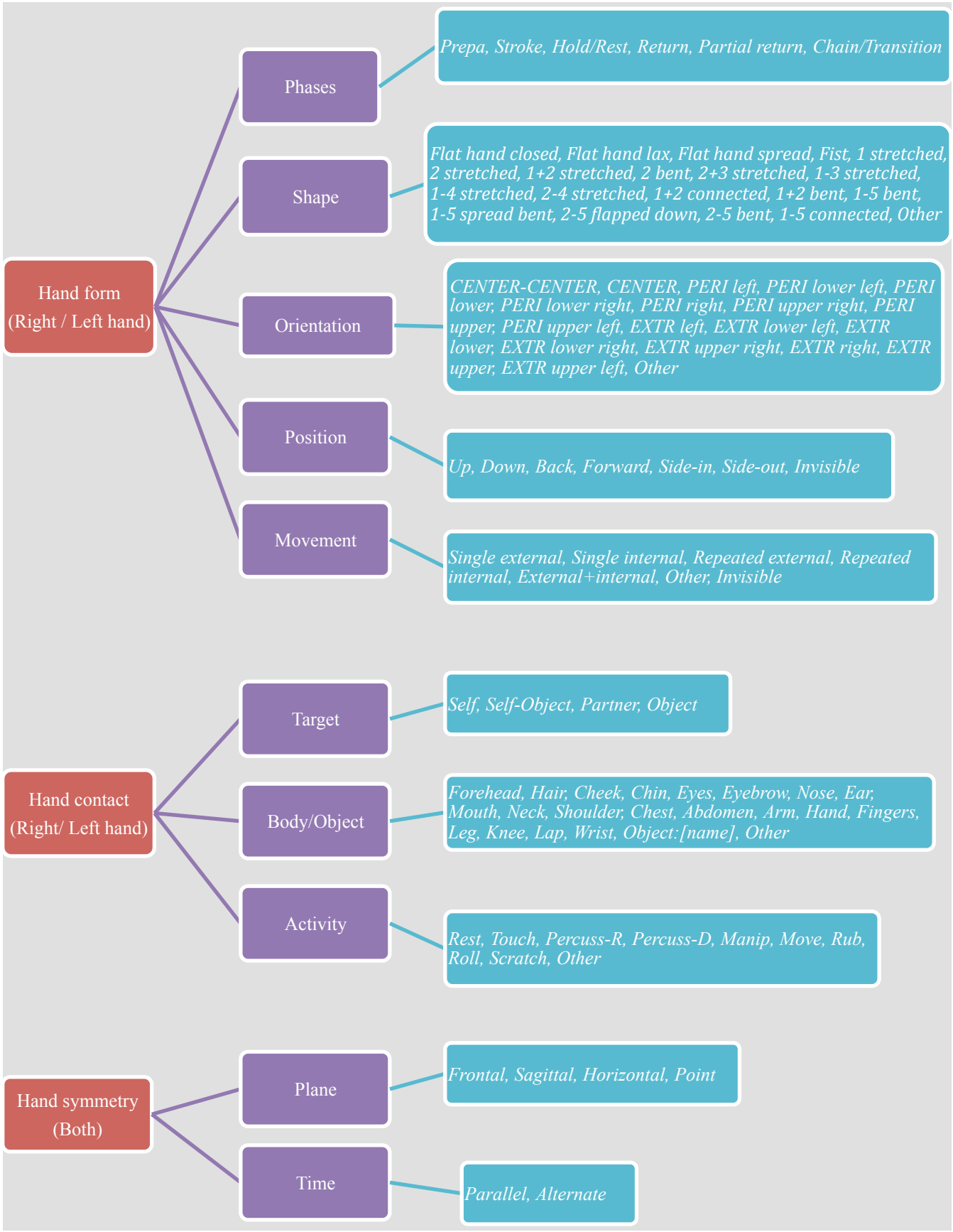


Figure 2. Tiers' structure and dependency between physiological parameters for hand gestures (in ELAN)

2.2. Guidelines per Tier

2.2.1. Phases [RHand-Ph and LHand-Ph]

On the "Phases" Tiers, type of hand movement is coded. A Controlled Vocabulary is used to segment hand activity into movement phases. Each category is described below. The "Phases" Tiers are annotated for each hand independently.

Preparation [Prepa]	Preparation phases include movements that precede the execution of a stroke, for example by taking the hand from rest position (e.g. in the lap) to the beginning point of the stroke. The change from preparation to stroke is often marked by a change in movement direction or handshape.
Stroke [Stroke]	Strokes are potentially meaningful movements. This includes any movement, either external (movement of the hand, as a whole, from one location to another) or internal (change of handshape and position, or movement of the fingers within a single handshape), with the exception of preparatory and return movements, or transitional movements between two strokes (see categories below). Each stroke includes one type of movement: external movement along a single path and/or the same internal movement. Both internal and external movements may be repeated within a stroke, as long as they involve the same path/same type of internal movement.
Hold, Rest [Hold]	Lack of hand movement or rest position. The hand is held static between two movement phases. 🛠 *How long does the hand need to be inactive for a hold to be annotated? In the current annotations, the shortest hold is 120 msec or 3 frames (1 frame = 40 sec); there is no limit for the maximum duration.*
Return [Return]	Return phases delimit movements following a stroke, when the hand returns to rest position, for example. Similar to the preparation phase, the return is often marked by a change in movement direction or handshape. "Recoil" phases are included within this category. Such phases often result from a return phase, for example, when hands are back to the rest position on the laps (cf. Ferré, 2012).
Partial Return [Return-P]	Partial return is marked when there is a return segment, but it is unfinished, for example because the hand does not reach rest position.
Chain, Transition [Chain]	Chain segments are transitional movements between two strokes or between two phases that cannot be classified as preparation or return phases, that is, which are ambiguous.

Annotation principles

Segmentation into phases

On the "Phases" Tier, hand activity is segmented into strokes (potentially meaningful gestures/actions from a semantic-pragmatic perspective) and holds, as well as preparation and return phases preceding and following strokes. On the child Tiers of the "Phases" Tier, each potentially meaningful hand movement (i.e., stroke) is further characterized according to four parameters: "Handshape", "Orientation", "Position" and "Movement".

Even though it is recommended to annotate one Tier at a time (to avoid any interpretative contamination from one parameter to another), the awareness of the type of movement at stake ("Hand-Mov") while segmenting hand gestures ("Hand-Ph") may serve as a help to decide where to put gesture boundaries. Indeed, recognizing one move as being external or internal, single or repeated, is often a prerequisite to decide between one single stroke with repeated moves (e.g., one single "External+Internal" stroke) or distinct moves (e.g., two subsequent strokes: "External" > "Repeated Internal").

What is a “Stroke”?

The “Stroke” is the most *potentially* meaningful part of the move, that is, which is *supposed* to convey meaning in the language interaction. Yet, its identification is very subjective and depends on the coder’s definition of what are the meaning potentials of gestures. In the CorpAGEst project, a move is considered to have a strong meaning potential when it plays a role in conveying a (partial) semantic (iconic, metaphoric, symbolic) or pragmatic meaning (beats, adaptors, interactive gestures, etc.).

How can I distinguish “Strokes” from the other phases?

Most of the time, it is possible to distinguish between binary moves, ternary moves and multiple moves. Two strokes can be contiguous, that means that they are not necessarily preceded or followed by a preparation, return or hold phasis. Examples of possible sequences: among others, “Stroke” > “Stroke” > “Return” > “Hold” OR “Prepa” > “Stroke” > “Chain” > “Stroke”.

- Binary moves: the “Stroke” is on the dynamic first part of the move (corresponding to a blurred image), including the *piek* of the gesture.
- Ternary moves: the “Preparation” phase in on the dynamic move (corresponding to the first blurred part of the image), the “Stroke” is on the *piek* of the gesture (corresponding to the stabilized and fixed image) and the next phase (or move) on the blurred image again.
- Multiple moves: the whole move is a “Stroke” because it cannot be divisible into segments. It generally correspond to quickly repeated (micro-)movements.

What can I do if one move is not divisible into phases? What is the minimum duration for a move to be divisible?

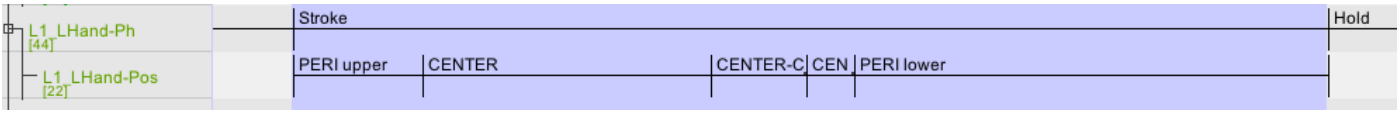
Not all movements can be segmented into preparation-stroke-return phases. In such cases, the whole movement is included in the stroke (e.g., quick binary moves). Furthermore, if the segmentation is possible, but the resulting segments would all be inferior to 250msec, only a single stroke annotation is added.

How to “split” one temporal segment into multiple annotations on a child Tier when it corresponds to one single Stroke on the parent Tier (with a “Time Subdivision Stereotype”)?

A first annotation on the child Tier will take up the whole length of the parent annotation. However, afterwards we can split the annotation on the child Tier selected, according to the following steps:

1) Create a new annotation on the child Tier (whose frontiers will correspond to the ones of the Stroke on the parent Tier)
2) Select the second part of the annotation that you want to split on the child Tier
3) Right click on the newly selected segment
4) Select "New Annotation Here"

You should get a new annotation from the beginning of your new selection to the end of the parent annotation. If you want a new annotation, make a new selection, etc.



Child Tiers of *Phases*

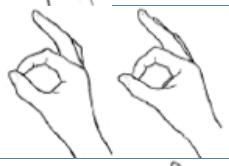
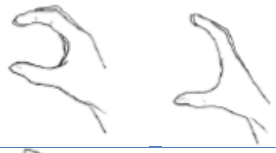
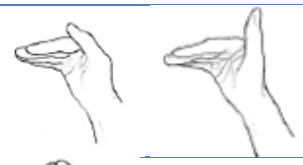
On the child Tiers associated with the "Phases" Tiers, the characteristics of the hand activity in each stroke phase are annotated. There are four child Tiers: "Shape" for handshape, "Location" for the location of the hand in gesture space, "Orientation" for the orientation of the palm, and "Movement" for the movement of the hand.

The child Tiers are related to the parent Tier with the "Time Subdivision" stereotype. This means that for each annotation on the parent Tier, several annotations of differing length can be created on the child Tier. Thus it is possible to indicate several handshapes or location (as well as orientation and movement) for a single stroke, in case there is a change in any of these parameters.

Shape *[RHand-Sh and LHand-Sh]*

Handshapes are described using a Controlled Vocabulary, adapted from Bressem (2013).

Flat hand closed [Flat-Closed]	
Flat hand lax [Flat-Lax]	
Flat hand spread [Flat-Spread]	
Fist [Fist]	
First finger (thumb) stretched [1_Stretch]	
Second finger (index) stretched [2_Stretch]	
First and second fingers stretched [1+2_Stretch]	
Second finger (index) bent [2_Bent]	
Second and third fingers (index and middle finger) stretched [2+3_Stretch]	
First, second and third fingers stretched [1-3_Stretch]	

First, second, third and fourth fingers stretched (without the little finger) [1-4_Stretch]	
Second, third, fourth ad fifth fingers stretched (without the thumb) [2-4_Stretch]	
First and second fingers (thumb and index) connected [1+2_Connect]	
First and second fingers (thumb and index) bent [1+2_Bent]	
All fingers bent [1-5_Bent]	
All fingers spread bent [1-5_SpreadBent]	
Fingers flapped down [2-5_FlappedDown]	
Fingers bent [2-5_Bent]	
All fingers connected [1-5_Connect]	
Other [Other]	Handshape is unclear or not included in the previous categories
Invisible [Invisible]	Hand is not visible from the video signal

Position *[RHand-Pos and LHand-Pos]*

The hand's position in gesture space is determined based on the schema below (from McNeill, 1992). The elements of the Controlled Vocabulary correspond to this schema.

Speakers often assume postures that are less straight as the one in the illustration: they may lean to the side, etc. In such cases the locations are identified with reference to the speaker's torso and the chair they are sitting in.

CENTER-CENTER [Center-C]
CENTER [Center]

PERI left [Peri-Left]
PERI lower left [Peri-LeftLow]
PERI lower [Peri-Low]
PERI lower right [Peri-RightLow]
PERI right [Peri-Right]
PERI upper right [Peri-RightUp]
PERI upper [Peri-Up]
PERI upper left [Peri-LeftUp]

EXTR left [X-Left]
EXTR lower left [X-LeftLow]
EXTR lower [X-Low]
EXTR lower right [X-RightLow]
EXTR upper right [X-RightUp]
EXTR right [X-Right]
EXTR upper [X-Up]
EXTR upper left [X-LeftUp]

Other [Other]

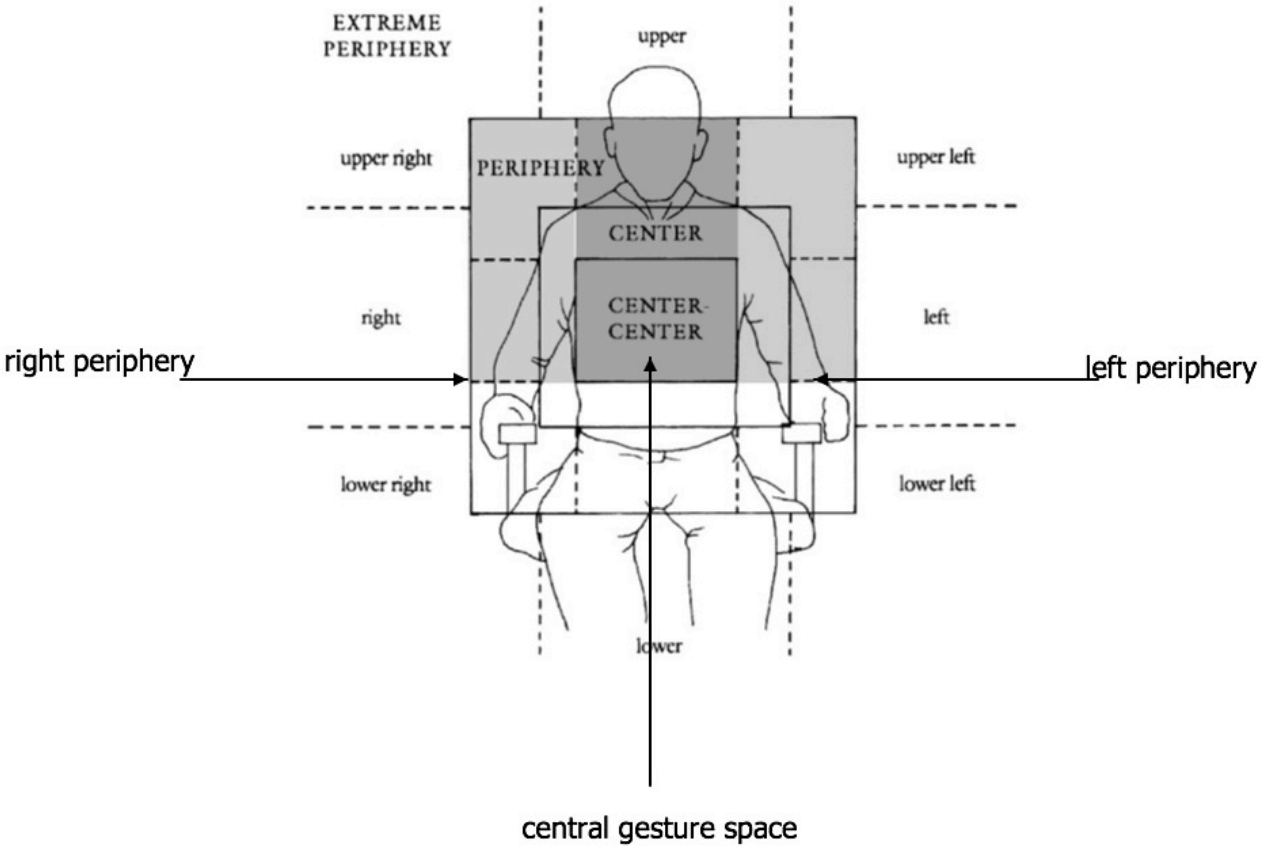
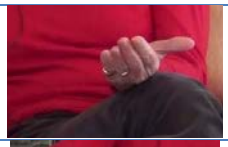
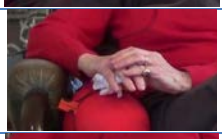
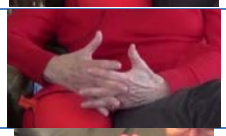
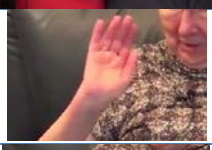
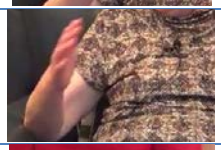
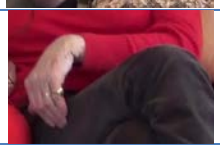


Figure from McNeill, 1992

Orientation *[RHand-O and LHand-O]*

The orientation of the hand is determined with reference to the palm. The elements of the Controlled Vocabulary are as follows:

[Up]	The palm faces upwards, towards the ceiling.	
[Down]	The palm is oriented towards the floor.	
[Back]	The palm is turned towards the speakers own chest.	
[Forward]	The palm faces forward, away from the speaker, towards the camera.	
[Side-In]	The palm is facing to the side, in the direction of the speaker's torso.	
[Side-Out]	The palm is turned to the side, away from the speaker's torso.	
[Invisible]	The orientation of the hand is not visible.	

Movement *[RHand-Mov and LHand-Mov]*

The movement of the hand is characterized as internal, external or a combination of the two, and as single or repeated. The resulting combinations are:

Single external [Ext]	A single movement of the hand, as a whole, from one location to another.
Single internal [Int]	A single hand internal movement: change of handshape, or movement of the fingers within a single handshape.
Repeated external [Ext-R]	The same external movement is repeated one or more times, possibly with short pauses.
Repeated internal [Int-R]	The same hand internal movement is repeated one or more times, possibly with short pauses.
External+internal [Ext+Int]	An external move combines with internal move(s), possibly with repetition.
Other [Other]	Other type of movement.
Invisible [Invisible]	The movement of the hand is not visible.

2.2.2. Contact [RHand-Contact and LHand-Contact]

The "Contact" Tier is annotated independently for each hand, delineating contact between the hand and other body parts or objects. The parent Tier contains only empty annotations, indicating the presence of contact with at least one body part/object. The characteristics of the contact are described on the child Tiers.

Child Tiers of *Contact*

If there are multiple contact points, the most salient or newest contact is annotated. For example, if one hand is resting in the lap and the other touches/rests on it, the hand is annotated as in contact with the lap, unless there is activity to engage the other hand (e.g. ageDA1r-1_sample2 00:02:11 – 00:02:18). If there is continuous contact with an object/body part, and another contact is then added, the new contact is annotated. For example, in ageBN1r-1_sample2, the speaker holds a tissue in her hand. This contact is initially annotated, as well as when there is not other contact, but if the hand holding the tissue is touched by the other hand, this new hand contact is annotated.

Target *[RHand-Contact-T and LHand-Contact-T]*

The target Tier contains annotations describing the actor that is contacted:

Self [Self]	Contact is made with the speaker's body, even if covered by clothing (unless acting on pieces of clothing directly).
Self-Object [Self-O]	Contact is made with an object that is attached to the speaker's body in a semi-permanent manner. E.g. Clothing (if it is clear that it is specifically the clothing that is being touched and not the body part covered by the clothing), jewelry, glasses.
Partner [Partner]	Contact is made with the partner's body.
Object [Object]	Contact is made with an object that is not attached to the speaker's body. E.g. Chair, glass, tissue.

Body/Object *[RHand-Contact-BO and LHand-Contact-BO]*

This Tier specifies further the body part or object that is in contact with the hand. The following items are included in the Controlled Vocabulary (closed list):

[Forehead]	[Mouth]	[Leg]
[Hair]	[Neck]	[Knee]
[Cheek]	[Shoulder]	[Lap]
[Chin]	[Chest]	[Wrist]
[Eyes]	[Abdomen]	[Object]
[Eyebrow]	[Arm]	[Other]
[Nose]	[Hand]	
[Ear]	[Fingers]	

If an object or self-object is contacted, the Controlled Vocabulary item "Object" is selected, and followed by a colon and the name of the object (closed list with open text):

E.g. [Object:Tissue], [Object:Ring]
🛠 In order to create the free description part of the annotation, hold down "Shift" and double click on the annotation.

Activity *[RHand-Contact-A and LHand-Contact-A]*

On the activity Tier, the activity of the hand is annotated (while in contact with another entity). The elements of the Controlled Vocabulary are described below.

Resting [Rest]	Hand in contact with another entity, without activity.
Touching [Touch]	Hand/finger touches something.
Percussion, Rhythmic [Percuss-R]	Percussive movement of hand/finger(s), even if only a single beat. No change of location. In the case of a single beat, percussion can be distinguished from touching by being in contact with other body part/object, then breaking and re-establishing contact. In the case of touching, there is no initial contact, and the peak of the movement establishes contact, which is then again broken.
Percussion, Distributed [Percuss-D]	Percussive movements in different spatial locations (multiple beats by definition).
Manipulation [Manip]	Hand manipulates object/body part. E.g. playing with folds in clothing or with ring.
Move [Move]	Object is moved from one location to other.
Rub [Rub]	Moving with a continuous (relatively "linear") contact (activity of rubbing, stroking, slipping, etc.)
Roll [Roll]	Rolling, turning, rotating, while in contact.
Scratch [Scratch]	Scratching movement.
Other [Other]	For example, hand moves slightly from wrist: ageDA1r-1_sample1 00:00:34.360 Also, when contact is still or re- established before/after a gesture, when the hand is not at rest anymore but some parts of it are still in contact with something else, and the movement cannot be qualified as rub, “Other” is used.

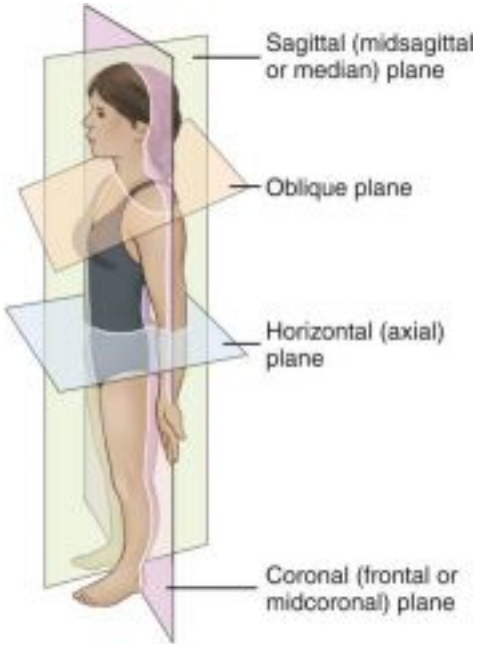
2.2.3. Symmetry [Hand_Symmetry]

On this Tier, symmetry of the two hands is annotated. Symmetry is present when the hands have quite similar handshape and their movement is symmetrical. The parent Tier serves to indicate the presence of symmetry (empty annotation), and the type of symmetry is described on the child Tiers.

Child Tiers of *Symmetry*

Plane *[Hand_Symmetry-Pl]*

On this Tier, the plane of the symmetry is annotated. The following figure illustrates the main planes in relation to the body, and the Controlled Vocabulary is listed below.



http://www.run3d.co.uk/announcements/what-is-pronation

[Frontal]	The movement of the hands is symmetrical along the frontal plane.
[Sagittal]	The movement of the hands is symmetrical on the sagittal plane.
[Horizontal]	The movement of the hands is symmetrical along the horizontal (transverse) plane.
[Point]	The hands move symmetrically around a single point.

Time *[Hand_Symmetry-S]*

The temporal symmetry and synchronization of hand movement is annotated on this Tier, that is, whether the movement is parallel or alternate.

[Parallel]	The hands are synchronized, showing parallel movements.
[Alternate]	The hands alternate and are not strictly synchronized (e.g., similar simultaneous moves, but going to an opposite direction).

3. Body gestures

3.1. Tier structure

Head [Head-F]
Torso [Torso]
Right Shoulder [RShoulder]
Left Shoulder [LShoulder]
Shoulders Symmetry [Sh_Symmetry]
Legs Position [Legs-Pos]
Legs Form [Legs-F]
Right Foot Movement [RFoot]
Left Foot Movement [LFoot]
Feet Symmetry [Feet_Symmetry]

Table 5. Tiers' structure for the annotation of body gestures (in ELAN)

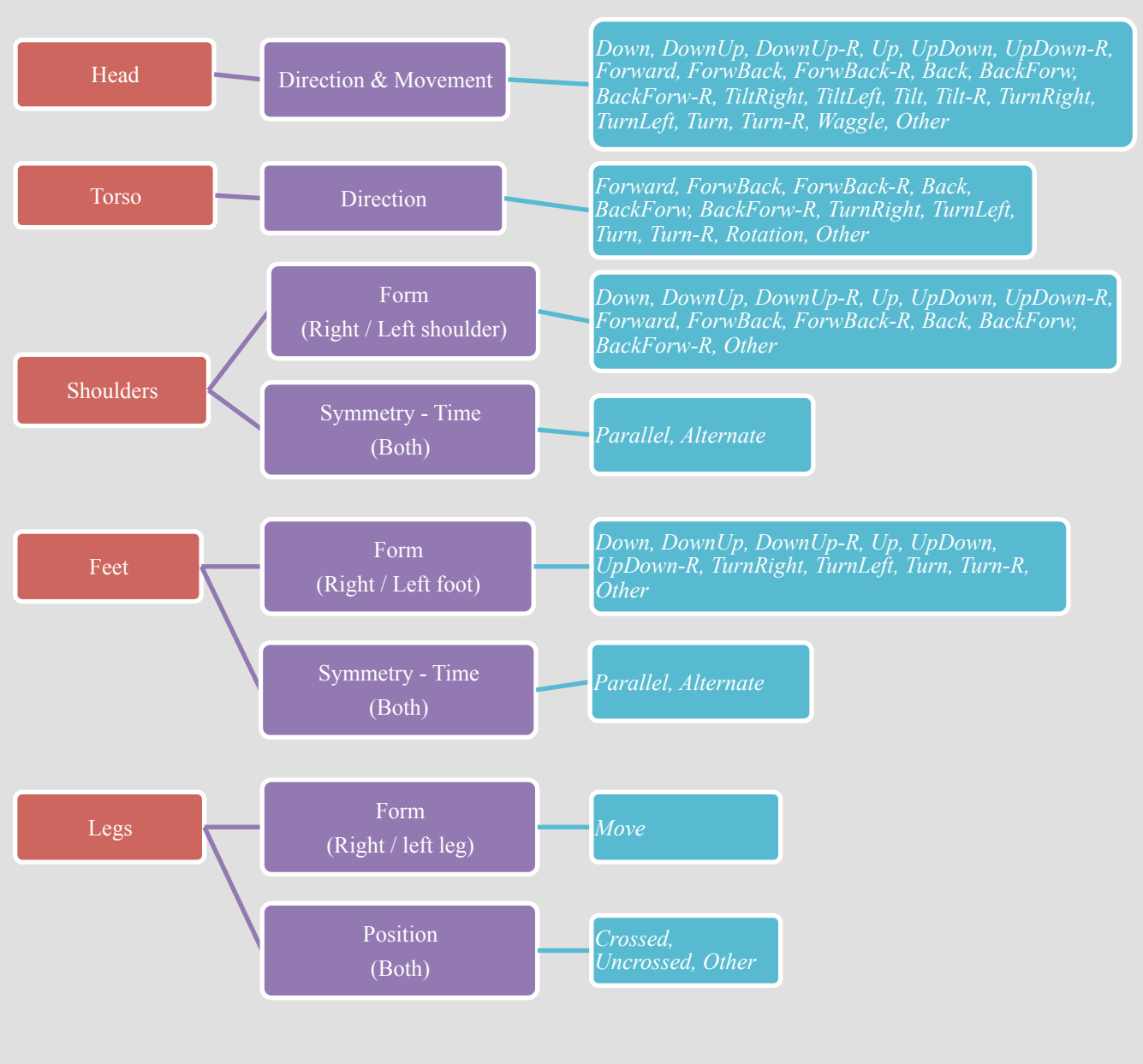


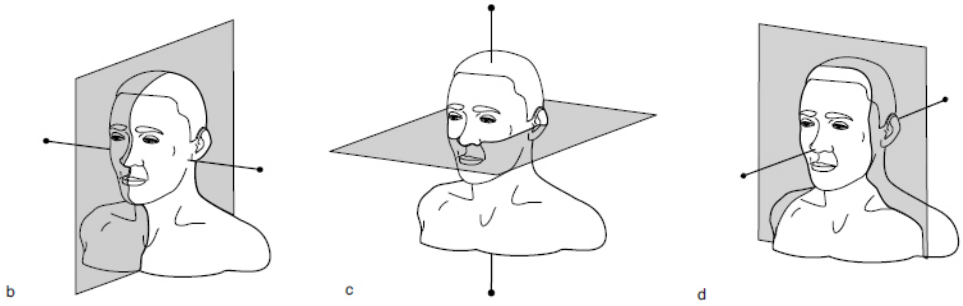
Figure 3. Tiers' structure and dependency between physiological parameters for body gestures (in ELAN)

3.2. Guidelines per Tier

3.2.1. Head [Head-F]

The description of head moves inspired by the MUMIN scheme (Allwood *et al.*, 2005, 2007) has been adapted in such a way to adhere more accurately to the "objectivity" principle adopted by form-based approaches to gesture, according to which what is annotated first is the physical move (without any interpretation of its potential meaning or function at that stage).

In line with systematic coding of body posture (e.g. Dael *et al.*, 2012), head moves are described according to the position and direction of the head in a three-dimensional space, with respect to the three orthogonal body planes (see figure below): a movement in the *sagittal* plane occurs relative to a *mediolateral* axis (perpendicular to the sagittal plane – e.g. "TiltLeft" or "Turn"); a movement in the *horizontal* plane occurs around a *vertical* axis (perpendicular to the horizontal plane – e.g., "UpDown-R"); and a movement in the *frontal* plane occurs relative to an *anteroposterior* axis (perpendicular to the frontal plane – e.g., "Back").



Three fundamental planes (b. sagittal, c. horizontal, d. frontal)
and three axes (b. mediolateral, c. vertical, d. anteroposterior) related to the human head
http://pocketdentistry.com/4-mechanical-principles-associated-with-removable-partial-dentures/

Annotation principles

What are the criteria to distinguish between two consecutive head gestures?

A change observed in the form and direction (e.g. "TiltRight", "TurnLeft", "Back", "Up", etc.) is usually sufficient to distinguish between two consecutive single head gestures. There are nevertheless particular cases, where it is the duration that will be used as the decision criterion:

- **Binary moves:** The limit for the maximum duration is 500ms for the whole move. If more than 500ms, then it must be segmented in two distinct moves (e.g., "Up" > "Down"). If one of the two segments remains too short to be noted (less than 250 ms.), the whole move is tagged as a single binary move (e.g., "UpDown").
  Rem.: When a move is less than 250 ms. but nevertheless appears to be *salient* in a quite obvious manner to the coder, it can be noted as a single move in this particular context.
- **Repeated binary moves**: At least one of the move within the repetition segment must be less than 250ms. In other words, the duration of at least one of the internal move (for example, "Up" or "Down") must be less than 250ms to be annotated as a single repeated "UpDown" move.
  Rem.: When binary moves are repeated but not exactly in the same manner, that is, if a significant change in *velocity* and/or *amplitude* is observed, then two distinct moves must be annotated (e.g., "Turn-R" > "Turn"). Again, if a change in *velocity* and/or *amplitude* is observed, but that no difference is expected in the meaning conveyed, then one single annotation value is attributed all along this move ("Turn-R").

What can I do when the gesture combines several *simultaneous* moves?

A. If two or more simultaneous external moves = "mixed external moves" (e.g. "Turn" and "Back"), adopt the following reasoning: If two external movements simultaneously occur without being able to decide between one or the other (that is, both seem to be meaningful), then they are both noted by alphabetical order in the annotation span (e.g., "Back+Down"): (i) select the first value automatically by means of the Controlled Vocabulary; (ii) add the second value manually by means of the "Shift"+"Double-click", see above).

Rem.: Ideally, one complex move cannot consist of more than two different moves. The two labels must be linked by the "+" character, to make every mixed annotation easily retrieved from ELAN *a posteriori* search.

B. If external move with simultaneous internal repeated movements (e.g. "Back" and "Turn-R"): When a single or binary move simultaneously combines with repeated internal movements (not necessarily salient), the annotation selected in the Controlled Vocabulary must be the one of the external move (e.g. "Back"). Then, the annotation of the internal repeated move (e.g., "UpDown-R") is added manually by maintaining the "Shift" button while double-clicking on the annotation to be changed. This means that it doesn't necessarily follow the alphabetical order. The resulting annotations would be, for example: "Down+DownUp-R", "Tilt+DownUp-R" or "DownUp+Turn-R".

Downward moves	
Down Move [Down]	Single downward head move (i.e. downward move followed by another move or a static head position). The direction of the move is from an "upper" position to a "lower" one on the vertical axis.
Down-Up Move (Nod) [DownUp]	Binary down-up move: downward move with a return phase.
Repeated Down-Up Moves (Nods) [DownUp-R]	Multiple, repeated quick head down-up move.
Upward moves	
Up Move [Up]	Single head upward move (i.e. upward move followed by another move or a static head position). The direction of the move is from a "lower" position to an "upper" one on the vertical axis.
Up-Down Move (Jerk) [UpDown]	Binary up-down move: upward move with a return phase.
Repeated Up-Down Moves (Jerks) [UpDown-R]	Multiple, repeated quick head up-down move.
Forward moves	
Move Forward [Forward]	Move of the head forwards. This is valid for moves of the head only, not when the move is initiated from the whole trunk (we do not follow Allwood *et al.*, 2005, 2007 with this respect).
Forward-Backward Move [ForwBack]	Binary forward-backward move: forward move with a return phase.
Repeated Forward-Back moves [ForwBack-R]	Multiple, repeated quick head forward-backward move.
Backward moves	
Move Backward [Back]	Move of the head backwards. This is valid for moves of the head only, not when the move is initiated from the whole trunk (we do not follow Allwood *et al.*, 2005, 2007 with this respect).
Backward-Forward Move [BackForw]	Binary backward-forward move: backward move with a return phase.
Repeated Backward-Forward moves [BackForw-R]	Multiple, repeated quick head backward-forward move.
Tilt moves	
Tilt to the Right (Sideway) [TiltRight]	Single move of the head leaning on the right side.
Tilt to the Left (Sideway) [TiltLeft]	Single move of the head leaning on the left side.
Tilt Return (Sideways) [Tilt]	Binary move of the head leaning on one side with a return phase.
Repeated Tilts (Sideways) [Tilt-R]	Multiple, repeated quick move of the head leaning from side to side.
Turn moves	
Side-turn to the Right [TurnRight]	Single move of the head turning to the right side.
Side-turn to the Left [TurnLeft]	Single movement of the head turning to the left side.
Side-turn Return [Turn]	Binary rotation of the head towards one side with a return phase.
Side-turn Repeated (Shake) [Turn-R]	Repeated, quick rotation of the head from one side to the other.
Waggle/Wobble	
Waggle/Wobble [Waggle]	Move of the head back and forth, side to side. It is like a mixture of shake and backward or forward move (see the Indian "wobble").
Other [Other]	Perceivable move, which is difficult to determined and cannot be classified according to the Controlled Vocabulary and principles adopted.

3.2.2. Torso [Torso]

Move Forward [Forward]	Forward move of the torso.
Move Backward [Back]	Backward move of the torso.
Move Sideways [Side]	Movement of the torso leaning on one side.
Rotation [Rotation]	Rotation of the torso toward one side.
Other [Other]	Another type of move, not in the above-listed categories.

3.2.3. Shoulders [RShoulder and LShoulder]

There are two main Tiers for each shoulder, "Left" and "Right".

Shoulder Up [Up]	A single movement of the shoulder up, usually coming from a neutral position (rest position).
Shoulder Down [Down]	A single movement of the shoulder down, usually coming from a neutral position (rest position).
Shoulder Up-Down [UpDown]	A movement of the shoulder upwards immediately followed by a downward movement. This annotation is only valid for quick binary moves (where at least the "Up" or the "Down" is less than 250ms).
Repeated Up-Down Moves [UpDown-R]	A multiple, repeated "Up-Down" binary move. This annotation is only valid for quickly repeated binary movements (where at least one move – "Up" or "Down" – is less than 250ms).
Shoulder Down-Up [DownUp]	A movement of the shoulder down, immediately followed by a upward movement. This annotation is only valid for quick binary moves (where at least the "Up" or the "Down" is less than 250ms).
Repeated Down-Up Moves [DownUp-R]	A multiple, repeated "Down-Up" binary move. This annotation is only valid for quickly repeated binary movements (where at least one move – "Up" or "Down" – is less than 250ms).
Move Forward [Forward]	A movement of the shoulder forward.
Return Forward-Backward Move [ForwBack]	A movement of the shoulder forwards immediately followed by a backward movement. This annotation is only valid for quick binary moves (where at least the "Forward" or the "Back" is less than 250ms).
Repeated Forward-Backward Moves [ForwBack-R]	A multiple, repeated "Forward-Back" binary move. This annotation is only valid for quickly repeated binary movements (where at least one move – "Forward" or "Backward" – is less than 250ms).
Move Backward [Back]	A movement of the shoulder backward.
Return Backward-Forward Move[BackForw]	A movement of the shoulder backwards immediately followed by a forward movement. This annotation is only valid for quick binary moves (where at least the "Forward" or the "Back" is less than 250ms).
Repeated Backward-Forward Moves [BackForw-R]	A multiple, repeated "Backward-Forward" binary move. This annotation is only valid for quickly repeated binary movements (where at least one move – "Forward" or "Backward" – is less than 250ms).
Other [Other]	A different type of movement than the above-mentioned.

Annotation principles

Annotation of laterality: what is "Right" and what is "Left"?

The "Right Shoulder" and "Left Shoulder" Tiers refer to the body parts of the speaker (from his/her perspective), and do not correspond to the "right" or "left" side of the screen (from the point of view of the annotator).

Which shoulder moves have to be taken into account?

- When the shoulder move is due to the activity of breathing/laughing, the move is not taken into account because it is not initiated from this part of the body but a consequence of the breath/laugh activity.

- Particular case: When a “Down” move corresponds to a return phase from a “Up” move, it has to be understood as being intimately linked to the physiological and articulatory constraints (law of gravity!), and therefore must not be selected for annotation. Nevertheless, if the “Down” move is obviously perceivable by the annotator as being part of a binary “Up-Down” move (here, velocity of the “Down” move is a criterion of selection), it has to be taken into account.
- Segmentation of the “Up” into “Up-Down”: when less than 250ms per segment/move, the phase is selected as an isolated move annotated “Up-Down”: the two phases are integrated into a single binary “Up-Down” move (e.g., where the “Up” phase is 238ms and the “Down” is 538ms, it must be annotated as a single “Up-Down” move).
  Exception: When an “up” or “down” move is less than 250ms but still clearly visible, then it needs to be taken into account for annotation.

3.2.4. Shoulders’ Symmetry [Shoulder_Symmetry]

On this Tier, the annotation of every symmetric move of the two shoulders follows from a two-step procedure. First, every synchronous move is automatically detected by means of searching every (partial) overlap of the Left and Right shoulder’s move, thus creating a new Tier comprised of all overlapping shoulders’ moves (be they symmetric or not). Secondly, every annotation is manually checked and a child Tier is created if any symmetry is detected between the two shoulders.

As for the hands, the parent Tier serves to indicate the presence of symmetry (empty annotation), and the type of symmetry is described on the child Tiers. However, as the shoulders’ moves are simpler than hands’ moves (in amplitude and form), there is only one child Tier dedicated to the temporal dimension distinguishing between parallel and alternate moves.

Child Tiers of *Symmetry*

Time *[Shoulder_Symmetry-S]*

The temporal symmetry and synchronization of shoulder movement is annotated on this Tier, that is, whether the movement is parallel or alternate.

[Parallel]	The shoulders are synchronized, showing parallel movements.
[Alternate]	The shoulders alternate and are not strictly synchronized (e.g., similar simultaneous moves, but going to an opposite direction).

3.2.5. Feet [RFoot and LFoot]

There are two main Tiers for each shoulder, “Left” and “Right”. Similarly to head moves, feet moves are described according to the position and direction of the foot in space, with respect to two orthogonal body planes: a movement in the *sagittal* plane occurs relative to the *mediolateral* axis (perpendicular to the sagittal plane – e.g. “Turn”); a movement in the *horizontal* plane occurs around the *vertical* axis (perpendicular to the horizontal plane – e.g., “UpDown-R”). Note that the *frontal* plane – which relates to possibly occurring Backward and Forward moves – has not been included in the tagset for feet, since these moves were too rare to be considered as such in the analysis (then annotated “Other” if any observed).

Down Move [Down]	Single downward foot move (i.e. downward move followed by another move or a static foot position). The direction of the move is from an “upper” position to a “lower” one on the vertical axis.
Down-Up Move [DownUp]	Binary down-up move: downward move with a return phase.
Repeated Down-Up Moves [DownUp-R]	Multiple, repeated quick foot down-up move.
Up Move [Up]	Single foot upward move (i.e. upward move followed by another move or a static foot position). The direction of the move is from a “lower” position to an “upper” one on the vertical axis.

Up-Down Move [UpDown]	Binary up-down move: upward move with a return phase.
Repeated Up-Down Moves [UpDown-R]	Multiple, repeated quick foot up-down move.
Side-turn to the Right [TurnRight]	Single move of the move turning to the right side.
Side-turn to the Left [TurnLeft]	Single movement of the foot turning to the left side.
Side-turn Return [Turn]	Binary move of the foot towards one side with a return phase.
Side-turn Repeated [Turn-R]	Repeated, quick move of the foot from one side to the other.
Other [Other]	Perceivable move, which is difficult to determined or cannot be classified according to the Controlled Vocabulary and principles adopted.

Annotation principles

Annotation of laterality: what is "Right" and what is "Left"?

The "Right Foot" and "Left Foot" Tiers refer to the body parts of the speaker (from his/her perspective), and do not correspond to the "right" or "left" side of the screen (from the point of view of the annotator).

Which foot moves have to be taken into account?

- Less perceivable feet moves are taken into account, because their narrower amplitude does not *a fortiori* imply a less salient semantic-pragmatic weight (as compared to other articulators). For example, every movement of the toes in the shoe, if perceivable, has been annotated, as it may potentially act as a beat in the interaction flow (see the multimodal annotation scheme for pragmatic functions developed by Bolly & Crible, 2015).
- Segmentation of the "Up" into "Up-Down": when less than 250ms per segment/move, the phase is selected as an isolated move annotated "Up-Down": the two phases are integrated into a single binary "Up-Down" move (e.g., where the "Up" phase is 238ms and the "Down" is 538ms, it must be annotated as a single "Up-Down" move).
  Exception: When an "up" or "down" move is less than 250ms but still clearly visible, then it needs to be taken into account for annotation.

Annotation of "Down" moves that are perceived as very "slow" return moves

- When a "Down" move corresponds to a return phase from a "Up" move, it has to be understood as being intimately linked to the physiological and articulatory constraints (law of gravity!), and therefore must not be selected for annotation. Nevertheless, if the "Down" move is obviously perceivable by the annotator as being part of a binary "Up-Down" which is potentially meaningful as a whole, it has to be taken into account.

How to distinguish between "Down" moves that are saccadic (one single move with a quick pause) or that consist in two distinct consecutive moves?

- As for the other articulators, a saccade is defined as an abrupt movement that consists in a change of position that does not entail a change of any physiological parameter (e.g., direction, amplitude, velocity). This is often characterized by the presence of a very quick pause within a continuous move. In such cases, the move is annotated as a single move and must not be divided into two distinct (repeated) moves. Yet, if any perceivable and salient change in velocity, the move has to be segmented into two distinct moves of the foot.

How to annotate a move of the foot when the leg is moving at the same time?

- When the foot simultaneously moves with the leg without any change of its position or direction (for example, it remains in a flex-position with a regular angle) it is not taken into account, because the move is not actively initiated from the part of the body concerned.

3.2.6. Feet's Symmetry [Feet_Symmetry]

On this Tier, as for the shoulders, the annotation of every symmetric move of the two feet follows from a two-step procedure. First, every synchronous move is automatically detected by means of searching every (partial) overlap of the Left and Right foot's move, thus creating a new Tier comprised of all overlapping foot moves (be they symmetric or not). Secondly, every annotation is manually checked and a child Tier is created if any symmetry is detected between the two feet.

As for the hands, the parent Tier serves to indicate the presence of symmetry (empty annotation), and the type of symmetry is described on the child Tiers. However, as the feet moves are simpler than hands' moves (in amplitude and form), there is only one child Tier dedicated to the temporal dimension distinguishing between parallel and alternate moves.

Child Tiers of *Symmetry*

Time *[Feet_Symmetry-S]*

The temporal symmetry and synchronization of feet movement is annotated on this Tier, that is, whether the movement is parallel or alternate.

[Parallel]	The feet are synchronized, showing parallel movements.
[Alternate]	The feet alternate and are not strictly synchronized (e.g., similar simultaneous moves, but going to an opposite direction).

3.2.7. Legs' move [RLeg-Mov and LLeg-Mov]

There are two main Tiers for each leg, "Left" and "Right". Every move of the legs has been identified and labeled with the "Move" tag.

[Move]	Visible movement of one leg.

Annotation principles

Annotation of laterality: what is "Right" and what is "Left"?

The "Right Leg" and "Left Leg" Tiers refer to the body parts of the speaker (from his/her perspective), and do not correspond to the "right" or "left" side of the screen (from the point of view of the annotator).

Which legs' moves have to be taken into account?

Due to the constrained body position of the speakers (who sit in chairs), moves of the legs mostly correspond to a change in the global body posture in the CorpAGEst data. For that reason, no precise description of the move has been added in the template. The label "Move" is therefore attributed to any visible move of the leg.

3.2.8. Legs' position [Legs-Pos]

The legs' position has been characterized as being "Crossed" or "Uncrossed" (or "Other" for undetermined cases), independently of the "Legs' move" annotation, all along the video sample under scrutiny.

[Crossed]	The legs are crossed.
[Uncrossed]	The legs are not crossed.
[Other]	Undetermined position.

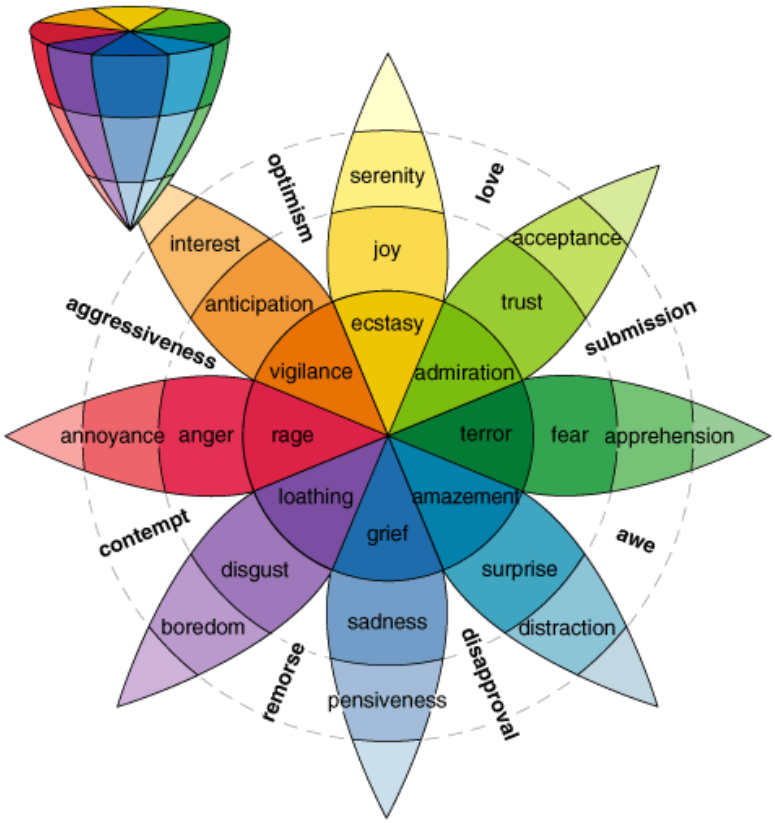
4. Emotion from the face

4.1 Emotion labeling

It is the Plutchik's multidimensional model (1980, 2001) that serves as a basis for the attribution of emotion to facial expressions within the framework of the CorpAGEst Marie Curie project. The model is based on eight primary emotional dimensions, which are organized in polarity dyads (e.g., *Ecstasy* as opposed to *Grief*), declined into several combinations (e.g., *Love* resulting from the combination of *Trust* and *Joy*), and nuanced according to their degree of intensity in tryads (e.g., *Acceptance* – *Trust* – *Admiration*, from weakest to strongest). It is recommended to be aware of polarity values and intensity levels while annotating emotions (see Table below), insofar as emotions are linked and identified by contrasting them with each other during the visualization task.

Negative affect (weakest > strongest)	Positive affect (weakest > strongest)
Apprehension > Fear > Terror Distraction > Surprise > Amazement Pensiveness > Sadness > Grief Boredom > Disgust > Loathing Annoyance > Anger > Rage Neutral in intensity: Submission, Awe, Disapproval, Remorse, Contempt, Aggressiveness, Nervousness, Disappointment	Attention > Interest > Vigilance Serenity > Joy > Ecstasy Acceptance > Trust > Admiration Neutral in intensity: Optimism, Love, Nostalgia

The Controlled Vocabulary comprises 32 emotions (from the Plutchik's model – see Figure below), with 3 additional recurrent emotions (*viz. Nervousness*, *Disappointment* and *Nostalgia*) that emerged from the video data analysis.



Plutchik's circumplex and wheel of emotions (reproduced from www.6seconds.org)

See below the 35 emotion tags listed by alphabetic order (with their corresponding level of intensity and polarity value):

1. Acceptance [Acceptance]	Intensity: low	Valence: positive
2. Admiration [Admiration]	Intensity: high	Valence: positive
3. Aggressiveness [Aggressiveness]	Intensity: neutral	Valence: negative
4. Amazement [Amazement]	Intensity: high	Valence: negative
5. Anger [Anger]	Intensity: medium	Valence: negative
6. Annoyance [Annoyance]	Intensity: low	Valence: negative
7. Apprehension [Apprehension]	Intensity: low	Valence: negative
8. Attention (here: low level of interest) [Attention]	Intensity: low	Valence: positive
9. Awe [Awe]	Intensity: medium	Valence: negative
10. Boredom [Boredom]	Intensity: low	Valence: negative
11. Contempt [Contempt]	Intensity: medium	Valence: negative
12. Disappointment [Disappointment]	Intensity: neutral	Valence: negative
13. Disapproval [Disapproval]	Intensity: neutral	Valence: negative
14. Disgust [Disgust]	Intensity: medium	Valence: negative
15. Distraction [Distraction]	Intensity: low	Valence: negative
16. Nervousness [Nervousness]	Intensity: neutral	Valence: negative
17. Ecstasy [Ecstasy]	Intensity: high	Valence: positive
18. Fear [Fear]	Intensity: medium	Valence: negative
19. Grief [Grief]	Intensity: high	Valence: negative
20. Interest, Anticipation (here: medium level of interest) [Interest]	Intensity: medium	Valence: positive
21. Joy [Joy]	Intensity: medium	Valence: positive
22. Loathing [Loathing]	Intensity: high	Valence: negative
23. Love [Love]	Intensity: medium	Valence: positive
24. Nostalgia [Nostalgia]	Intensity: neutral	Valence: positive
25. Optimism [Optimism]	Intensity: neutral	Valence: positive
26. Pensiveness [Pensiveness]	Intensity: low	Valence: negative
27. Rage [Rage]	Intensity: high	Valence: negative
28. Remorse [Remorse]	Intensity: neutral	Valence: negative
29. Sadness [Sadness]	Intensity: medium	Valence: negative
30. Serenity [Serenity]	Intensity: low	Valence: positive
31. Submission [Submission]	Intensity: neutral	Valence: negative
32. Surprise [Surprise]	Intensity: medium	Valence: negative
33. Terror [Terror]	Intensity: high	Valence: negative
34. Trust [Trust]	Intensity: medium	Valence: positive
35. Vigilance [Vigilance]	Intensity: high	Valence: positive
Other [Other]		

4.2 Guidelines per Tier

4.2.1. Facial emotions

“Emotions are grounded in the here-and-now experience of communicating participants. (…) [T]he recognition of emotional patterns – that is, the attribution of one specific emotion to physiological patterning – is a complex process (Russell et al., 2003: 334) depending on the situational context, on the affective state, social and cultural identity of the participants, as well as on their (individual and shared) background knowledge. (…) [F]acial expressions are particularly recognized as a major conveyance of stance in a situation of communicative interaction.” (Bolly & Thomas, 2015)

Annotation principles

Steps in the annotation process

The multi-level annotation of emotion was performed as follows: (i) annotation of the physiological parameters for the face; (ii) annotation of emotions expressed through the face (without any recourse to the sound signal); (iii) annotation of the relation between the tagged emotion and the contextual information (taking into account gestures and linguistic information).

How many coders?

All annotations were done by one investigator and partly crosschecked by a second investigator mainly during the learning phase, in order to develop, improve and stabilize the coding scheme. The second investigator also served as control for uncertain and ambiguous cases. Note, however, that the reproducibility of the annotation procedure is very limited, mainly due to the highly subjective dimension of emotion perception. It is therefore recommended either to assign the annotation process to one single coder (favoring homogeneity across the annotated data) or to develop a systematic control of annotation by (at least) another coder on the same data (favoring reliability of the model).

With or without access to contextual cues and sound signal?

The annotation of emotions was done without any access to the previously annotated physiological parameters, in order to guarantee maximal objectivity in what can be recognized from face only. This methodological choice is directly connected to particular research objectives, which are at the core of one pilot study carried out within the framework of the CorpAGEst project (see Bolly & Thomas, 2015). However, this Controlled Vocabulary could obviously apply to any other audio and video data, taking into account the sound signal, the contextual cues and the whole body, in a *de facto* multimodal approach to interaction.

How to identify "emotional units"? Where to put the boundaries?

The criterion is fully subjective and perceptive. The coder is asked to identify one "emotion" by focusing only on facial expressions while annotating the video data. As a consequence, the boundaries of emotion tags do not correspond to the ones of physiological features identified for facial expressions (*viz.* "Eyebrow", "Eye", "Gaze", and "Mouth"). Emotional spans are thus mostly comprised of numerous physiological tags (e.g., one single emotion identified as expressing *trust* may include successive eye-closing moves and changes in gaze direction).

Child Tiers of *Facial emotions*

Semantic relationship with context

This Tier points to the semantic relation between emotions perceived from the face and their context of appearance (including the whole body, the linguistic and extralinguistic context). The Controlled Vocabulary is includes 5 types of contextual semantic relations (mainly inspired by Colletta *et al.*, 2009):

Redundant [Redundant]	A similar emotion (even though not necessarily synchronous) is expressed from the face and from the linguistic context.
Complementary (to linguistic information) [Complementary]	The facial emotion is compatible with and adds some value to the linguistic information conveyed (e.g., modalization, emphasis, hedge, specification, elaboration, etc.).
Contradictory [Contradictory]	The facial emotion is not compatible with the linguistic information conveyed.
Independent [Independent]	There is no relation between the two modes, which fulfill their proper function in the language interaction.
Accordant (with extralinguistic information) [Accordance]	Facial emotions are in accordance with information transmitted by the extralinguistic context at large (e.g., as a reaction to external stimuli such as noises).

References

Allwood, Jens, Cerrato, Loredana, Dybkjaer, Laila, Jokinen, Kristiina, Navarretta, Constanza and Patrizia Paggio (2005). *The MUMIN multimodal coding scheme (v3.3)*. *NorFA yearbook 2005*: 129-157.

Allwood, Jens, Cerrato, Loredana, Jokinen, Kristiina, Navarretta, Constanza and Patrizia Paggio (2007). The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation* 41(3-4): 273-287.

Dister, Anne, Francard, Michel, Geron, Geneviève, Giroul, Vincent, Hambye, Philippe, Simon, Anne Catherine and Régine Wilmet (2007 [2009]). *Conventions de transcription régissant les corpus de la banque de données Valibel*. Available online at: http://www.uclouvain.be/valibel.html

Boersma, Paul and David Weenink (2014). *Praat: Doing phonetics by computer* [Computer program]. Version 5.4.01, retrieved 9 November 2014.

Bolly, Catherine T. and Dominique Boutet (2015, submitted). The multimodal CorpAGEst corpus: Keeping an eye on pragmatic competence in later life.

Bolly, Catherine and Ludivine Crible (2015). From context to functions and back again: Disambiguating pragmatic uses of discourse markers. Paper presented at the panel session "Anchoring utterances in co(n)text, argumentation, common ground" (Org.: K. Fischer, M. Alm), *14th International Pragmatics Conference (IPra)*, 26-31 July 2015, Antwerp, Belgium.

Bolly, Catherine T., Crible, Ludivine, Degand, Liesbeth and Deniz Uygur-Distexhe (forthc.). Towards a Model for Discourse Marker Annotation in spoken French: From potential to feature-based discourse markers.

Bolly, Catherine T., Crible, Ludivine, Degand, Liesbeth and Deniz Uygur-Distexhe (2015). MDMA. Identification et annotation des marqueurs discursifs "potentiels" en contexte. *Discours* 16. Online journal: http://discours.revues.org/index.html

Bolly, Catherine T., Gabarró-López, Silvia and Laurence Meurant (2015). Pragmatic gestures at the gesture-sign interface. Nonmanuals and palm-up gestures among older Belgian French speakers and French Belgian Sign Language signers". Paper presented at the International Workshop *Nonmanuals at the Gesture Sign Interface (NaGSI)*, 9-10 October 2015, University of Göttingen, Germany.

Bolly, Catherine T. and Anaïs Thomas (2015). Facing Nadine's speech. Multimodal annotation of emotion in later life. In Kristiina Jokinen & Martin Vels (eds.), *Proceedings of the 2nd European and the 5th Nordic Symposium on Multimodal Communication August 6-8, Tartu, Estonia*. Linköping: Linköping Electronic Conference Proceedings 110, 23-32, http://www.ep.liu.se/ecp_home/index.en.aspx?issue=110.

Bressem, Jana (2008). Notating gestures. Proposal for a form based notation system of coverbal gestures. *Unpublished manuscript.* Available online at: http://janabressem.de.

Bressem, Jana (2013). A linguistic perspective on the notation of form features in gestures. In Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill and Sedinha Teßendorf (eds.), *Body – Language – Communication (Vol. 1)* (Handbooks of Linguistics and Communication Science 38/1). Berlin, New York: Mouton de Gruyter, 1079-1098.

Colletta, Jean-Marc, Kunene, Ramona N., Venouil, Aurélie, Kaufmann, Virginie and Jean-Pascal Simon (2009). Multi-track annotation of child language and gestures. In Michael Kipp, Jean-Claude Martin, Patrizia Paggio and Dirk Heylen (eds.), *Multimodal corpora* (*Lecture Notes in Computer Science* 5509). Berlin, Heidelberg: Springer, 54-72.

Dael, Nele, Marcello Mortillaro and Klaus R. Scherer (2012). The body action and posture coding system (BAP): Development and reliability. *Journal of Nonverbal Behavior* 36(2): 97-121.

Ferré, Gaëlle (2012). Critères de segmentation de la gestualité co-verbale. In *Proceedings of JEP-TALN-RECITAL 2012, Atelier DEGELS 2012 : Défi GEste Langue des Signes, Grenoble*, 9-21.

Goldman, Jean-Philippe (2011). EasyAlign: An automatic phonetic alignment tool under Praat. In *Proceedings of InterSpeech*, September 2011, Firenze, Italy.

Jokinen, Kristiina, Nishida, Masafumi and Seiichi Yamamoto (2009). Eye-gaze experiments for conversation monitoring. In Proceedings of the 3rd *International Universal Communication Symposium (IUCS'09)*. New York: ACM, 303–308.

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

Müller, Cornelia, Bressem, Jana and Silva H. Ladewig (2013). Towards a grammar of gestures: A form-based view. In Cornelia Müller, Alan CIenki, Ellen Fricke, Silva H. Ladewig, David McNeill and Sedinha Teßendorf (eds.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction* (Vol.1, Chapter 45). Berlin, Boston: De Gruyter Mouton, 707-733.

Plutchik, Robert (1980). *Theories of Emotion. Vol. 1: Emotion: Theory, Research, and Experience*. New York: Academic.

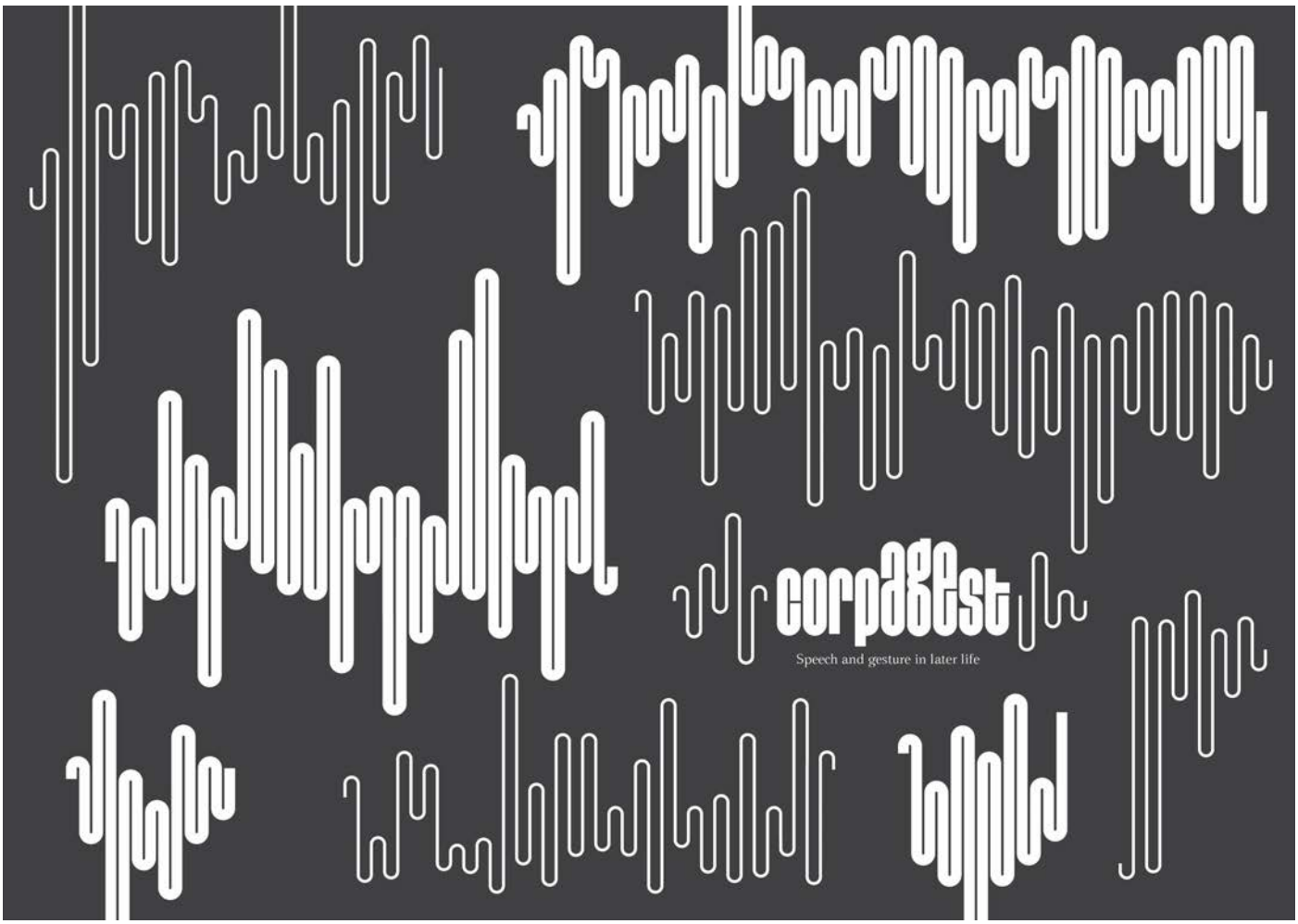
Plutchik, Robert (2001). The nature of emotions. *American Scientist* 89: 344-350.

Wittenburg, Peter, Brugman, Hennie, Russel, Albert, Klassmann, Alex and Hans Sloetjes (2006). Elan: A professional framework for multimodality research. In *Proceedings of LREC 2006*, Fifth International Conference on Language Resources and Evaluation. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands, http://tla.mpi.nl/tools/tla-tools/elan/.

Corpus

Bolly, Catherine T. (2013). *CorpAGEst. Multimodal corpus for the elderly's language*. Louvain-la-Neuve, Paris: Université catholique de Louvain (Valibel - Discours et Variation) and CNRS (UMR 7023 Structures Formelles du Langage), http://corpagest.org.

CLARe. *Corpora for Language and Aging Research*. Website: http://wikis.fu-berlin.de/display/clare/HOME.



Original Author: Catherine Bolly
With the collaboration of: Delphine Belin, Dominique Boutet, Julie Kairet, Alysson Lepeut, Anna Sáfár, Anaïs Thomas

Contact: catherine.bolly(at)uclouvain.be; cbolly(at)uni-koeln.de
Web: http://corpagest.org