"Russian Pear Chats and Stories": Manual Annotation Guide

(Version 14.12.2018)

Alla O. Litvinenko, Julia V. Nikolaeva, Andrej A. Kibrik, Olga V. Fedorova[1]

1. Introduction

As the main material for manual annotation in the Russian Pear Chats and Stories corpus, we used the data taken by three individual cameras which made full face footage of the three participants of the recording ("individual videos").

As supplementary material, we used the data obtained by the cover shot camera ("cover shot video"). Comparing the footage taken from two different angles allows us to define the boundaries of separate movements, as well as each movement's trajectory, amplitude, and other parameters with more precision.

It is important to note that the primary manual annotation is done entirely on a kinetic basis, excluding the verbal and prosodic components of the recording and thus any knowledge of the contents of the conversation. This approach ensures creating independent annotations for all channels, so that they can be further studied in their correlation without creating a loop. Thus, the first stage of the manual annotation procedure is carried out with the sound of the cover shot video muted, and it is only later that the speech context is used for the annotation of functional objects and for verification.

Annotation is performed using ELAN Linguistic Annotator developed by the Max Planck Institute for Psycholinguistics. This software can be downloaded via the following link: https://tla.mpi.nl/tools/tla-tools/elan/download/ ELAN has a vast inbuilt manual in English which not only offers a detailed description of the software interface and functions but also proposes useful tips on the annotation process.

For the annotation process, the following media pack is used:

(i) 2 video files:

- cover shot camera footage where all participants appear in the shot (a file with a W letter in the name, or a W-file);
- one of the individual files:
  - footage from the Narrator's individual camera (N-file)
  - footage from the Commentator's individual camera (C-file)
  - footage from the Reteller's individual camera (R-file)

(ii) a ready template for ELAN software (.eaf) alongside with its utility file (.pfsx)

Consulting the cover shot footage is necessary when the video taken by an individual camera does not provide enough information to estimate gesture range and velocity (e.g., some gesture is made directly towards the individual camera or straight away from it). If the computer capacity does not allow for simultaneous opening of both video files in ELAN, it is advised to remove the cover shot video from the list of Linked files (ELAN > Edit menu > Linked files) and to open it separately using any other software with a frame-by-frame view option (e.g., VirtualDub: http://www.virtualdub.org/).

The annotation scheme is built on sequences of hand movements and stillness (inaction) periods. Separate hand movements form bigger units, of which the most important ones are communicatively significant objects, or manual gestures.

Apart from gestures, the annotation system also includes:

- various non-speech-motivated movements, so-called adaptors, usually aimed at maintaining the speaker's physical comfort;

[1] To cite this version:
Litvinenko, A.O., Nikolaeva, Ju.V., Kibrik, A.A., Fedorova, O.V. "Russian Pear Chats and Stories": Manual Annotation. Version 14.12.2018. http://multidiscourse.ru

- periods of maintaining the same reference ("neutral") hand position, or manual postures;
- movements aimed at changing or correcting the hand position, or manual posture changes and accommodators.

The annotation procedure is described step-by-step in Section 2.

Detailed instructions on the kinetic units' segmentation at different levels and the annotation rules are explained in Section 3.

Section 4 (appendixes) offers concise descriptions of the necessary ELAN tiers set along with their characteristics and corresponding controlled vocabularies.

2. The annotation procedure

2.0. General description

The annotation procedure can be described as a series of several successive stages; at each stage, kinetic units of different levels are annotated (see Figure 1 below). The annotation is carried out in a 'bottom-up' direction, starting from separate movements and the stillness intervals between them, then proceeding to functional objects (gestures, adaptors, posture changes and posture accommodators) consisting of one or more movements, and then to more complex structures comprised of one or several functional units (postures, gestures, and movement chains).

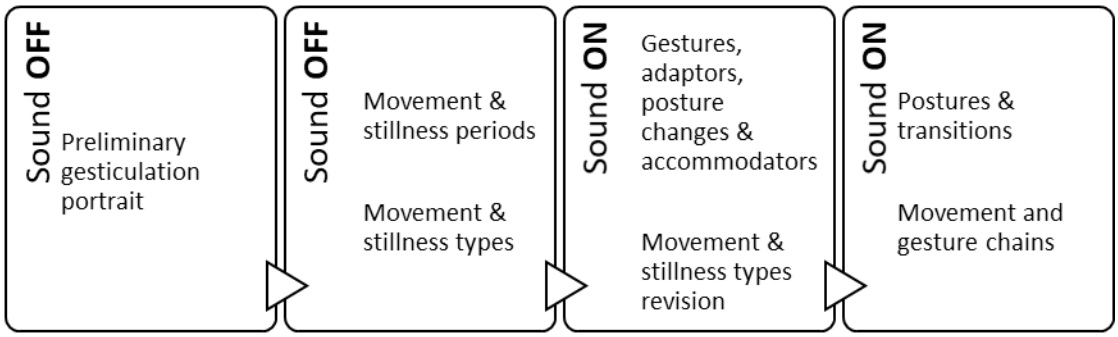

Figure 1. Annotation procedure stages (Preliminary Stage > Stage 1 > Stage 2 > Stage 3).

Below we will briefly describe each of the depicted stages.

2.1. Preliminary stage and preliminary gesture portrait

The purpose of the preliminary stage is the visual detection of the kinetic patterns that are typical for the speaker: estimation of the average velocity, amplitude, and effort for gestures and adaptors; the frequency and typical duration of stillness periods; and the list of typical neutral positions and posture changing movements. This information allows annotators to provide similar descriptions in similar cases and to describe difficult situations using the same approach. The preliminary stage of annotation is carried out with the sound muted.

Below you will find a list of sample questions that may help an annotator to compile a preliminary gesticulation portrait.

- How many gestures does the speaker use? (many, average, few)
- How many articulated (structured) adaptors does the speaker use? (many, average, few)
- How many subtle (unstructured) adaptors does the speaker use? (many, average, few)
- What is the typical duration of stillness between the movement chains? (high, middle, low)
- What is the typical velocity of the movements? (high, middle, low)
- What is the typical amplitude of the movements? (high, middle, low)

- What are the typical neutral positions? (For example: hands laying palm down on the knees; arms crossed at the chest; palms pressed against the chair; chin resting on the fist; hands staying in the air in front of the torso, etc.)
- Are there any recurrent (repeating) gestures? (For example: spreading one's hands; pointing one's elbow at the interlocutor; snapping fingers; moving an arm with an open palm towards the interlocutor; rotating cupped hand with the knuckles facing the interlocutor, etc.)
- Are there any frequent repetitive adaptors? (For example: coughing and covering one's mouth; adjusting glasses; fumbling with the hem of the dress; rubbing knees; moving thumbs back and forth with clasped hands, etc.)

Such a list helps to treat difficult cases in the same way, using the same approach, during the annotation process. This list can be expanded or corrected as the annotation progresses, and the decisions can be reviewed accordingly; in this case, all previous entries of some revised occurrence should also be corrected. For instance, in the preliminary stage, some typical movement might be listed as an adaptor, but later the annotator after re-evaluating various factors in context might reconsider and treat it as a gesture; the list helps to keep track of all similar movements and to revise them accordingly.

2.2. The first stage

At the first stage of annotation, the whole flow of kinetic behavior is segmented into separate movements with the stillness periods between them. The annotation is done without relying on the speech content (with sound muted), thus minimizing the influence of speech unit boundaries on the decisions regarding kinetic unit boundaries. At this stage, the annotator's decisions are based on the visual evaluation of different kinetic characteristics of a movement, as well as various motion detection software data, if available.

2.3. The second stage

At the second stage, we annotate second level units (gestures, adaptors, posture changes and posture accommodators), assigning all the corresponding characteristics (handedness, phase structure, multi-strokes or rebounds for gestures; adaptor type for adaptors, etc.).

At this and in the next stage, annotation is done with the sound on, and the speech content is used for the verification of the earlier annotation. Boundaries and the types of basic level units may also be corrected; however, it is preferable to avoid situations when speech unit boundaries influence the manual ones. However, in some difficult situations that allow different interpretations, semantic verification can lead to a recharacterization of some secondary movement (preparation, retraction) into a stroke, or to the division of one multi-stroke into two or more separate strokes, etc.

In all such cases, relying on the information from other channels is acceptable when, with all other factors being equal, the kinetic characteristics allow for two equally valid solutions in the same context.

For example, the speaker makes two very similar movements in a row with little difference in their velocity and amplitude, and the annotator has to make a decision as to whether the given sequence consists of a multi-stroke or two separates strokes (and, consequently, two gestures). If from the point of view of kinematics this sequence is not smooth enough to be considered a single movement, but its elements do not differ enough to constitute two separate movements, then the speech content and/or the prosodic organization of the corresponding verbal context can be taken into consideration.

2.4. The third stage

The third stage serves for annotating manual postures, and also for creating gesture chain and movement chain tiers, using built-in ELAN functions.

The resulting network of units can be seen in Figure 2. Starting from the basic first level units in the middle, the system expands to the second level functional units and then to the third level composite units.

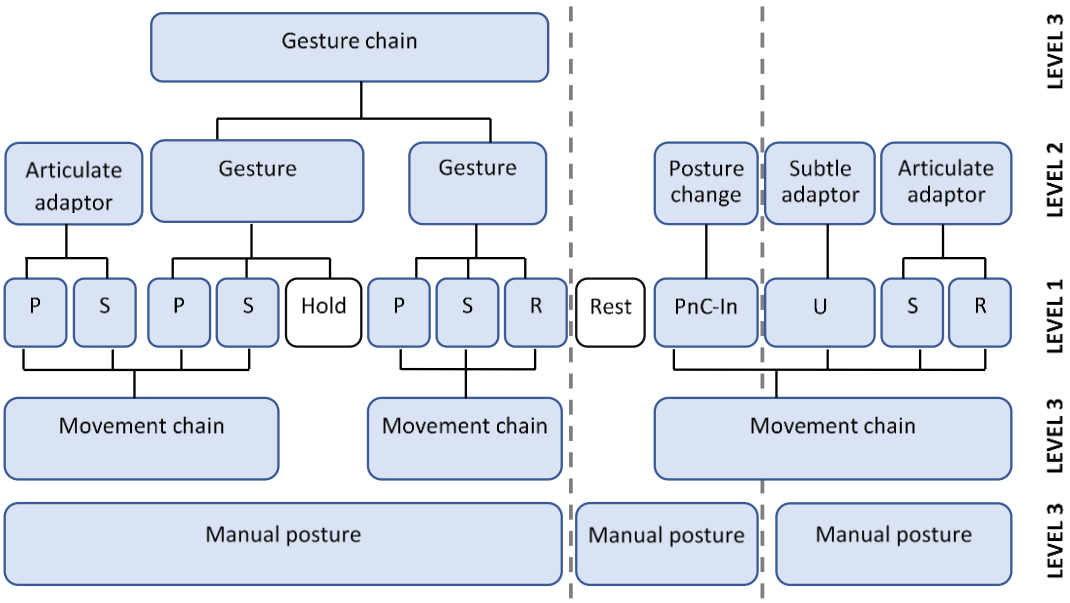

Figure 2. Manual units of different levels. Movement types: P – preparation; S – stroke, R – retraction, U – unstructured, PnC-In – independent position change.

In Figure 3 below, the system of ELAN tiers designed for manual movement annotation can be found; tiers are grouped into levels for illustrative purposes. Apart from these tiers, the system includes the mComments tier, assigned for annotators' comments.

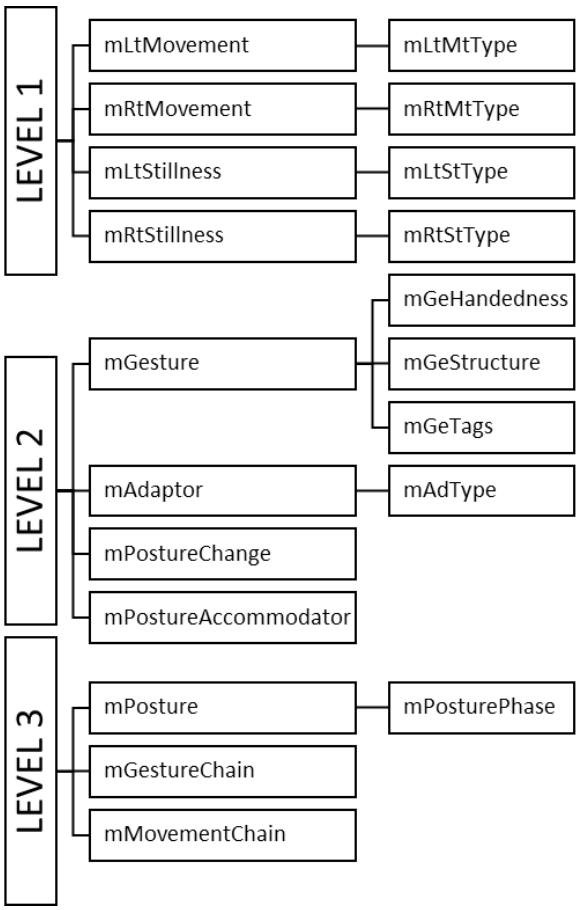

Figure 3. ELAN manual annotation tiers system.

In the following section, we will have a closer look at each of these levels, discussing the segmentation rules for the corresponding units and their characteristics.

3. Segmentation rules and unit characteristics

3.1. First level: movements and stillness intervals

3.1.1. Segmentation of the kinetic flow into movements and stillness intervals

At the first stage of the annotation, the whole flow of kinetic behavior is divided into separate movements between which there are periods of stillness when the hands and arms stop in a certain position, whether neutral or marked. This division is carried out based on formal (visual) grounds without turning to the speech content (the sound is off).

A separate manual movement is a fragment of kinetic flow united by a single effort, trajectory, direction, velocity pattern, hand shape, etc. Usually this is a unit that corresponds to a phase of a gesture or an adaptor.

The set of kinetic characteristics used for defining the boundaries of movements in the course of the annotation process is described below in Table 1.

Articulator (acting part of the upper limb)	Part of the limb with which the movement is performed: fingers, hand, forearm, arm
Hand shape	Hand shape & configuration: • flat hand (open palm), slightly bent hand, curved hand (fingers at an angle of 90 ° to the palm), hand clenched into a fist • thumb up, thumb lies over a fist, thumb lies along the knuckles, thumb is hidden in a fist • one/two/several fingers sticking out...
	Palm orientation: inside, outside, up, down, inside-up, inside-down...
Movement	Type: directional, object-painting, back and forth, rotation...
	Trajectory: straight, arch, circle, spiral, S-wave, figure-shaped, other...
	Direction: left, right, up, down, away from body, towards body, diagonally...
	Amplitude: large, medium, small (absolute and relative — in comparison with the typical one for the speaker)
	Velocity pattern: constant speed, acceleration, deceleration
Gesture space location	Gesture space zone: see diagram below (center-center, center-up, center-right, center-left, center-down, periphery-top...)
	Proximity to the body: on the body, close to the body, far from the body
Tension	Articulator tension: +/– (absolute and relative — in comparison with what is typical for the speaker)

Table 1. Kinetic characteristics of motion

The scheme in Figure 4 below (taken from David McNeill's book "Hand and Mind: What Gestures Reveal about Thought", 1992) shows how the gesture space is divided into zones, from the center to the extreme periphery.

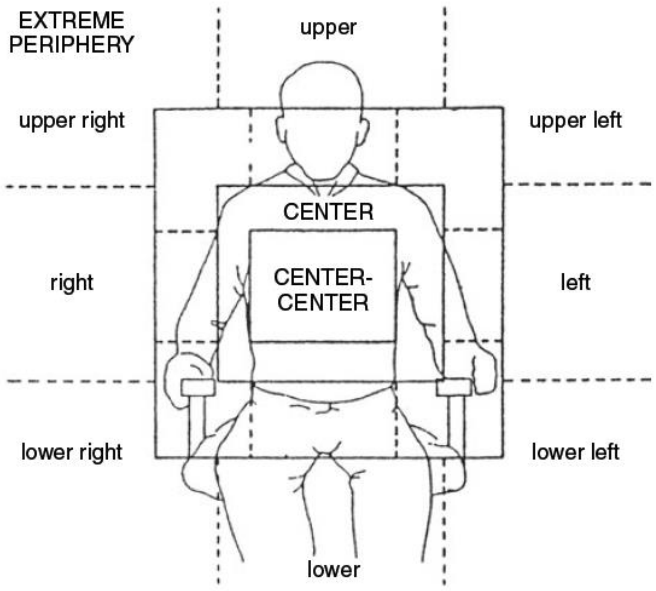

Figure 4. Scheme of the gesture space zones.

A point of a noticeable shift in the characteristics described above indicates a boundary between movements. They also help to separate gestures or adaptors in a sequence: the more characteristics change at the same time, and the more significant the change is, the higher the probability that this is not only a shift from one movement to another (or from one phase to another) but also from one gesture/adaptor to another. For the majority of speakers, a simultaneous change of two or three characteristics points to a boundary between the movements, while a change of four or more characteristics signifies a boundary between larger objects such as gestures or adaptors.

Movements can be significant (large) or insignificant (small). Movement is considered significant if at least one of the following conditions is fulfilled:

- A hand shifts in space for a distance exceeding half a hand
- The basic hand configuration changes (see Table 1)
- The position of the thumb changes (see Table 1)

When the hands are put/clenched together, or the fingers are intertwined (crossed), a change of the upper object is also considered significant (e.g., ‘left hand on top of the right’ changes to ‘right hand on top of the left’, the same for thumbs and index fingers).

Movements of a shorter range but still ‘exceeding the width of a finger’ are considered insignificant and comprise part of the U-type intervals (see below). Under certain conditions, they are allowed in Hold and Frozen stillness intervals (“drifting”, “tension shakes”).

The more expressed and contextually noticeable the kinetic characteristics of a certain movement (for example, “the hand is clenched into a fist; the thumb is sticking out; there is tension; there is acceleration; the hand is moving downwards; the amplitude is from the middle of the chest to the middle of the torso”), the more it is articulated (structured). If the characteristics are less expressed and bear little difference from the surrounding context, its articulation becomes weaker.

Two independent tiers, mLtMovement and mRtMovement, are used for the annotation of left and right hand movements, respectively. The movements are annotated separately and independently for each hand. Annotations in these tiers are numbered using the automatic function in ELAN: Tier > Label and Number Annotations (with the Prepend leading zero option). First comes the code of the speaker (N, C, or R), then the prefix (-mLtMt- or -mRtMt-) depending on the tier, and, finally, the number itself; for instance N-mLtMt001, C-mRtMt021, or R-mLtMt121.

Within the annotation scheme developed for this project, the following kinds of stillness (inaction) exist:

- Complete absence of any motion
- Micromovements ("less than a finger width") of any kind
- Passive movements triggered by other factors (shaking with laughter; tension shakes when a hand is being held in one position in the air for some time; shakes from the impact of or made by the other hand; hands *drifting* as a result of gravity; so-called *echoing*, when the inactive hand duplicates the movement of the active hand, thus producing subtle automatic movements)

For the segmentation and classification of stillness periods, it is highly important to define what a neutral position is.

A neutral position is a position which (almost) does not require any specific effort on the part of the speaker and to which they return having finished yet another fragment of communication.

As opposed to a neutral position, a marked position requires a noticeable muscular effort. A marked position can either be a part of a gesture or an adaptor, or it can form a so-called Frozen interval.

As a rule, each speaker has a customary set of neutral positions, and this set has a hierarchical structure.

The absolute neutral position (*rest* position) is a sign of a (temporary) withdrawal from active communication or refusal to join in (for example: a relaxed posture, hands are relaxed and resting on the knees, the gaze might not be directed at the speaker).

A middle-level neutral position is a hand position that can be considered neutral within a relatively continuous part of a conversation. For example, a speaker leaning forward to the interlocutor or sitting back in the chair; elbows or hands are resting on the knees or on the chair. In this level the hands might not stay in the neutral position for long or at all; speakers often return to such positions only to signify the termination of a communicatively significant movement. Crossed arms, folded hands with interlocked fingers and other similar positions are also very frequent middle-level neutral positions.

And finally, a local neutral position is a position to which the speaker returns within a short fragment of the conversation (as a rule, within a movement chain). A typical example of this position is the hands kept in the air in front of the speaker's chest with elbows pressed to the torso; the speaker returns the hands to this position between gestures. A local neutral position differs from a marked one by the higher degree of the hands' relaxation and by visibly less tension. Within the local context, such a neutral position serves as a reference point that separates gestures or gesture sequences from each other (see Figure 5).

Figure 5. A local neutral hand position.

Typical neutral positions for each speaker are established at the preliminary stage of the annotation procedure. The most important feature of a neutral position is the lack of tension or any visible muscular effort even if the hands do not rest on the knees/hips but remain in the air.

Here is a list of the most frequent neutral positions:

- Hands resting on a surface (knees, thighs, chair)
- Arms crossed at the chest
- Hands folded or hanging relaxed near the chest in the center of the gesture space, with the elbows down, usually propped against the sides of the torso or against the back of the chair

Thus, when dividing the flow into separate movements and periods of stillness, the annotator relies on the following parameters:

- Position (neutral vs. marked) in which the movement preceding stillness ended
- Lack of considerable differences between the initial and final positions of the potential interval of stillness
- Tension or lack thereof (for instance, even hands resting on one's knees might be viewed as being in a marked position if the hands are visibly tense)
- Structural connection (or the lack thereof) between the movements following and preceding the particular interval of stillness (for instance, a retraction after a stroke, or a stroke after preparation).

All stillness intervals should be homogeneous as far as their characteristics are concerned. Two different stillness intervals in a row are not allowed in the scheme; there must be some movement between them that had caused the change of the stillness type.

Two independent tiers, mLtStillness and mRtStillness, are used for annotating stillness periods for each hand separately and independently. In the mLtStillness and mRtStillness tiers, annotations are numbered using the built-in ELAN option, as in N-mLtSt001, C-mRtSt021, or R-mLtSt121 (where the first letter is the speaker's code, followed by the tag -mLtSt- or -RtSt-, and after that the number itself).

3.1.2. Structural types of movements

Even in the first stage of the annotation procedure, conducted on a purely kinetic basis before turning to the speech content, an annotator can determine the structural type of a separate movement which later will be interpreted functionally as a part of a larger kinetic unit. The annotation scheme allows for the following structural types of movements.

- Stroke (S) is a well-articulated (within the given context) significant movement which is contextually independent; the motion itself serves its own purpose, as opposed to being necessary for another motion to be executed. Further on, a stroke will be interpreted as a functional core of a gesture or of an adaptor.
- Preparation (P) is a significant directional movement which serves as a structural transition to the onset of a stroke from a neutral position or from the end of the previous movement.
- Retraction (R) is a significant directional movement which serves as a structural transition from the end of a stroke to a neutral position or to the onset of the next movement. Retraction can be full (ending in a neutral position) or unfinished (interrupted by the next motion, either preparation or a stroke). The core structural attribute of a retraction is its direction towards a neutral position. Retraction also often entails tension reduction (visible relaxation).
- Independent position change (PnC-In) is a simple independent significant movement from one neutral manual position to another. Sometimes independent position changes may be incomplete (not ending in a new neutral position); such failures are typically caused by interrupted preparations or retractions. In such cases, the following sequences of units may appear: [PnC-In – Frozen – P] or

[R – Frozen – PnC-In]. If two hands simultaneously change position, the resulting position change movements are considered independent.

- Dependent position change (PnC-Dp) is a compulsory change of manual position which takes place when the movement of one hand is impossible without the other hand changing its position as well. For instance, to lift a hand which is covered by the other hand, one has to lift the upper hand first. The criterion for dependency is the existence of a shared point in the trajectory where the hands touch (usually it is either the onset or the end of the movement). Dependent position changes appear along with a preparation or a retraction of a one-handed gesture/adaptor when one hand makes a gesture/adaptor, and the other is forced to shift. In some cases, a dependent position change appears when one hand 'allows' the other to pass because the inactive hand stays in the way of the active hand. Dependent position changes can also be unfinished (due to the same reasons as independent position changes).
- Unstructured movement (U) is a movement or a series of movements where the beginning and the ending of the interval have no significant differences, and no transitional point between them differs significantly from the beginning and the ending. Unstructured movements can be described in terms of weak articulation, low effort and low velocity; these parameters also remain relatively unchanged during the whole interval.
- Other is a label for the movements of a combined nature that cannot be classified as any of the above-mentioned types. As a rule, these movements either combine wide-ranged directional movement with subtle weakly articulated movements (e.g., lowering hands and rubbing them simultaneously) or form a long series of weakly articulated movements where subtle changes accumulate in such a way that the final position of the interval turns out to be considerably different from that of the beginning, although there is not a single substantial movement between them.

Similar movements are annotated the same way irrespectively of whether they are a part of a gesture or of an adaptor. For example, a directed preparatory lift of a hand in order to perform the main movement should be treated as preparation for all similar instances.

Apart from simple single-stroke gestures, there are also gestures with so called multi-strokes. Multi-stroke refers to a homogenous series of repeated movements, united by the same trajectory, constant effort and the same velocity pattern. Typical examples of multi-stroke gestures are clapping hands, twisting a finger at the temple, waving, etc. Movements which constitute a multi-stroke are not only kinetically but also functionally united; however, at this stage multi-strokes are annotated strictly on their structural basis (for a discussion of the most typical segmentation difficulties and possible solutions, see Section 3.1.4. below).

Movement type annotation is done in the dependent mLtMtType and mRtMtType tiers. For each annotation in the parent tiers mLtMovement and mRtMovement, a new annotation in the dependent tiers is added where the structural type is chosen from a closed list of values. This list can be found below.

- S (Stroke)
- P (Preparation)
- R (Retraction)
- PnC-In (Independent position change)
- PnC-Dp (Dependent position change)
- U (unstructured) — unstructured (subtle) movement
- Other — combined movement

At this stage of the annotation procedure, these types are assigned purely based on visual analysis. The annotator compares the kinetic characteristics of the juxtaposed movements,

estimating their roles in the kinetic context. The resulting segmentation and classification is verified during later stages using the speech content and information from other discourse channels.

3.1.3. Stillness interval types

Relying on the list of stillness characteristics discussed earlier, the annotator assigns one of the following types to each stillness interval.

- *Rest* is a relaxed stop in a neutral position.
- *Hold* is a stop in a marked position within the boundaries of a potential gesture/adaptor (i.e., after a preparation or a stroke).
- *Frozen* is either a stop in a marked position outside the gestures/adaptors (after a retraction or an unfinished position change), or a visibly tense stop in a neutral position.

From a functional point of view, a hold is a part (a phase) of a gesture or an adaptor, and as such, it carries part of the information load (see below the section on the phase structure of gestures). However, during the first stage of annotation, one should rely strictly on the structural characteristics of the movements, and that is why the choice between a hold and a frozen interval is done based on the preceding context with the following recommendations.

- Hands halted in a marked position after preparation or after a stroke are considered a hold interval, by default (all other factors held equal).
- Hands halted in a marked position after a retraction or after an unfinished position change are considered a frozen interval.

Stillness type annotation is carried out in the mLtStType and mRtStType dependent tiers by choosing one of the values from the closed list (Rest, Hold, Frozen).

A more detailed discussion of difficult cases involving assigning stillness types can be found in Section 3.1.4.

3.1.4. Segmenting the kinetic flow and assigning movements and stillness types: Resolving issues

Since at this stage the annotation is made without turning to the speech content, and the functional units (gestures, etc.) are not annotated yet, in ambiguous cases both the segmentation process and assigning the types of movements or stillness may cause difficulties. However, for a first approximation, such issues can be resolved even without turning to the speech content. In order to achieve this, a series of standard solutions has been developed; at the later stages of the annotation procedure, the accuracy of these solutions is checked and validated.

Multiple low-amplitude movements, especially of the adaptive kind (rubbing, scratching, etc.), are often difficult to categorize. When choosing between a multi-stroke (S) and an unstructured interval (U), one should rely on the amplitude, the equivalence of the elements, and on the constant velocity pattern; one should also take into account secondary movements before and after the problematic one.

- If the amplitude of the movement is approximately half a hand size or more, the movement is considered a stroke.
- If the amplitude is less than half a hand size, but the movement consists of identical repeated elements and/or has a rhythm, then it is also considered a stroke (one of the typical examples is tapping).
- If there is an obvious preparation before the movement or an apparent retraction after it, such a movement is considered a stroke. For this test, preparation is a significant movement from a neutral position to a marked one, and retraction is a significant movement from a marked position to a neutral one. Typical examples of such a stroke are tapping or scratching one’s nose.
- In the rest of the cases, any small movement is qualified as unstructured.

When gesticulating actively, a speaker can move from one stroke to another using only one transitional movement without reaching any neutral position between the strokes. In this

case, such a secondary transitional movement is interpreted as a preparation of the subsequent stroke and not as a retraction of the preceding one.

Some types of strokes may include integrated elements of preparation (in the beginning) or retraction (in the end). For example, in the beginning of a pointing gesture, the hand simultaneously begins the change of the configuration and the movement itself to the final point of the stroke. In many gestures of the 'knocking' or 'striking' type, where the stroke ends in a neutral position, the relaxation of the hand starts when it is still moving. As a result, there is no clear boundary between the end of the preparation and the onset of the stroke, or between the end of the stroke and the beginning of the retraction; the corresponding kinetic changes occur alongside the initial and final stages of the stroke trajectory. In such cases, preparation or retraction is not segmented as a separate movement, and the whole combined movement is tagged as a stroke (S). However, when later describing the structure of the gesture, this case is described as a combined stroke of the P-S or S-R type correspondingly.

The sequence [PnC-In – S] takes place only when either the position change movement ends in the neutral position or the stroke includes its own preparation (i.e., belongs to the P-S subtype). Otherwise, any transitional movement that moves the hand from a neutral position to a marked one before the stroke is considered preparation.

The sequence [S – PnC-In] takes place only when the stroke either ends in a neutral position or includes its own retraction (i.e., belongs to the S-R subtype). Otherwise, any transitional movement that moves the hand from a marked position to a neutral one after the stroke is considered a retraction.

Retraction can be partial: having begun from the ending of a stroke to the neutral position (for example, to the center of the gesture space), at some point the speaker overlaps it by transitioning to a new movement. In this case, the movement to the point where the trajectory changes is considered a retraction, and the following one a preparation or a stroke.

If a preparation or a retraction is interrupted with a stop in a marked position, then only the movement immediately adjacent to the stroke, the pre-stroke hold, or to the post-stroke hold is considered preparation or retraction accordingly. The premature start of the preparation (before the interruption) or the second part of the retraction (after the interruption) is annotated as an independent position change (PnC-In), and the stop in a marked position is annotated as a Frozen interval. Thus we get such sequences as [PnC-In – Frozen – P – S] or [S – R – Frozen – PnC-In]. The longest possible sequence of phases within one gesture is P H S H R.

When annotating multi-strokes at the stage of segmentation, the main criterion for considering a movement to be a multi-stroke (as opposed to a series of singular strokes) is the homogeneous kinetic character of the entire sequence: constant trajectory, amplitude, hand shape, and velocity throughout the duration of the potential stroke. If there is a preparation before a multi-stroke, it significantly differs from similar movements within the stroke. For example, in a 'hammer' gesture, the amplitude of the first lift of the hand is usually much higher than the amplitude of the intermediary lifts of the hand between the downward striking motions. If the first preparatory motion is not different from the rest, while the velocity and the amplitude of the raising and striking are homogeneous throughout the given interval, then such a movement is considered a multi-stroke without preparation. This happens quite often within gesture chains, especially with gesture repetitions.

Apart from multi-strokes, there is also a phenomenon called a rebound, which is a short bouncing of the hand, following the same trajectory as the main movement. Rebounds usually take place at the end of high-effort strokes and retractions. They are specifically tagged S Rebound and R Rebound (see below the corresponding section on the rules of annotation of gesture characteristics). A rebound can be multiple or singular. A stroke with a multiple rebound is also annotated as one united movement since a rebound is impossible on its own (without a preceding stroke) and it does not have its own semantics separate from the main stroke. A multiple rebound is tagged as Multi-Rebound.

Splitting a combination of two movements into P and S as opposed to P-S, or S and R as opposed to S-R, as well as the choice of structural interpretation (P S or S R?) in many cases depends on the semantics of the gesture, although there are some kinetic factors and technical solutions which contribute to creating a systematic and unambiguous annotation. Such solutions are listed below.

- In general, a smooth fragment united by the same effort, velocity pattern, and trajectory should not be split during the segmentation.
- In a combination of movements such as 'up and down' or 'down and up', the boundary is by default set at the point where the trajectory changes (at the top or at the bottom, correspondingly).
- If in a combination of movements of the 'up and down' type the second movement is made with a greater effort, this combination is interpreted as a 'striking' or 'knocking' gesture (the majority of rhythmical beat gestures also work this way). In such cases, the second movement (down) is considered a stroke and the preceding 'up' movement is preparation. In the opposite case, if it is the first movement (upwards) that is characterized by a bigger effort this first movement is tagged as S or P-S and is followed by a retraction. There are some beats directed upwards and slightly towards oneself that have this structure.
- If a movement of the 'up and down' or 'down and up' type is absolutely homogeneous in velocity and effort, then it is not divided into parts (similar to a multiple stroke). This type of movement is called "Shuttle" and is specifically tagged in the later stages.
- If between two movements with different trajectories there is no clear breaking point, but there is a fragment with a smooth change of trajectory and hand shape, then this transitional fragment belongs entirely to one of two movements and is not broken down into parts. As a rule, in this case the boundary is placed at the point of a noticeable change in velocity. For example, in the combination "raising the hand > turning the hand smoothly by 90 ° > horizontal palm strike", the velocity usually increases either at the onset of the turn or at the end of it, and the onset of the stroke should be determined by this reference point.
- If during a smooth change of trajectory the hand changes shape, then the onset of the stroke may be determined by the beginning of the change in the configuration of the hand, but splitting smooth homogeneous fragments on such grounds is to be avoided.
- In open palm movements (open palm towards the interlocutor), the first movement (upwards and/or away from oneself) is considered a stroke (S or P-S), followed by a retraction.
- An arc or a circle-type movement with homogenous velocity and effort is not divided into parts. If in the onset of the stroke with such trajectory there is a small preparatory fragment (for example, a change in the shape of the hand simultaneously with the onset of the movement), this movement is classified as P-S, but not split.

When characterizing stillness, the following types of difficulties may occur.

The choice between rest and hold in the presence of some weak tension. As a rule, rest is characterized not only by the neutrality of the position, but also by the relaxation of the hand, and the hold is primarily characterized by its marked position, but also by tension (although it is not necessary for a hold). In ambiguous cases, one should turn to the two components of the preliminary gesture portrait: a) how frequently the given position of the hand is used, and if the speaker returns to it; b) how relaxed the speaker's hands are in general when they do not move, both inside and outside the movement chains. Some speakers can keep their hands in an unusual position which is comfortable for them (for example, the hand rests on the knee, but only at the

fingertips) for a long time. During the active phase of communication, some speakers can experience constant slight tension for a considerable time, even when the hands are fully resting on their knees or on the chair. In such cases, the typical position should be considered neutral, and some small amount of persistent tension has to be dismissed as insignificant.

The choice between hold and frozen states in the absence of transitional movements. The most ambivalent is the case when there are no preparations and retractions in the sequence, and the stop in a marked position occurs immediately after a stroke (less often before it). In these cases, the stillness interval is, by default, considered to be structurally connected with the previous movement and is annotated as a hold after a stroke or a preparation, and as frozen in other cases.

The sample annotation in ELAN obtained at the end of the first stage is shown in Figure 6.

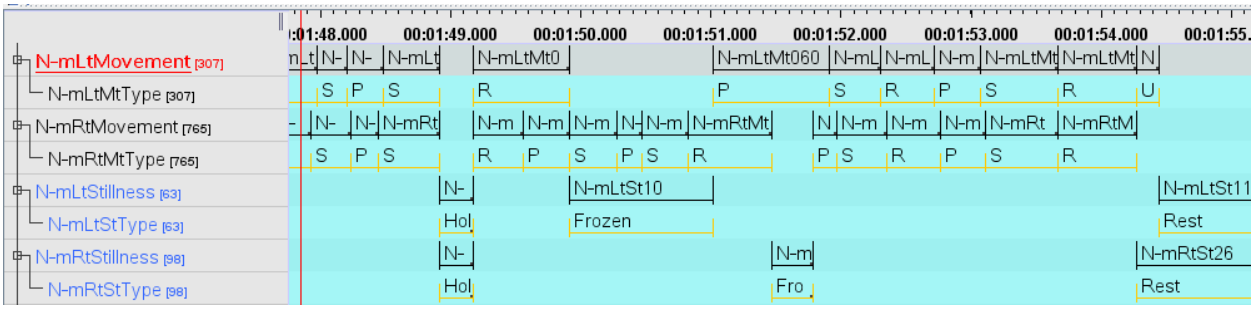

Figure 6. Sample annotation of movements and stillness in ELAN.

Further on, the results of the first stage of annotation will be verified at the second, or functional, stage.

3.2. Second level: Gestures, adaptors, posture changes and accommodators

3.2.1. General description of the second level units

Kinetic flow segmentation into gestures, adaptors, posture changes and posture accommodators is performed relying on speech context (with the sound on), but based on the same kinetic principles as the segmentation into movements. At the boundaries of the second level units, more parameters change than at the boundaries of their constituent movements.

The structural units that have been allocated in the previous stage are now reinterpreted as functional components of the bigger units: the phases of gestures or adaptors, as well as posture changes and accommodators. Strokes become semantic and kinetic cores of the gestures or adaptors, and preparations, retractions, and hold intervals become their secondary phases. Position change movements are interpreted as posture changes and posture adaptors.

Articulate adaptors (Type 1) consist of the same phases as gestures. The stroke of an adaptor is the central component where the purpose of the adaptor is realized (scratching the nose, rubbing the knees, etc.).

Subtle, weakly articulated adaptors (Type 2) consist of U-type intervals.

Position change movements at this stage are reinterpreted as components of either a posture change or a posture accommodator.

Below we shall discuss the rules for annotating all these units in greater detail.

3.2.2. Gesture segmentation

For each gesture, it is important to determine which part of the upper limb is involved in it: whether it is the fingers only, the entire hand, the forearm or the whole arm. As a rule, the bigger the articulator, the less important the change in some parameters for smaller parts. For example, for most forearm gestures, the change of the hand shape is important, but the exact position of one finger is not.

While comparing various parameters of adjacent movements, it is necessary to determine which part of the hand these parameters do characterize each time, and therefore which of them are essential and which are not essential for this particular gesture.

At this stage, the boundaries of movements, initially identified without reference to the speech context, can be adjusted, along with their functions. However, one should make sure that the boundaries of speech and manual units remained independent. In order to achieve this, any information from other channels (primarily speech context, as well as oculomotor behavior, head movements, etc.) is taken into account only in cases where manual kinetic behavior allows two (or more) different equally applicable interpretations.

In case of a mismatch of the left- and right-hand phase boundaries (that is, desynchronization), gesture boundaries are determined in accordance with the special rules (see Section 3.3.2 below). The general principle is that in most cases the onset of a gesture is considered to be the beginning of its first phase, unless it reduces the stroke of the previous gesture in favor of the secondary phases of the next one.

Depending on the multiplicity of the stroke (the main movement of the gesture), gestures are classified as ordinary (with a single stroke), with a multiple stroke or with a multiple rebound.

A gesture with a multi-stroke is a gesture where the stroke is a series of rhythmic repetitive movements. Typical examples of a gesture with a multi-stroke are wagging a finger, twisting one's finger at the temple, clapping hands, etc. It is important to note that in cases like these, the whole series of movements is considered to be one stroke, since, if taken separately, the parts of a prototypical multi-stroke gesture lose the meaning that is inherent to the whole series. In addition to single and multiple strokes, there are also gestures with multiple rebounds. A multi-rebound differs from a multi-stroke by a substantial difference in the amplitude between the first stroke and its repetitions. The basic principles of segmentation of multiple strokes and rebounds are described above in Sections 3.1.2 and 3.1.4.

With a low velocity of gesticulation, one may experience difficulty in distinguishing a multi-stroke gesture from a series of repeated gestures. In such cases, an annotator must rely on both kinetic and semantic parameters.

In a multi-stroke gesture, at the stroke part boundaries only trajectory changes, and effort, amplitude and velocity stay the same. Hand shape and finger configuration may vary slightly, but not significantly. The amplitude of the parts in a multi-stroke is usually smaller, and the velocity is usually higher than that of other movements in the same context: the characteristic change in tempo occurs during the transition to the entire sequence of rhythmically repeated movements in its entirety, and then upon its completion. A multiple stroke possesses not only a fairly homogeneous internal structure, but also uniform semantics (as opposed to a repeated gesture).

Typical meanings conveyed by multi-stroke gestures

- An imitation of complex actions consisting of several identical movements: clapping hands, pedaling, wagging a finger, jumping up and down, etc.
- The multiplicity, intensity or prolongation of some component of the situation being described: "the boys continue walking in the direction of the tree" means describing a continuous action; "the gardener is picking pears" means repeating the same action many times.
- An emotional emphasis associated with the described event: the speaker's surprise and other emotions.
- The vagueness or inaccuracy of a chosen nomination without obvious hesitation in speech. The speaker does not exactly struggle to find a word but lets the interlocutor know that the chosen nomination is not perfect for the intended meaning.

The last two meanings sometimes combine, and a multi-stroke appears as a result of the fast homogeneous multiplication of an appellative gesture ("See what I am talking about!").

A series of repeated gestures is usually differentiated from a multi-stroke by the following signs: microscopic holds at the boundaries of the separate parts of the series, erratic velocity changes (deceleration between repetitions, and then acceleration), erratic amplitude changes (increase after a decrease), etc. In a multi-stroke, especially if a gesture lasts for some time, the hand may also slightly change shape, or the onset point of each individual repetition may slightly shift towards the neutral position. However, all these differences are visually estimated as insignificant and smooth, unlike a series of repeated strokes where such irregularities allow an annotator to annotate parts of a series precisely as separate gestures.

Typical contexts characterized by gesture repetitions

- Hesitation or the search for an accurate expression in the online mode (e.g., "[a bicycle] with that frame... the men's one... what's the name...": each component is accompanied by a repetition of a gesture depicting the diamond frame of a bicycle).
- Stroke-only repetitions with the shift of the onset point each time, thus advancing the trajectory. Such repetitions usually accompany verbal expressions like "went further", "goes down the stairs", "climbs up the tree", and so on. Such series can be oriented both horizontally and vertically.
- Repetitions of identical or close in form appellative gestures (open palm gestures and the like), that are usually associated with difficulties of discourse planning. In such cases, a gesture of this type can accompany every speech unit produced.

The multiplicity of the stroke is determined taking into account the articulator that is used for the gesture and its main movement. If a large stroke is accompanied by secondary smaller motions, that does not necessarily make it a multi-stroke. During the gesture in Figure 7, for example, the right hand moves from right to left, wriggling its fingers at the same time. As the whole hand is the main articulator and the fingers only produce additional movement, this is a single stroke, and not a multi-stroke.

Figure 7. The movement of the hand, accompanied by the movement of the fingers.

3.2.3. Desynchronization of hand movements: Overlaps

The boundaries of gestures are annotated in the mGesture tier (the same for both hands), and the gesture characteristics are indicated in the corresponding dependent tiers. In the main mGesture tier, the annotations are numbered using the built-in ELAN option, such as N-mGe001, C-mGe021, R-mGe121 (the first letter is the speaker's code, then comes the tag -mGe-, and after it the number itself).

By default, a gesture starts at the beginning of its first phase and ends at the end of the last phase.

However, the movements of the left and right hands are not always completely synchronous, which leads to overlaps of gestures or of their phases. In case of desynchronization, gesture boundaries are determined according to the following rules.

1. If the same gesture is started with one hand earlier than with the other, then by default, the earlier boundary of the two possible ones is chosen as the left boundary of the gesture.
2. If the post-stroke phases (hold or retraction) of the gesture n overlap with the initial phases of gesture n+1, then the left boundary of gesture n+1 is placed where its first phase begins, except for the case of a preparation or a pre-stroke hold overlapping with the stroke of the gesture n. In the latter case, the gesture n+1 starts at the end of the stroke n.
3. If two strokes overlap, the first gesture ends when the stroke of the second gesture begins.
4. If the left and right hand retractions of the last gesture in a chain do not end at the same time, the later (more to the right) end-boundary of the two possible serves as the end of the gesture.

Thus, in case of an overlap, the left (earlier) boundary is always taken as the start of the next unit, except for the case when the secondary phases overlap with the stroke of the previous gesture. In Figure 8 below, overlap examples can be seen: the next gesture preparation overlaps with the previous gesture retraction (in this case, gestures N-mGe102, N-mGe103, N-mGe104).

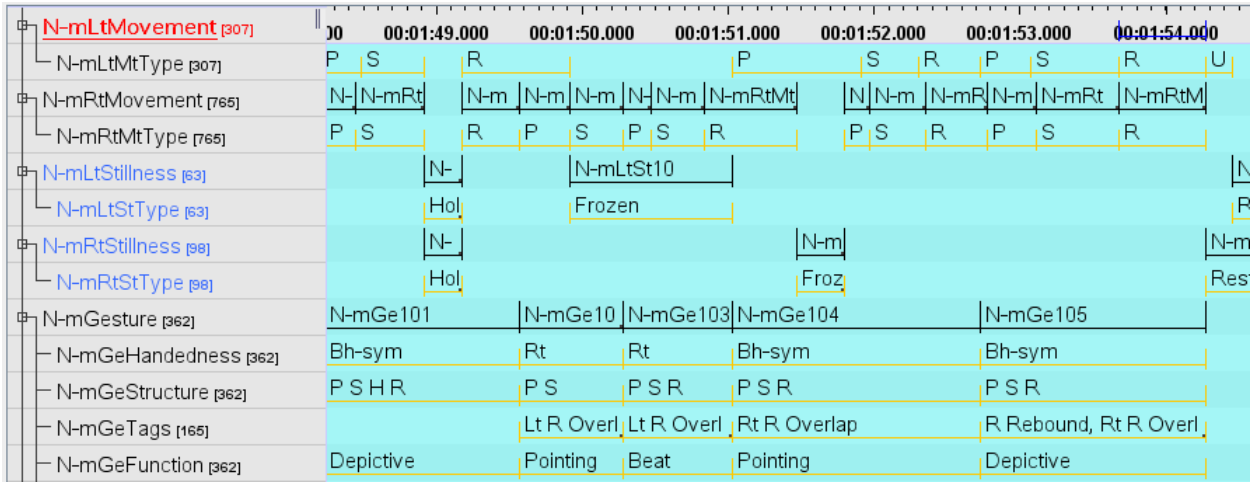

Figure 8. Desynchronization of hand movements in the recording pears22N.

While annotating adaptors, posture changes and posture accommodators, the same rules apply in any case of hand desynchronization.

3.2.4. Gesture characteristics

Handedness (mGeHandedness tier)

One of the main characteristics of a gesture is its handedness. Gestures are divided into one-handed (made by right or left hand) and two-handed (made by two hands) ones. In the latter case, the hands can move a) symmetrically about one of the body axes or planes (most often it is the sagittal plane, but other axes, planes and points of symmetry are also possible, especially for circular and rotative kind of gestures, such as imitating bike pedaling); b) together in the same way (in parallel or as one object, such as imitating the movement of windshield wipers or throwing a ball with two hands); or c) in two completely different ways (such as showing the Earth rotating around the sun or the process of writing on a palm with an imaginary pen).

This parameter is annotated in the mGeHandedness dependent tier by selecting one value from the predetermined list.

- Left handed gesture = Lt
- Right handed gesture = Rt
- Two-handed gesture (Bh):
  - Symmetric movement = Bh-sym
  - Identical movement = Bh-id
  - Different movements = Bh-dif

The types of two-handed gestures are described in more detail below.

Bh-sym is a gesture symmetrical about one of the points, axes or planes in the gesture space. Typical examples of symmetrical gestures are spreading hands, clapping hands, etc. The following gestures are also considered symmetrical: rotating hands around a point or around each other; gestures where one hand goes down and the other simultaneously rises up. Hand movements in a symmetrical gesture are mirror images of each other.

Bh-id is a gesture where the hands perform the same movement, move in parallel (for example, up and down or left and right) or move together as one object with a common trajectory (for example, holding an imaginary ball). One should bear in mind that all identical gestures where the hands move in parallel are also always technically symmetrical about the plane between the hands. Thus, the identity of a gesture makes a stronger statement, while the usual symmetry makes a weaker one. The majority of two-handed gestures are symmetrical, but identical gestures are much rarer.

Bh-dif is a gesture where the hands perform completely different movements or take a completely different shape. For example, one hand knocks on the palm of the other, one hand rotates around the other, etc.

The usual phase desynchronization, a slight difference in amplitude, tension or hand shape is not enough for a gesture to be classified as Bh-dif. Most people move their hands not always synchronously and not quite symmetrically, and often the dominant hand is tenser, takes a more rigid shape or moves with a larger amplitude. The degree of deviation of a specific gesture implementation from an “ideal” performance (i.e., from the two hands’ movements being absolutely identical) may vary. Therefore, the main criterion for defining a gesture as Bh-dif is the lack of a common trajectory and hand shape. At the level of interpretation, such a distinction can be understood as “in this gesture, hands do functionally different things”, or “the hands move in completely different ways”, and not just “the movements of the hands are not quite the same”. Desynchronization of the left and right hand movements does not affect this parameter (and is displayed only through independent annotation of the gesture phases in the corresponding tiers). Nor do small differences in hand shape and finger position mean that the gesture should be categorized as Bh-dif (see Figure 9).

Figure 9. Identical gesture (Bh-id) with a slightly different hand shape.

Another type of Bh-dif is the simultaneous coordinated execution of two gestures with different functions. For example, there are cases when one hand performs a descriptive gesture and the other performs a beat or a pragmatic gesture (for more details on gesture functions, see the corresponding subsection below). If two such gestures are executed at the same time and their phases completely or almost completely coincide, it is considered a special subtype of a two-handed gesture within the Bh-dif type. The main criterion for distinguishing such types of two-handedness from simple overlapping / desynchronization is the complete or almost complete simultaneity of the strokes.

For complex cases of heavy hand desynchronization causing doubts that there is not a single two-handed gesture, but two overlapping one-handed gestures, there is a special temporary handedness label Other. However, this label is temporary and is removed at the end of the annotation procedure.

Gesture phase structure (mGeStructure tier)

For each gesture, its phase structure is identified and annotated in the form of a phase code sequence. As has been mentioned above, a gesture may include the following phases: preparation (P), stroke (S), hold (H), and retraction (R).

In some two-handed gestures, the static phase (hold) might be comparable in importance to the dynamic phase (stroke); this happens when the active part of the gesture is performed only with one hand, and the other hand plays the role of a stationary object. For example, the speaker knocks with one hand on the motionless other hand; one hand represents an object, and the other moves to it, etc. In such cases, due to the additional functional load the hold phase is marked as a hold-stroke (Hs) in the gesture structure formula.

Several examples of the most typical phase combinations are listed below.

- **P S R**: preparation, stroke, retraction (the gesture begins and ends in a neutral position).
- **S R**: stroke, retraction (e.g., if after the previous gesture the hands have already moved to the necessary position, and preparation is not required).
- **P H S**: preparation, hold, stroke (and immediately after the stroke, there is a preparation for the next gesture, which is not included in this one).
- **P S Hs R**: preparation, stroke, hold-stroke, retraction.
- **S H**: stroke, hold (in some cases, a preparation is used only in the first gesture of the chain, and a retraction is used only in the last one).

The given list is not exhaustive, as it represents only the most frequent cases. Phases are listed in order of their appearance (beginning). If in a two-hand gesture two different phases begin at the same time (usually S and Hs), then a stroke is indicated first in the list.

If a gesture includes a combined movement, which is a combination of a preparation with a stroke or a stroke with a retraction, such combined phases are hyphenated in the gesture structure formula: P-S, S-R, P-S-R. The P-S-R combination might be found in the “up and down” gestures where both parts have equal effort and it is impossible to determine what part of the whole is a stroke.

If a gesture is truncated (i.e., its last phase is obviously not completed, which is a rare phenomenon but still a possible one), the truncated phase is marked with an equal sign in the gesture structure, for example: P S= or P=.

Tags (mGeTags tier)

The tags tier is used for indicating typical but not obligatory gesture characteristics, and also for indicating some frequent types of inter-gesture relations (repetitions, overlaps), if applicable. In an annotation in this tier, the applicable tags are entered manually and may be combined with each other; however, within the proposed scheme the list of tags available is limited (see below).

- **Shuttle**: a two-part gesture (“back and forth”, a P-S-R without any possibility to determine which part is a stroke).

- **Multi-S**: a gesture with a multi-stroke.
- **S Rebound**, **R Rebound**: a rebound at the end of the stroke or the retraction.
- **Multi-Rebound**: a gesture with a multiple rebound at the end of the stroke.
- **Long R**: a long retraction (consisting of two or more parts, without reaching a neutral position until the end of the last part).
- **Lt/Rt P/S/H/R Overlap**: an overlap of another gesture phase with the current gesture. The tag is used at the start of the overlap; if the overlap affects several gestures in a row, each of them is tagged. If two phases of the same adjacent gesture overlap with the current one, they are indicated within one tag, e.g. Lt S H Overlap. If two phases of different gestures (a preceding and a succeeding one) overlap with the current gesture, they are listed separately in chronological order, such as Lt R Overlap or Lt P Overlap.
- **Repeat**: a repeat of the previous gesture.
- **GeBreakOff**: truncation of a mostly accomplished gesture. This tag is used if there is a stroke in the gesture, but the stroke phase is not completed due to a premature transition to another gesture or another failure.
- **GeFalstart**: truncation of a gesture in which it is not possible to recognize a stroke phase (for example, the one that consists only of preparation).

It is preferable first to list the characteristics of the gesture itself (e.g., Shuttle, Multi-S), and then to use tags that describe the relationship of the current gesture with adjacent ones (e.g., Overlap, Repeat) in chronological order (the left context is described first, then the right context).

If necessary, the list of tags can be expanded depending on the research objectives.

Gesture functions (mGeFunction tier)

In the proposed annotation scheme, for each gesture its functions are annotated. The list of functions and rules can be found below.

- Depictive
- Pointing
- Beat
- Other
- Pragmatic (+/–), or a pragmatic / metaphorical gesture

The last function is not used separately but can be combined with any of the functions listed above.

For the process of annotation, it should be taken into account that functionally more complex gestures may include simpler functions as their elements. For example, a gesture depicting a character's movement may additionally include a pointing to the character itself, as well as a rhythmic component (beat), which allows one to synchronize the peak of the stroke with the corresponding speech part. To avoid an excess of tags, only a more complex function is indicated in such cases (for example, "Depictive" instead of "Depictive, Pointing"; "Pointing" instead of "Pointing, Beat"). "Pragmatic" can be combined with any other tag; in the case of a purely pragmatic gesture, it is tagged as Other, Pragmatic.

Figure 10 below shows a sample of annotated gestures and their characteristics.

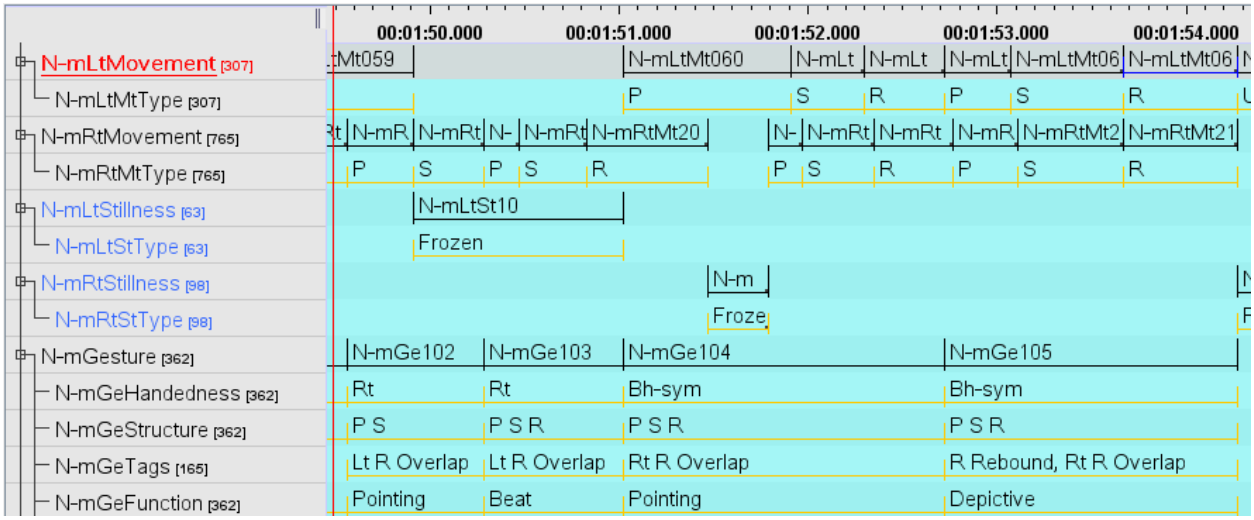

Figure 10. Sample annotation of gestures and their characteristics.

3.2.5. Adaptors and their types

Adaptors are annotated according to the same principles and rules as gestures, by comparing kinetic characteristics and functions and relying on the speech context.

In our corpus, adaptors of two main and two combined types are annotated.

Articulate adaptors (Type 1) are large, well-structured (articulated) movements. They tend to have a clear practical purpose which is not connected with speech or its perception: scratching, adjusting clothes, etc. At the level of movements, adaptors of this type consist of the same phases as gestures: preparation, stroke, retraction. Adaptors in general are much less likely to contain holds, but some asymmetrical two-handed adaptors tend to do so (for example, the speaker raises his hand and scratches it with the other hand while one arm is active and the other is in in a hold, as in two-handed Bh-dif gestures that contain the Hs phase).

Subtle adaptors (Type 2) are small, weakly articulated movements, sometimes with no internal structure, the practical purpose of which is hard to determine: rubbing hands, wiggling fingers, etc. At the level of movements, adaptors of this type consist of unstructured intervals (U). The initial and final position of the hands in such adaptors is by definition the same (there must not be a significant difference).

The kinetic properties of prototypical articulate adaptors (Type 1) are close to the kinetic properties of gestures:

- obvious effort;
- clear trajectory;
- high amplitude;
- noticeable acceleration / deceleration; and
- well-marked regular rhythm (for multiple movements).

Kinetic properties of prototypical subtle adaptors (Type 2) include:

- weak effort;
- unclear trajectory;
- low amplitude;
- lack of acceleration / deceleration; and
- lack of regular rhythm.

The two combined types of adaptors are more complex versions of the main ones.

Combined adaptors (Type 1+2) are essentially Type 1 adaptors complicated by additional weakly-articulated movements (for example, raising or lowering hands, accompanied by rubbing them).

Accumulating adaptors (Type 2+1) are Type 2 adaptors in which the gradual accumulation of changes as a result of multiple small movements leads to a significant difference between the initial and final positions of the adaptor. That means that an accumulating adaptor causes a

significant change in hand position but does not contain any significant-sized movements. Instead, accumulating adaptors correspond to the combined Other type of movements.

From a functional point of view, articulate adaptors tend to have a practical purpose related to the environment or physical discomfort (fixing clothes, scratching, stretching numb arms, etc.), and subtle adaptors have a purpose associated with emotional discomfort (fidgeting because of the necessity of sitting in one place for a long time, etc.). These goals can often be combined; thus, these functions serve as a basis for distinguishing adaptors from gestures, but when classifying adaptors into types, one should rely on kinetic properties rather than on functions.

The boundaries of the adaptors are marked in the mAdaptor tier according to the principles identical to the rules of annotating gestures (in the case of hand desynchronization, etc.). Annotations in this tier are numbered the same way as gestures, using the code -mAd-, for example C-mAd001, etc. The adaptor type is annotated in the mAdType dependent tier. Figure 11 below shows a sample annotation of adaptors and their types.

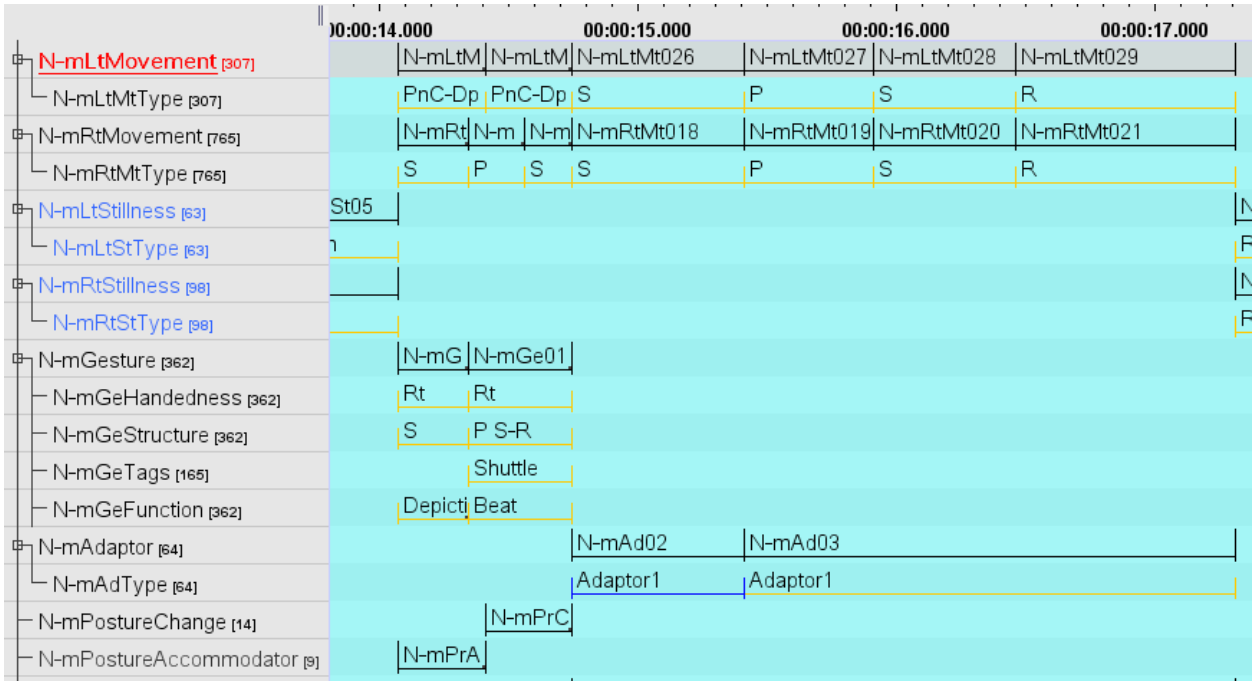

Figure 11. Adaptor annotation sample.

3.2.6. Posture changes and posture accommodators

At this level, contextually independent and dependent position change movements are interpreted as one of two functional phenomena: a posture change or a posture accommodator.

Posture changes are position change movements that result in taking a new (significantly different from the previous) neutral position for at least one hand (see Section 3.1.1 for significant and non-significant movements; to determine if the differences between positions are significant, the same criteria are applied). A manual posture change example can be seen in Figure 12.

Figure 12. Changing the manual posture (neutral position 1 at the left, neutral position 2 at the right).

Posture accommodators are position change movements that do not result in a new neutral position. That may happen because a hand returns to the same neutral position as before or because a position change movement is interrupted or not finished.

Both independent and dependent position changes can form either a posture change or a posture accommodator.

Posture changes and accommodators differ from gestures by their lack of communicative functions, and from adaptors by their lack of any practical purpose or an obvious connection with physical or emotional discomfort.

Posture changes are annotated in the mPostureChange independent tier, and the accommodators are annotated in the mPostureAccommodator tier using the same general rules as for gestures or adaptors.

Figure 13 below shows a sample of an accommodator and a posture change in a gesture context. As the right hand is performing a gesture, the left hand slightly changes its position at first, making a dependent position change at the level of movements and a posture accommodator at the functional level; this movement is not completed, as the left hand does not reach a neutral position. Then, alongside the second right-hand gesture, the left hand makes the second dependent position change movement, which at the functional level leads to a posture change.

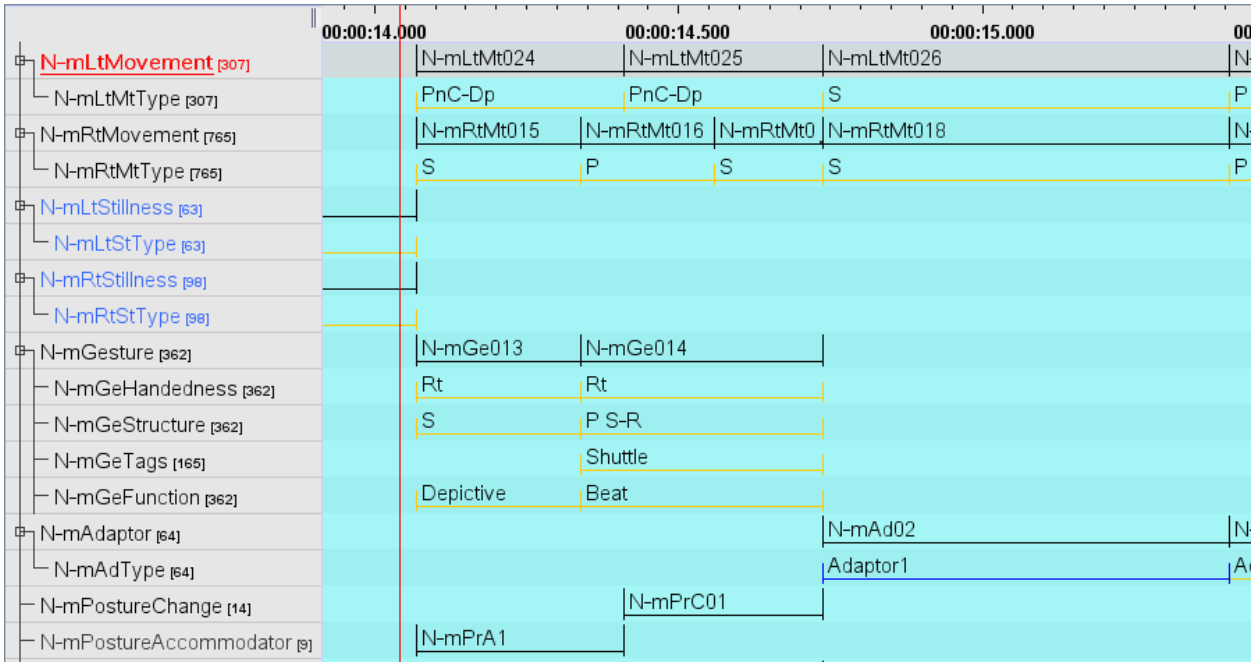

Figure 13. Posture accommodator and posture change annotation sample.

3.3. Third level: Manual postures and gesture chains

3.3.1. Manual postures and transitions

The segmentation of the kinetic behavior into manual postures is done at the last stage of the annotation procedure, as it relies on the finished annotation of individual movements, periods of stillness, gestures, adaptors, posture changes and posture accommodators.

A manual posture is an interval during which a speaker maintains the same neutral position, returning to it after having performed a gesture or an adaptor. Thus, the entire flow of kinetic behavior can be divided into manual postures without remainder.

Within one manual posture, there can be periods of stillness as well as a series of movements (gestures or adaptors) after which the hands return to the same neutral position as before. It is important that within one manual posture beginning from a neutral position A, no other neutral position, other than A, appears.

With each new neutral position, non-identical to the previous one, the current manual posture ends and the new one starts.

The interval between the last point where the neutral position A was taken and the first point where the neutral position B appears (i.e., a new posture begins) is called the transition phase. The transition phase can be continuous, consisting of a long movement; no neutral position is reached in this interval.

In order for neutral positions A and B to be recognized as significantly different, the differences between them must be significant (see Section 3.1.1. on significant movements; the same parameters are used to estimate the difference between neutral positions) For example, as a result of the transition phase, a hand lying on the knees should shift for at least half a hand distance; if during the transition the hand shape changes, this change must be significant (e.g., palm up > palm down, open fingers > fist; thumb over fist > thumb hidden in a fist). Insignificant changes are considered variations of the same posture.

Sometimes a new neutral position is taken by one hand only, while the other continues to gesticulate. In such cases, the posture boundary is also set at the point where the new neutral position has been reached, and the next posture will continue until either the motionless hand changes its posture again, or the second hand also takes a neutral position.

Thus, the identification of the manual posture boundaries and the transition phase boundaries is determined by the following two rules.

1. A posture boundary is set at the point where a new neutral position has been reached for the first time by at least one hand.
2. The transition phase begins at the last point where both hands took the initial neutral position and continues until the posture end.

As a rule, most manual postures have a constituting part that might be either a stillness period or a subtle (Type 2) adaptor that does not include preparations or retractions, consisting only of insignificant movements (U). However, sometimes a neutral position might be reached at the gesture boundaries without staying in it. In such cases, the point where a new neutral position is reached sets the boundary of the new posture.

In this case, as in the example described above, when one hand comes to a neutral position and the other does not, a posture consisting entirely of the transition phase is rendered acceptable.

Speakers who are inclined to move their hands from place to place several times before reaching a rest position can have several short postures of this kind in a row. Such postures would consist entirely of transitions and are formed by a single posture change, an adaptor or a gesture.

Manual postures are annotated in the independent mPosture tier and are numbered according to the general principles: N-mPr001, etc. All video material is to be divided into manual postures without remainder.

Transitions are marked in the mPrPhase dependent tier. They are not numbered and simply receive a standard Transition label.

The boundaries of any posture, as well as the boundaries of transitions, must coincide with one of the annotated kinetic objects in the tiers of movements.

Most often, a boundary between postures is placed at the following typical points: a) the end of position change movement that serves as a posture change at a functional level; b) the end of the retraction of a gesture or an adaptor. Less often, the boundary might be placed at the end of a stroke, if the stroke ends in a neutral position.

The beginning of a transition phase is usually the beginning of a gesture or an adaptor (Types 1, 1+2 or 2+1), after which the hands will no longer return to their original neutral position.

Sometimes when the hand movements are desynchronized, a new neutral position or positions is not reached simultaneously with both hands. In this case, between the points of reaching the neutral position with one hand and with the other, short postures appear that consist entirely of a transition phase. See Figure 14 below for postures mPr043, mPr044 and mPr045.

At the end of the Ge339 gesture, the right hand first comes to a neutral position, completing the mPr043 posture and starting the mPr044 posture. A little later, the left hand reaches a neutral position, completing the mPr044 posture and starting the mPr045 posture. Thus, the mPr044 posture consists entirely of a transition phase.

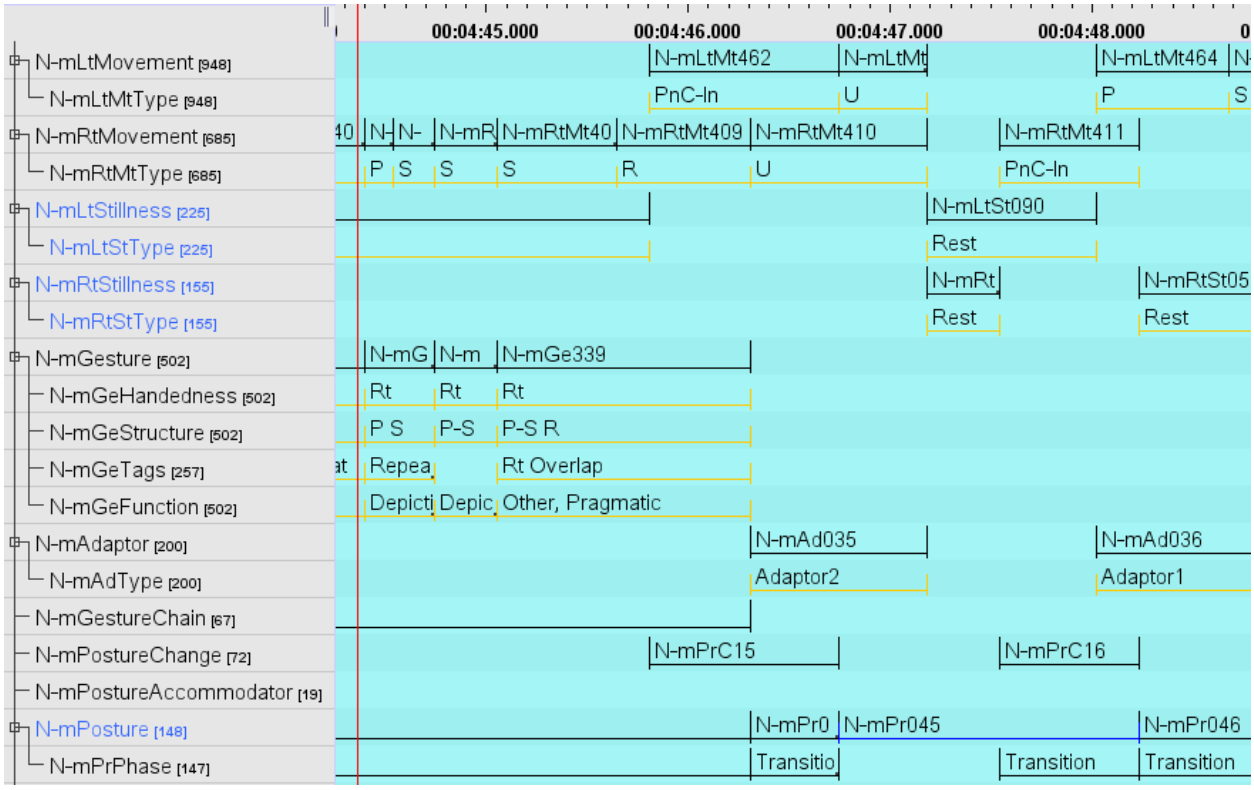

Figure 14. Manual postures annotation sample.

3.3.2. Gesture and movement chains

Gesture chains are uninterrupted sequences of gestures (including hold phases), and movement chains are uninterrupted series of movements.

The corresponding tiers (mGestureChain and mMovementChain) are the last to work through, as these annotations are produced semi-automatically, using the built-in ELAN tools (Merging tiers, Create Annotations From Gaps and other options). In order to put together these tiers, the full manual annotation of the movements and gesture tiers is needed. The chain tiers annotations are numbered in the standard way: N-mGeCn001, C-mMtCn021, etc.

3.3.3. Comments to the annotation

Annotators' comments are left in a separate independent mComments tier. This is necessary in all cases when arising difficulties significantly affect the annotator's decisions (primarily the decisions concerning phase and gesture boundaries). In the mComments tier, an annotation is created with the same boundaries as the commented object or the series of objects. The comments describe possible annotation options for a difficult fragment, indicating additional factors that are not included in the manual annotation directly but affect manual movements (e.g., laughter, other channels influence, drift or echo during stillness intervals, etc.).

4. Appendixes

4.1. General list of annotation tiers in ELAN

MANUAL ANNOTATION (hand gestures)				
Tier name	**Stereotype**	**Denoted phenomenon**	**Possible values**	**Comments**
First level of segmentation				
***-mLt/RtMovement**	None	Separate movement of the left/right hand	Unit code number	Separate tiers for the left and right hands
***-mLt/RtMtType**	Symbolic Association	Structural type of the right / left hand movement: preparation, stroke, retraction, etc.	One value from the closed list	Mandatory field. Separate tiers for the left and right hand
***-mLt/RtStillness**	None	Stillness interval of the left / right hand	Unit code number	Separate tiers for the left and right hands
***-mLt/RtStType**	Symbolic Association	Type of the left / right hand stillness interval: rest, hold, etc.	One value from the closed list	Mandatory field. Separate tiers for the left and right hand
Second level of segmentation				
***-mGesture**	None	Manual gesture	Unit code number	
***-mGeHandedness**	Symbolic Association	Gesture handedness	One value from the closed list	Mandatory field
***-mGeStructure**	Symbolic Association	Gesture phase structure	Formula composed of the predefined elements	Mandatory field
***-mGeFunction**	Symbolic Association	Gesture function	One or more values from the closed list	Mandatory field
***-mGeTags**	Symbolic Association	Additional characteristics of the gesture	One or more values from the closed list	
***-mAdaptor**	None	Manual adaptor	Unit code number	
***-mAdType**	Symbolic Association	Manual adaptor type	One value from the closed list	Mandatory field
***-mPostureChange**	None	Manual posture change	Unit code number	
***-mPostureAccomodator**	None	Manual posture accommodator	Unit code number	

Third level of segmentation				
***-mPosture**	None	Manual posture	Unit code number	
***-mPrPhase**	Included In	Transition phase from one manual posture to another	The only predefined value	
***-mGestureChain**	None	Manual gesture chain	Unit code number	
***-mMovementChain**	None	Manual movement chain	Unit code number	
Supplementary tier				
***-mComments**	None	Comments in free form	Free text	

4.2. Values in non-empty intervals

MANUAL ANNOTATION (hand gestures)			
Tier name and denoted phenomenon	**Value**	**Meaning**	**Comments**
First level of segmentation			
***-mLt/RtMovement** *Separate movement of the left/right hand*	***-mLtMt001** etc. ***-mRtMt001** etc.	Movement code number	Code number indicates which hand is moving
***-mLt/RtMtType** *Structural type of the right / left hand movement*	**P**	Preparation (of a gesture or an articulate adaptor)	
	S	Stroke (of a gesture or an articulate adaptor)	
	R	Retraction (of a gesture or an articulate adaptor)	
	PnC-In	Independent position change	
	PnC-Dp	Dependent position change	
	U	Unstructured movement	
	Other	Other	Used for combined movements
***-mLt/RtStillness** *Stillness interval of the left / right hand*	***-mLtSt001** etc. ***-mRtSt001** etc.	Stillness interval code number	Code number indicates which hand is still

	***-mLt/RtStType** *Type of stillness for the left / right hand*	**Rest**	Rest (stopping in an unmarked position)	
		Hold	Hold (stopping in a marked position inside gesture or an articulate adaptor)	
		Frozen	Frozen (stopping in a marked position outside gesture or an articulate adaptor)	

Second level of segmentation				
***-mGesture** *Manual gesture*		***-mGe001** etc.	Gesture code number	
	***-mGeHandedness** *Handedness of the gesture*	**Lt**	Left-handed gesture	
		Rt	Right-handed gesture	
		Bh-sym	Two-handed gesture with symmetrical trajectory for both hands	
		Bh-id	Two-handed gesture with identical trajectory for both hands (parallel movement, one object)	
		Bh-dif	Two-handed gesture with different trajectories for both hands	
		Other	Other variants	Temporary tag for non-standard cases
	***-mGeStructure** *Gesture phase structure*	**P __**	Preparation within the gesture phase structure	
		__ S __	Stroke within the gesture phase structure	
		__ H __	Hold within the gesture phase structure	
		__ Hs __	Hold-stroke within the gesture phase structure	
		__ R	Retraction within the gesture phase structure	
		P-S ___	Combination of preparation and stroke within the gesture phase structure	
		___ S-R	Combination of stroke and retraction within the gesture phase structure	

	***-mGeFunction** *Gesture function*	**Depictive**	Depictive gesture	**Pragmatic** tag is not used separately, but can be combined with any of the above-mentioned functions
		Pointing	Pointing gesture	
		Beat	Beat gesture	
		Other	Other	
		Pragmatic	Pragmatic / metaphoric gesture	
	***-mGeTags** *Additional characteristics of the gesture*	**Shuttle**	Two-part gesture (back and forth, P-S-R with non-detectable internal boundaries)	
		Multi-S	Gesture with a multiple stroke	
		S Rebound, **R Rebound**	Rebound at the end of a stroke or retraction	
		Multi-Rebound	Gesture with a multiple rebound at the end of the stroke or retraction	
		Long R	Long retraction (in two or more steps)	
		Lt/Rt P/S/H/R Overlap	Gesture overlaps an indicated phase of another gesture of the indicated (left / right) hand	
		Repeat	Repetition of the previous gesture	
		GeBreakOff	Gesture truncation	
		GeFalstart	Gesture false start	
***-mAdaptor** *Manual adaptor*		***-mAd001** etc.	Adaptor code number	
	***-mAdType** *Manual adaptor type*	**Adaptor1**	Articulate adaptor	
		Adaptor2	Subtle adaptor	
		Adaptor1+2	Combined adaptor (complicated Type 1 adaptor)	
		Adaptor2+1	Accumulating adaptor (complicated Type 2 adaptor)	
		Other	Other variants	Temporary tag for non-standard cases

***-mPostureChange** *Manual posture change*	***-mPrC001** etc.	Manual posture change code number	
***-mPostureAccommodator** *Manual posture accommodator*	***-mPrA001** etc.	Manual posture accommodator code number	

Third level of segmentation			
***-mPosture** *Manual posture*	***-mPr001** etc.	Manual posture code number	
***-mPrPhase** *Transition phase from one manual posture to another*	**Transition**	The transition phase between the last occurrence of the neutral position A and the first occurrence of the neutral position B	
***-mGestureChain** *Manual gesture chain*	***-mGeCn001** etc.	Manual gesture chain code number	
***-mMovementChain** *Manual movement chain*	***-mMtCn001** etc.	Manual movement chain code number	
Supplementary tier			
***-mComments**	*Text*	Comments to the annotation	