methods

Annotating Hand Movements in Multichannel Discourse: Gestures, Adaptors and Manual Postures

Alla O. Litvinenko
Institute of Linguistics RAS, Moscow, Russia

Andrej A. Kibrik
Institute of Linguistics RAS, Moscow, Russia;
Lomonosov Moscow State University, Moscow, Russia

Olga V. Fedorova
Lomonosov Moscow State University, Moscow, Russia;
Institute of Linguistics RAS, Moscow, Russia

Julia V. Nikolaeva
Lomonosov Moscow State University, Moscow, Russia;
Institute of Linguistics RAS, Moscow, Russia

Abstract. This paper discusses the process of hand movement annotation in the context of multichannel discourse, based on a multichannel corpus of Russian conversations. The detailed annotation scheme that has been developed for the corpus is described, along with the corresponding annotation procedure. Hand movement in its entirety is treated as a complex combination of different (in)activities that must first be segmented and described as more elementary kinetic units, based on their kinetic characteristics (effort, velocity, trajectory, movement direction, etc.). After this first step, the segmented units or units' combinations may be interpreted as gestures or other types of manual behavior (stillness intervals of different kinds, hand posture changes, adaptors and so on), relying not only on their kinetic form, but also on their function in discourse. The resulting multilevel annotation may be used for researching interdependencies between manual behavior and other channels in multichannel discourse.

Correspondence: Alla O. Litvinenko, allal1978@gmail.com, Institute of Linguistics, Russian Academy of Sciences, 1 Bolshoi Kislovsky Lane, 125009, Moscow, Russian Federation; Andrej A. Kibrik, aakibrik@gmail.com; Olga V. Fedorova, olga.fedorova@msu.ru; Julia V. Nikolaeva, julianikk@gmail.com

Keywords: multimodality, multichannel discourse, multichannel corpus, hand movements, manual gesture, manual posture, adaptors, gesture annotation, annotation scheme

Acknowledgements. This study is supported by the Russian Science Foundation, Project #14-18-03819.

Received October 15, 2017, accepted June 25, 2018.

Introduction

The Western tradition of understanding body movements and gestures as an important part of communication goes back to Classical Antiquity. Since then, gesture has been extensively discussed as a major component in the art of rhetoric, as a possible form of a universal language, and as a medium in which natural language was first formed (detailed reviews on the history of the topic can be found in Graf, 1994; Dutsch, 2013; Kendon, 2004; Müller, Ladewig, & Bressem, 2013). Nonetheless, during the 20th century, due to the domination of structuralist methods and views in linguistics, gestures were seen as a component of speech or language usage only, as opposed to the true language system, and, as a result, were mostly excluded from linguistic studies (for some exceptions see Pogodin, 1913; Yakubinsky, 1923/1986; Dobrogaev, 1931; Pike, 1967). Meanwhile, important studies of gestures and body movements started within the fields of psychology and anthropology (Efron, 1941/1972; Ekman & Friesen, 1969). Among the pioneers who proposed studying human kinetic behavior, and hand gestures in particular, were Ray Birdwhistell (1970) and Adam Kendon (1972). Kendon demonstrated that communicative body motions possess similar structure and hierarchy to speech (1972), and later postulated the idea of gesture and language being two sides of a single process of utterance (Kendon, 1980). Despite that, most linguists still left gestures outside of their area of interest, and psychologists for a time concentrated on the idea of body movements being a form of non-verbal communication (Hinde, 1972; Argyle, 1975; Scherer & Ekman, 1982; Feldmann & Rimé, 1991). However, not only Kendon, but also David McNeill (1979, 1985) challenged the concept of non-verbal nature of gesticulation, which led to the active scientific debates on the topic in the second half of the 1980s. These discussions became the turning point for cognitive linguists' interest in gestures. In 1992, McNeill formulated his integrated theory of gestures and speech, and in the subsequent decades there has been increasing research of this area (Cienki 1998a, 1998b; Müller, 1998; McNeill, 2000; McNeill & Duncan 2000 Kita, 2003; Kendon, 2004; McNeill, 2005; Cienki & Müller, 2008a, 2008b; Kopp, Bergmann, & Wachsmuth, 2008; Müller & Tag, 2010; Müller et al. (Eds.), 2013; Abner, Cooperrider, & Goldin-Meadow, 2015; inter alia).

Studying various aspects of the interaction between gesture and speech requires significant data accumulation and analysis, and therefore the development of annotation systems and corpora. Numerous gesture annotation schemata have been designed, with different advantages, limitations and theoretical bases, depending on specific research questions and goals, and on the available material and resources; see, for instance, Multimodal Score annotation scheme for ANVIL™ (Caldognetto, Poggi, Cosi, Cavicchio, & Merola, 2004), CoGestT system for TASX-annotator (Trippel et al., 2004); MUMIN coding scheme for ANVIL™ (Allwood, Loredana, Jokinen, Navarretta, & Paggio, 2007); Michael Kipp's system for ANVIL™ (Kipp, Neff, & Albrecht, 2007); LASG system (Bressem, Ladewig, & Müller, 2013); NEUROGES scheme (Lausberg & Sloetjes, 2015); PAGE project (Karpiński, Jarmołowicz-Nowikow, & Czoska, 2015). Detailed overviews of the existing gesture annotation schemata were provided by Wagner, Kopp, and Malisz (2014) and by Karpiński et al. (2015); Duncan, Rohlfing, and Loehr (2013) reviewed the most popular multimodal annotation tools used for gesture annotation. Ide and Pustejovsky (2017) offer a thorough review of various annotation systems and principles, as well as various annotation tools.

This paper discusses the process of hand movement annotation in the context of multichannel discourse, belonging to the field of multimodal (multichannel) linguistics that sees various formal means used in natural communication — words, prosody, manual gestures, head and body movements, oculomotor behavior and so on — as being equally significant (Gibbon, Mertins, & Moore, 2000; Kress, 2002; Granström, House, & Karlsson, 2002; Scollon, 2006; Kibrik, 2010; Müller et al. (Eds.), 2013; Kibrik, 2018a; inter alia).

The exploration of multichannel communication should be based on the analysis of each channel, considered individually. Furthermore, the stream of behavior in each channel must be understood as a series of units, down to the elementary units produced by interlocutors (Wagner et al., 2014; Kibrik, 2018a). In a multichannel corpus, one should strive to apply similar principles to annotating various kinetic channels (hand, head and body movements, for example), as long as these channels do not display certain special properties that prevent such similarity.

The annotation scheme and the procedure presented here have been developed as a part of the Russian Pear Chats and Stories corpus project, based at the Institute of Linguistics of the Russian Academy of Sciences (www.multidiscourse.ru). The project participants set a goal to develop and test a flexible annotation scheme that could be used for a variety of research questions on the interaction of discourse channels, such as the temporal coordination between discourse units of different channels and levels; individual variations in gesticulation; speech disfluencies in relation to gesticulation; and so on. The current work also relies on the previous experience in developing Russian spoken corpora (Kibrik & Podlesskaya, 2009; http://www.spokencorpora.ru/) and takes into account various gesture studies and multimodal corpora created within the Russian linguistic school of thought (Kreydlin, 2002, 2010; Nikolaeva, 2004, 2009; Grishina, 2017; www.ruscorpora.ru/murco-authors.html).

The corpus includes Russian discussions of the Pear Film (Chafe, 1980; www.linguistics.ucsb.edu/faculty/chafe/pearfilm.htm) and consists of two parts. The first one was recorded in 2015; it contains 24 sessions (or 'recordings') with 96 Russian native speakers aged 18–36 (34 men and 62 women); the total duration is about 9 hours, with the length of a recording varying from 12 to 38 minutes (average 24); the total word count is about 100,000 words. The second part of the corpus was recorded in 2017; it includes 16 recordings with 64 Russian native speakers (16 men and 48 women); its total duration is about 6 hours, with the length of a recording varying from 8 to 41 minutes (average 21); the total word count is approximately 60,000 words. Each recording includes two monologic (narrative) parts and one interactive (discussion) part; the four participants have fixed roles: Narrator, Commentator, Reteller and Listener. In each session, three frontal video cameras (100 frames per second and 1392 × 1000 pixels)

Figure 1. Manual gesture.

Figure 2. Articulate manual adaptor, type 1 (adjusting glasses).

recorded three main participants, and these data were later analyzed in ELAN (tla.mpi.nl/tools/tla-tools/elan). All the participants have given written consent to using their recordings and images for scientific research. For a more detailed description of the multichannel corpus design, see (Fedorova, Kibrik, Korotaev, Litvinenko, & Nikolaeva, 2016; Kibrik, 2016, 2018b).

At the time of this publication, the sample subcorpus that can be found at the project website consists of three recordings with nine participants and a total length of 1 hour. The subcorpus has been used as a testing ground for the present annotation scheme.

In the following sections of this paper, we first discuss the key issues that arise while trying to provide consistent manual (hand) gesture annotation and the types of kinetic behavior that are involved in the gesticulation process. Second, we describe the proposed annotation procedure, segmentation guidelines and the set of ELAN tiers that might be used for implementing the scheme. Finally, we briefly address several typical annotation issues and the practical ways to resolve them that we have developed for the corpus.

Manual Kinetic Behavior: Gestures and Kinetic Background

In order to perform valid and consistent gesture annotation, one needs to address two major issues. The first one is developing adequate segmentation rules and procedure, ensuring that one obtains comparable units of analysis. This issue has been discussed in detail in a number of studies (McNeill, 1992; Kita, van Gijn & van der Hulst, 1998; Kipp, 2004; Wagner et al., 2014; inter alia); in the subsequent sections of this paper we outline the segmentation decisions we have made for our corpus.

The second issue is a quest for systematic distinction between gestures and other types of kinetic behavior. This issue can be resolved on functional (Kreydlin, 2002; Kendon, 2004; Nikolaeva, 2004; McNeill, 2013; Grishina, 2017) or formal grounds (Nikolaeva, 2009; Müller, Bressem, & Ladewig, 2013; Bressem et al., 2013). We employ a combined approach to gesture identification, relying both on the kinetic parameters of movements and on their functioning in discourse. The motor behavior of a communicating person, taken as a whole, should be seen as kinetic background, against which gestures stand out as figures. Moving one's hands, head and body requires effort, and if a movement has features such as distinct shape, location and motion patterns, the result of the corresponding effort is recognized as communicatively meaningful (Müller & Tag, 2010). Thus, we understand *gesture* as an articulatory effort with a primary discourse function: adding to speech content (e.g. iconic or metaphoric gestures) or addressing discourse production processes (facilitating verbalization, appealing to an interlocutor, regulating turn-taking, adding to rhythmical structure, etc.). Gestures possess internal structure and may consist of several phases (Kendon, 1980, 2004; see below for details), a *stroke* being the constituting phase, both kinetically and functionally.

An example of a manual gesture can be seen in Figure 1: the speaker represents a gardener looking at a pear in his hand, simultaneously saying "He looks at it".

The kinetic background consists of several types of activities and/or motionlessness. The most obvious part of it is *resting* (being motionless in a rest position), and here we observe much variation between speakers. Some of them tend to be motionless for extended periods of time between relatively long gesticulation series; others tend to either rest or speak and gesticulate for short periods, and some tend to almost never be completely motionless, going from gesticulation to other types of movement and back again. Any speaker has a number of typical hand positions (neutral positions) that s/he returns to when gesticulation ends; some of these positions are used for rest, while others serve as boundary markers between gestures. Retaining a particular hand position as a starting point for movements makes a *manual posture*; specific hand movements aimed at changing a manual posture are *posture changes*.

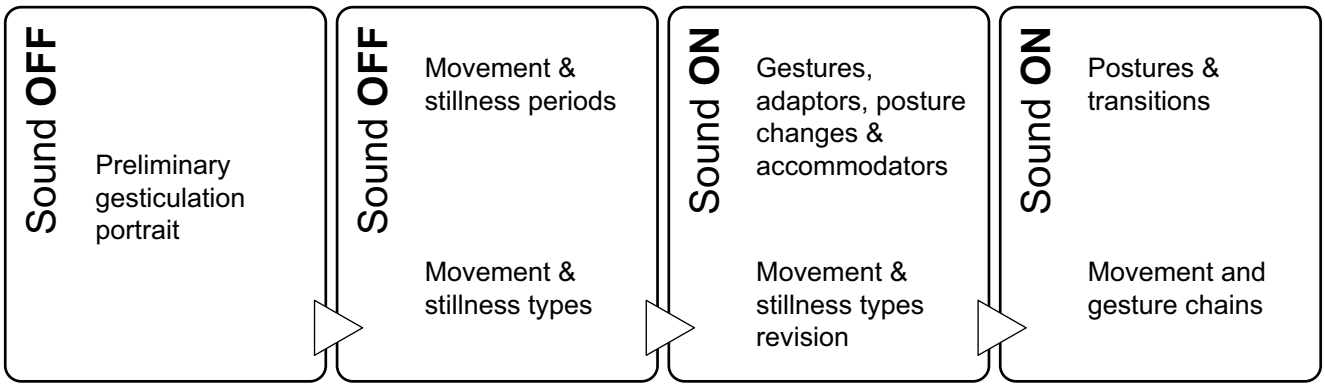

Figure 3. Annotation procedure stages (Preliminary Stage → Stage 1 → Stage 2 → Stage 3).

Another important part of the kinetic background is formed by various non-speech-motivated movements, so-called *adaptors* (Ekman & Friesen, 1969), that involve touching one's body (self-adaptors) or manipulating external objects (object-adaptors). A detailed overview of this domain can be found in Żywiczyński, Wacewicz, and Orzechowski (2017); their research shows that the increased use of adaptors may be connected to turn-taking.

"Articulate", or "type 1", adaptors are well-articulated movements that are structurally close to gestures, as they have similar phases (see below). These adaptors usually have a clear non-speech-related purpose, like scratching an itch, adjusting clothes/hair/glasses, etc. An example of an articulate adaptor can be seen in Figure 2. "Subtle", or "type 2", adaptors are characterized by weak articulation; they consist of small repetitive motions that do not possess regular internal structure. These movements often seem to be stress relievers or signs of anxiety; the most typical examples are continuous rubbing of the fingers or fiddling with a pen.

Systematic differentiation between the kinetic background and gesticulation requires not only being able to describe both in more or less objective terms, but also taking into account individual differences. To address this issue, we employ the notion of a speaker's gesticulation portrait (Litvinenko, Nikolaeva, & Kibrik, 2017a; Kibrik & Fedorova, 2018; cf. prosodic portraits in Kibrik, 2009), which allows us to aggregate information on a particular speaker's kinetic habits. In order to describe a speaker's gesticulation portrait and to quantify its parameters, we need to include the aforementioned types of motion and stillness in the annotation scheme.

Annotation Scheme: Overall Description and Procedure

The annotation scheme includes two coordinated components: a system of annotation tiers for kinetic units of different hierarchical levels and a procedure of annotation as implemented by an expert. Partial descriptions of the proposed scheme have already been presented (Litvinenko, Nikolaeva, & Kibrik, 2017a, 2017b). However, since those publications, the segmentation rules have been significantly updated, and the procedure has been fully reworked. Previous versions of the scheme involved singling out annotated objects (gestures and/or adaptors) from the kinetic background at the very first stage of the annotating process, and after that stage, the "background" part was ignored. While applying this scheme to the corpus, we found it insufficient, as it relied heavily on the annotators' ability to recognize units correctly before analyzing their structure and functions. This led to increased inconsistency between annotators and between the annotations created for different speakers.

The latest version of the annotation scheme, presented here, proceeds from the basic distinction between motion and stillness (or inaction) to more complex structures at the next levels. Likewise, the new procedure involves segmenting manual kinetic behavior as a whole into basic units before moving on to differentiating gestures from other types of kinetic behaviors and describing their functional qualities. Both the procedure and the tier system are organized hierarchically and involve moving "bottom up", from lesser structural units (movements and stillness intervals) to more complex functional ones (e.g. gestures) and then to the units' sequences and combinations (e.g. gesture chains).

The annotation procedure can be described as a series of at several successive stages, or passes, (see Figure 3), of which one is preliminary and three correspond to the three levels of kinetic units (Figure 4), to the three levels of the annotation scheme and to the three groups of annotation tiers in ELAN (Figure 5).

The goal of the preliminary stage is to establish typical kinetic patterns for a speaker: to estimate normal velocity, amplitude and effort for gestures and adaptors, the frequency and average duration of stillness periods, and to capture the most common neutral positions and posture changing patterns. This part of the procedure is not further formalized, yet it allows annotators to be more consistent in analyzing the data of a given speaker. For example, if a speaker regularly produces high-effort gestures with multiple rebounds (recoils), each of those movements would be described as a single gesture, rather than a series involving the first gesture and a number of subsequent shorter beat gestures. The result of the preliminary pass must be a preliminary gesticulation portrait of the speaker, aiding in unit identification at further passes.

The first stage serves to divide the manual motion flow into separate movements and stillness periods (thus annotating first level kinetic units) and to assign

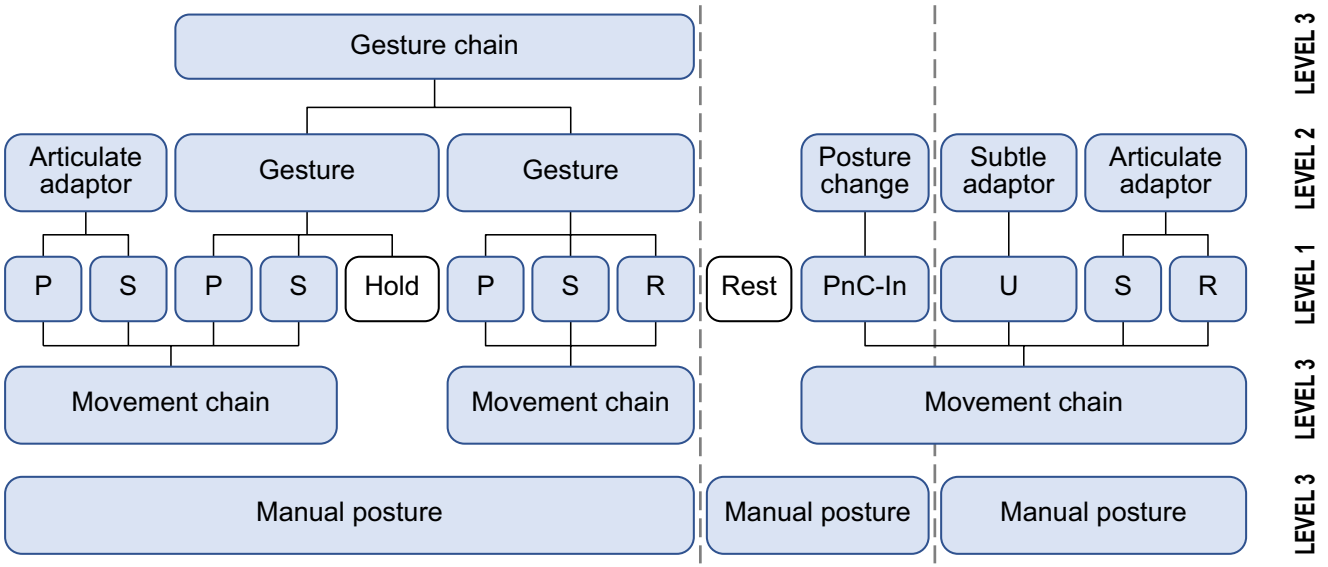

Figure 4. Manual units of different levels. Movement types: P — preparation; S — stroke, R — retraction, U — unstructured, PnC-In — independent position change.

their structural types (e.g. preparation, stroke, retraction, position change, unstructured period, etc.; see the next section for details). Both the preliminary and first stages are done without relying on the speech content (with sound muted), thus minimizing the influence of speech unit boundaries on the decisions regarding the kinetic unit boundaries.

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). A detailed system of tags enables one to systematically describe gesture overlaps and gesture repetitions (Litvinenko et al., 2017a; http://multidiscourse.ru/annotation/). This pass is performed with the sound on, in order to allow for verification of the previous annotation with the help of the speech context. The boundaries of the first level units may also be adjusted, if needed, but within certain limits: the adjustments are allowed only when an annotator must choose between two segmentation versions, depending on interpretation, all other factors being equal. For example, we may see two very similar movements in a row with a slight difference in speed or amplitude and have to decide if this is one multi-stroke gesture or two separate strokes (and two gestures). Suppose from the point of view of the kinetic structure as such that this series is not smooth enough to be confidently analyzed as one repetitive stroke, but at the same time its parts are not different enough to be confidently annotated as two independent strokes. In such instances, the annotator may take into account the speech content and prosodic structure, in order to resolve the issue.

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 4. 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.

Annotation Scheme: Kinetic Units and Segmentation Guidelines

In this section, we discuss in detail the kinetic units of each level and the guidelines for their segmentation and annotation. The system of ELAN tiers we propose for annotating manual behavior is designed to correspond with the hierarchy of the kinetic units and with the annotation procedure stages. This system of tiers grouped by level can be found in Figure 5.

In the end of the section, a brief review of the ELAN tier system is provided.

Each stage of the annotation scheme should be completed before moving on to the next one; however, depending on the research goals, the second and third stages might be done in more or less detail. For example, one might code only gestures or only adaptors on the second stage or skip postures or movement chains on the third. The scheme is designed in such a way that it might be applied in parts, expanding or abridging in accordance with the annotators' needs, given that the first stage is fully completed, and all the first level kinetic units are annotated.

First Annotation Level: Movements, Stillness Intervals and Their Types

At the first, most basic level, we annotate the simplest motion units (or *movements)* for each hand separately, along with their structural types.

The criteria for the segmentation and identification of movement units have been discussed in the literature (McNeill, 1992; Kita et al., 1998; Ladewig & Bressem, 2013; Lausberg & Sloetjes, 2015). As for the identification of manual movements, we use the criteria proposed by Kita et al. (1998), with some additions and modifications that will be described below.

What an interlocutor (or an annotator) sees as an articulatory effort can be described with a set of specific kinetic parameters: hand shape and orientation, movement trajectory, amplitude and direction, velocity pattern (acceleration, deceleration, steady speed), location in gesture space and, for repeated movements, rhythm

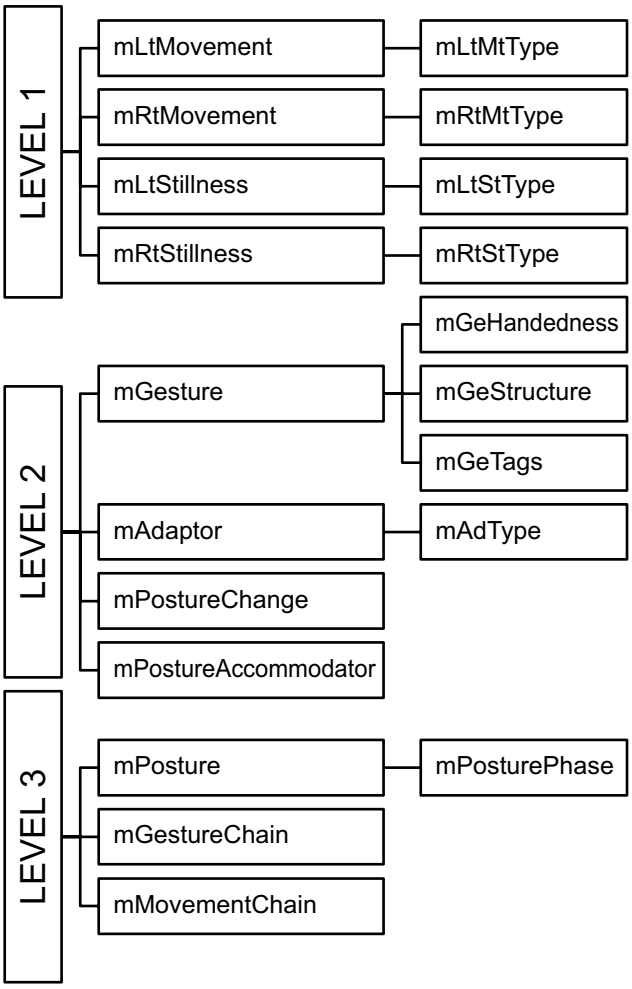

Figure 5. ELAN manual annotation tier system.

(or lack thereof). Combined, these kinetic characteristics form a motion pattern. When dividing the ongoing stream of manual behavior into a series of separate movements, annotators identify the points where kinetic features change. A simultaneous change in two or more parameters is interpreted as a movement boundary; a change in several (usually four or more) features points to a higher-level unit boundary. In addition to abrupt direction and velocity pattern changes that are discussed by Kita et al. (1998), we also take into account significant (equal to 1/2 hand size or more) amplitude and location changes. For example, if the same hammering motion is repeated twice, first with high amplitude, and then with a much smaller one, we annotate it as two separate movements, even if there is no change in speed and trajectory between repetitions.

According to Kita et. al. (1998), the onset of a movement is used as the beginning of the movement unit, and the point where the hand first comes into a resting position is considered the gesture unit's end. If a hand rebounds (recoils, in Kita's terminology) in the process or settles into a more comfortable position, this part is excluded from the gesture unit. However, this approach is reasonable if one tries to annotate gestures only; if we aim to segment and annotate the kinetic flow as a whole, any movement unit should count as a structural part of some higher-level unit. Hence, we also use the onset of any movement for its left boundary, and any movement that is not a result of a separate effort (e.g. hand rebound) is treated as a part of the same movement that caused it (e.g. retraction). Thus, for the right boundary of any unit we use either the beginning of the next unit or the point where a hand stops moving altogether.

Each of the segmented movements must be further classified as one of the structural types, depending on its functioning in the motion flow. Later in the procedure, these types shall be interpreted as *gesture phases* (Kendon, 1980, 2004; Ladewig & Bressem, 2013), *adaptor phases*, *posture changes* or *posture accommodators* (see the next subsection) and reviewed accordingly if needed. However, during the first pass the annotator relies only on kinetic characteristics and does not yet differentiate between functional units (gestures, adaptors, etc.). The most articulate movements that demonstrate more exerted force (more velocity and/or tension) and are kinetically independent (the motion itself serves as its own purpose, as opposed to being necessary for another motion to be executed) are classified as *strokes* (S). Transitional movements from a neutral position or from a previous gesture to a stroke are considered *preparations* (P), and a *retraction* (R) is a transitional movement between a stroke and a neutral position. In some cases, retractions may be unfinished as a speaker starts the next motion and a neutral position is never reached. Nonetheless, such cases are also coded as retractions, with the movement direction defining their structural role.

Certain sequences of movements with weak articulation do not undergo segmentation; they belong to a separate type of *unstructured* movement (U). A movement is considered unstructured if it is simultaneously small in amplitude (less than 1/2 hand size), involves little effort and has no regular trajectory or hand shape pattern.

There are two types of *position change* movements, depending on their role in the movement flow. The first type occurs anywhere in the flow and serves to move a hand from one neutral position to another; these are *independent* position changes, PnC-In in our notation. The second type serves to adjust one hand's position in order to extract, move or reposition the other hand as needed for producing another movement; these are *dependent* position changes, PnC-Dp. These movements occur when the starting or ending position of a movement involves putting two hands together and moving one hand requires moving the other. If both hands change position simultaneously, the movement is considered independent for each of them.

Apart from movements, there are obviously periods of non-movement, or stillness. All stillness intervals must also be annotated and classified. Staying in a neutral position with relaxed hand (s) is considered a *rest*. In contrast, hands staying in any marked (non-neutral and/or tense) position constitutes either a *hold* (before or after a stroke, usually after a preparation or before a retraction when these phases are present) or a *frozen* interval (appearing after a retraction or before a preparation). On the next annotation level, hold intervals are interpreted as gesture phases, and frozen intervals remain independent units. This interpretation may be corrected later when gesture boundaries are defined and verified by speech content.

Thus, at the first annotation level, we get the ongoing kinetic stream produced by a speaker segmented into basic units and code these units' structural types depending on the way they relate to each other.

Second Annotation Level: Functional Units

At the second level, we annotate functional units — individual *gestures* (= Kendon's gesture phrases), *adaptors*, *posture changes* and *posture accommodators*. These units can be one-handed or two-handed; on this level they are annotated for both hands combined. Gesture structure and the annotation of gesture characteristics (handedness, multi-strokes, etc.) have been described in detail in our previous work (Litvinenko et al., 2017a).

Structural movements and stillness types that were coded during the previous pass must now be interpreted as gesture or adaptor phases, and then reviewed using speech content. The stroke movement functions as the nucleus of a gesture and conveys most of its meaning. It is an obligatory phase for any uninterrupted gesture. Other phases (preparation, hold and retraction) may be absent. Note that we do not posit so-called independent holds (Kita et al., 1998), as it is hard to prove that a hold has an independent meaning and is not just emphasizing the preceding stroke. In some asymmetrical gestures (e.g. drawing with the finger of one hand on the palm of the other hand), one hand may perform a stroke and another a hold; thus, some two-handed gestures can have only one stroke and an equally significant hold, but a hold without a stroke is not enough to form a gesture.

Some stillness intervals with hands staying in a marked position are structurally and functionally ambiguous. With preparation or retraction absent, it is sometimes hard to differentiate between a hold and a frozen interval, and an annotator may use speech and data from other channels to support one of the possible interpretations. However, by default, we consider any halting in a marked position a part of the previous context. For example, without a following retraction, and if speech content does not suggest otherwise, a post-stroke stillness interval should be coded as a post-stroke hold (and a part of the corresponding gesture). Without a preceding preparation, a pre-stroke marked stillness would be considered a frozen interval.

Articulate (type 1) adaptors have the same phase structure as gestures. They include a stroke — a main part that serves to achieve the adaptor's goal (e.g. to adjust glasses or to scratch the nose) — and subsidiary phases needed to move the hand from a neutral position and to move it back. Subtle (type 2) adaptors consist of undivided U movements only.

At this level, position change movements are interpreted as one of the two functional phenomena: a posture change or a posture accommodator. *Posture changes* are position change movements that result in taking a new neutral position (thus starting a new posture, see below) for at least one hand. Both independent and dependent position changes can form a posture change. In contrast, *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. Dependent position changes are interrupted more frequently than independent ones, but it is possible for both types.

The second level's units are functional; here the structural units of the first level get interpretation as the constituting parts of larger entities.

Third Annotation Level: Manual Postures, Movement Chains and Gesture Chains

At the third level we annotate composite objects that consist of the second level's functional units.

Gesture chains are uninterrupted series of gestures; the continuity and quantity of such chains give information both on the speaker's personal habits and on the gesticulation aspects in various discourse situations.

Movement chains are uninterrupted series of movements. These chains combine gestures with other movements and exclude any periods where both hands are still; the movement chain characteristics are important for understanding kinetic background and a person's kinetic behavior as a whole.

Manual postures in the proposed notation system represent a more complex phenomenon. A *manual posture* is an interval during which a hand assumes a certain neutral position (A), lasting until the moment when another neutral position (B), nonidentical to A, is taken. If a hand leaves position A and returns to it after a series of movements, it is treated as the same posture. Any manual posture consists of two parts: a "stable" phase where the starting neutral position is maintained and/or returned to by both hands, and a transition phase which starts when a hand leaves position A and never returns to it later. As a result, the analyzed kinetic flow is divided into postures without remainder. For now, we annotate only transition phases in a dependent posture phase tier.

As with the second level units, the posture tier is used for annotating postures for both hands combined; therefore, rules must be applied for any cases of asynchronous posture changing. As soon as any hand reaches a new neutral position, a new posture starts; as soon as any hand leaves its starting position (A) for the current posture, the transition phase starts. This approach allows annotators to clearly differentiate between the "stable" part where both hands maintain their starting position and the transition intervals when at least one hand has no neutral position that can be seen as the starting point for movements.

Sometimes hands assume one neutral position after another without any "stable" interval; in such cases, a manual posture may consist of a transition phase only, and this is a point of significant individual variation. Some people change postures more often than others; some tend to change postures simultaneously with both hands, while others do this alternately with their left and right hands, which results in shorter postures and more numerous transitions.

Implementing the Scheme in ELAN

The scheme is implemented as an ELAN template with a system of tiers, including necessary tier types and the corresponding controlled vocabularies. Detailed technical instructions for ELAN users can be found at the project website, along with the latest scheme descriptions and annotation samples.

The independent tiers for various objects (movements, stillness periods, postures, posture changes and accommodators, gestures, adaptors, gesture and movement chains) are attached to the time axis. Subordinate tiers are used for annotating object features and other parameters. These tiers, combined with a

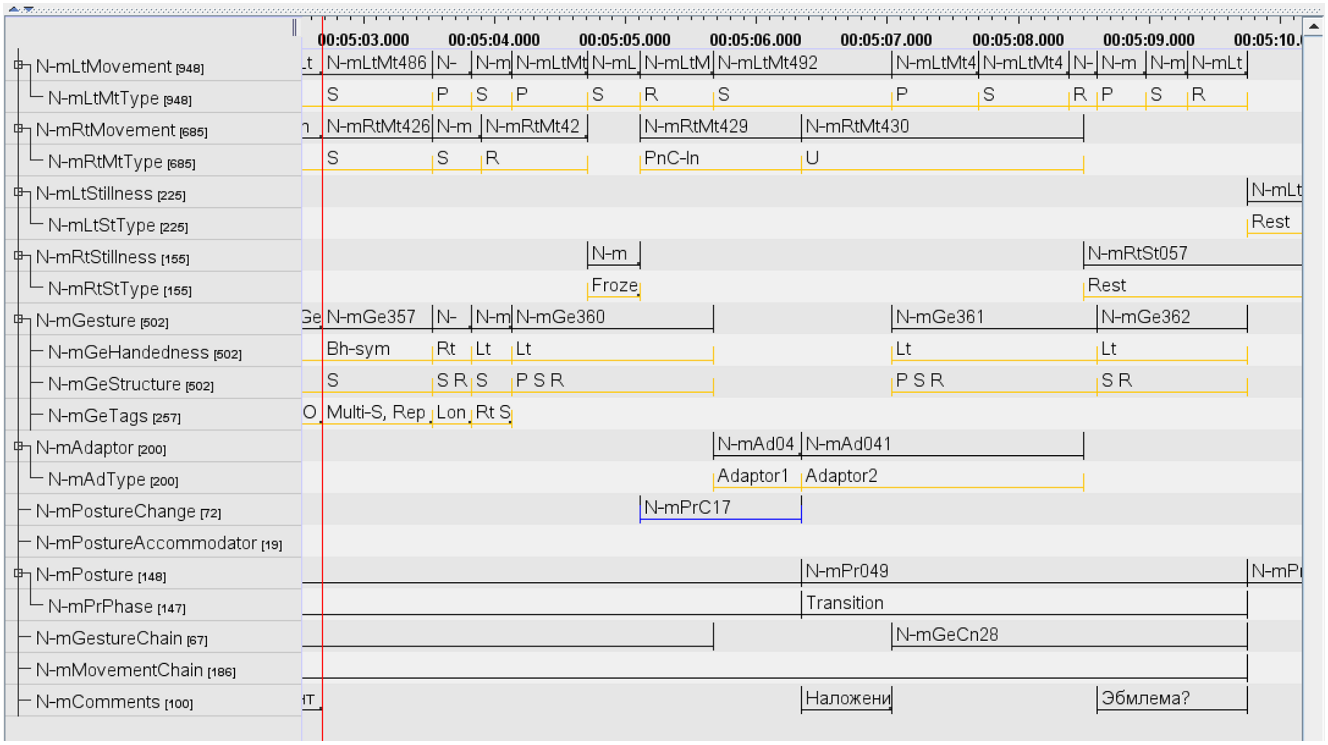

Figure 6. Hand movement annotation in ELAN.

detailed tag system, describe various gesture characteristics, such as handedness, phase structure, structural types of stroke (e.g. multi-strokes), gesture repetitions, the degree of synchronization between hands, etc. As left and right hand movements are often asynchronous, a set of formal rules has been designed in order to establish gesture boundaries and to annotate gesture overlaps (see Litvinenko et al., 2017a or the project website).

An example of the resulting annotation in ELAN can be seen in Figure 6.

Gestures vs. Adaptors: Resolving Issues

The task of systematic differentiation between gestures and adaptors is resolved both at the formal and functional levels. Prototypical gestures are mostly performed off the body in their stroke phase and do not have any "practical" purpose. Articulate adaptors (type 1), while having a phase structure similar to gestures, demonstrate less effort in stroke and often feature a simpler trajectory and hand shape in the auxiliary phases, as movement is directed straight to the target point of the effort or to the rest position. Subtle adaptors (type 2) feature weak articulation, on-body movement, slower speed and small amplitude. By definition, they do not shift far from the starting point, which is usually a resting position. In difficult cases, it is recommended to refer to the typical patterns of the given speaker, and to treat more ambiguous cases by analogy. This strategy is especially useful for movements with weakened articulation. For example, if a speaker uses a specific recurrent gesture with a half-curled hand off the body towards other communicators (a frequent variation of the common "palm up open hand gesture", see Müller, 2004), even a loose version of this typical movement should be coded as a gesture, especially if this interpretation is supported by the speech context. If a speaker does not normally use such gestures and speech does not help, such an ambiguous fragment should be coded as a subtle adaptor.

Some of the depictive gestures may actually reenact adaptors in the discourse space; in this case only the speech content allows one to identify them as gestures. The narrator in Figure 7 is talking about a boy who fell off of a bicycle and then lowered his sock to check if he had hurt his leg; the narrator's descriptive gesture with her right hand on her own leg ('he lowers his sock') may look like an adaptor.

The distinction between gestures and adaptors is not absolutely clear-cut and sometimes has to be a matter of the coders' interpretation. Experienced annotators interpret particular movements as either gestures or adaptors with a fair degree of confidence. There are some dubious instances that, for the sake of categorical interpretation, must be treated as either this or that. Comments may be added to tag difficult instances.

Previous versions of the scheme have been partially tested for the match of independently coded phases, and we have been confronted with the same difficulties, as discussed by Kita et al. (1998). In particular, sequences of short movements with small amplitudes and less effort are relatively difficult to analyze and therefore become the cause of discrepancies. The reasons for most disagreements are a) the choice between a single complicated stroke as opposed to a preparation-stroke combination; b) the choice between a multi-stroke and a series of smaller independent strokes. The current version of the scheme should significantly lessen such problems, as it involves a more formal segmentation procedure reducing interpretation influences on the segmentation process.

To minimize possible inconsistencies and interpretation dependencies, all recordings are analyzed by at least two different annotators (the main coder and the reviewer) and all ambiguous cases are discussed in detail until an agreement is achieved.

Figure 7. 'The boy lowers his sock' gesture looking like an adaptor.

Conclusion

Detailed hand movement annotation is essential for understanding the interdependencies between manual gesticulation and other channels in multichannel discourse. In order to properly analyze co-speech gestures, we must compare them to other types of manual kinetic behavior, including adaptors and posture changes. Moreover, individual variation in this domain is very high, and in order to compare different participants' gesticulations we need to create gesticulation portraits, describing an individual's kinetic habits.

The underlying principles of kinetic process description are important for understanding the way multichannel discourse works. Defining gesture as a figure that stands out against a kinetic background brings us to the concept that in order to understand gesture we must also thoroughly investigate other types of kinetic behavior. The proposed annotation scheme not only provides a basis for a detailed analysis of manual movements and helps to take individual variation into account, but may also be applied to other kinetic channels, thus building a foundation for exploring inter-channel dependencies. A systematic description of motion and stillness, postures and adaptor-like movements exists in all kinetic channels, and probably in some way can also be applied to the verbal and prosodic channels.

The annotations based on the coding scheme presented here or on its previous versions have already been used for studies of temporal coordination of gestures and speech (Fedorova et al., 2016) and for research on gesticulation portraits (Kibrik & Fedorova, 2018). A multichannel approach to communication entails significant changes in linguistic research methods and theory, introducing new questions in such areas as turn-taking in dialogue, distribution of information and temporal coordination between different channels, defining basic (minimal, elementary) communication units, etc. (Kibrik, 2018b). The proposed method of multichannel communication description and multichannel corpus design may be valuable for future studies in these areas.

References

Abner, N., Cooperrider, K., & Goldin-Meadow, S. (2015). Gesture for linguists: A handy primer. *Language and Linguistics Compass, 9*(11), 437–451. doi:10.1111/lnc3.12168

Allwood, J., Loredana, C., Jokinen, K., Navarretta, C., & Paggio, P. (2007). The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena. *Language Resources and Evaluation, 41*(3–4), 273–287. doi:10.1007/s10579-007-9061-5

Argyle, M. (1975). *Bodily communication.* New York: International Universities Press.

Birdwhistell, R.L. (1970). *Kinesics and context.* Philadelphia: University of Pennsylvania Press. doi:10.9783/9780812201284

Bressem, J., Ladewig, S.H., & Müller, C. (2013). Linguistic Annotation System for Gestures (LASG). In C. Müller, A. Cienki, E. Fricke, S.H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body — Language — Communication: An international handbook on multimodality in human interaction* (Vol.2 (pp.1098–1124). Berlin: Mouton.

Caldognetto, E.M., Poggi, I., Cosi, P., Cavicchio, F., & Merola, G. (2004). Multimodal score: An ANVIL based annotation scheme for multimodal audio-video analysis. In J.-C. Martin, E.D. Os, P. Kühnlein, L. Boves, P. Paggio, & R. Catizone (Eds.), *Proceedings of the LREC-Workshop on multimodal corpora* (pp.29–33). Lisbon, Portugal: Retrieved from http://www.lrec-conf.org/proceedings/lrec2004/ws/ws10.pdf.

Chafe, W. (Ed.). (1980). *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production.* Norwood, NJ: Ablex.

Cienki, A. (1998a). Metaphoric gestures and some of their relations to verbal metaphorical expressions. In J.-P. König (Ed.), *Discourse and cognition: Bridging the gap* (pp.189–204). Stanford, CA: Center for the Study of Language and Information.

Cienki, A. (1998b). Straight: An image schema and its metaphorical extensions. *Cognitive Linguistics, 9*(2), 107–150. doi:10.1515/cogl.1998.9.2.107

Cienki, A., & Müller, C. (2008). Metaphor, gesture and thought. In R.W. Gibbs (Ed.), *Cambridge handbook of metaphor and thought* (pp.483–501). Cambridge: Cambridge University Press.

Cienki, A., & Müller, C. (Eds.). (2008). *Metaphor and gesture.* Amsterdam: John Benjamins Publishing Company. doi:10.1075/gs.3

Dobrogaev, S.M. (1931). Uchenie o reflekse v problemakh yazykovedeniya [Reflex theory and issues in linguistics]. In N.Y. Marr (Ed.), *Yazykovedenie i materializm [Linguistics and materialism].* (Issue 2 pp.105–173). Leningrad: Sotsekgiz. (In Russian).

Duncan, S., Rohlfing, K., & Loehr, D. (2013). Multimodal annotation tools. In C. Müller, A. Cienki, E. Fricke, S.H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body — language — communication: An international handbook on multimodality in human interaction* (Vol.1. pp.1015–1022). Berlin: Mouton.

Dutsch, D. (2013). The body in rhetorical delivery and in theatre — An overview of classical works. In C. Müller, A. Cienki, E. Fricke, S.H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body — language — communication: An international handbook on multimodality in human interaction* (Vol.1, pp.329–343). Berlin: Mouton.

Efron, D. (1972). *Gesture, race and culture.* Paris: Mouton. (Originally published 1941).

Ekman, F., & Friesen, W.V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*(1), 49–98. doi:10.1515/semi.1969.1.1.49

Fedorova, O.V., Kibrik, A.A., Korotaev, N.A., Litvinenko, A.O., & Nikolaeva, J.V. (2016). Vremennaya koordinatsiya mezhdu zhestovymi i rechevymi edinitsami v multimodalnoy kommunikatsii [Temporal coordination between gestural and speech units in multimodal communication]. In *Computational linguistics and intellectual technologies. Papers from the Annual International Conference "Dialogue"* (Vol.15, pp.159–170). Moscow: RGGU. (In Russian).

Feldmann, R.S., & Rimé, B. (Eds.). (1991). *Fundamentals of nonverbal behavior.* Cambridge: Cambridge University Press.

Gibbon, D., Mertins, I., & Moore, R.K. (Eds.). (2000). *Handbook of multimodal and spoken dialogue systems: Resources, terminology and product evaluation.* Berlin: Springer.

Graf, F. (1994). Gestures and conventions: The gestures of Roman actors and orators. In J. Bremmer, & H. Roodenburg (Eds.), *A cultural history of gesture* (pp. 36–58). Cambridge: Polity Press.

Granström, B., House, D., & Karlsson, I. (Eds.). (2002). *Multimodality in language and speech systems.* Dordrecht: Kluwer.

Grishina, E.A. (2017). *Russkaya zhestikulyatsiya s lingvisticheskoy tochki zreniya. Korpusnye issledovaniya [Russian gesticulation in linguistic perspective. Corpus studies].* Moscow: LRC Publishers. (In Russian).

Hinde, R.A. (Ed.). (1972). *Nonverbal communication.* Cambridge: Cambridge University Press.

Ide, N., & Pustejovsky, J. (Eds.). (2017). *Handbook of linguistic annotation.* Netherlands: Springer. doi:10.1007/978-94-024-0881-2

Karpiński, M., Jarmołowicz-Nowikow, E., & Czoska, A. (2015). Gesture annotation scheme development and application for entrainment analysis in task-oriented dialogues in diverse cultures. In G. Ferré and M. Tutton (Eds.), *Proceedings of GESPIN 2015 Conference* (pp. 161–166). Nantes, France.

Kendon, A. (1972). Some relationships between body motion and speech: An analysis of an example. In A.W. Siegman, & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177–210). New York: Elsevier.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M.R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–227). The Hague: De Gruyter Mouton.

Kendon, A. (2004). *Gesture: Visible action as utterance.* New York: Cambridge University Press. doi:10.1017/cbo9780511807572

Kibrik, A.A. (2009). Prosodicheskie portrety govoryashchikh [Speakers prosodic portraits]. In A.A. Kibrik, & V.I. Podlesskaya (Eds.), *Corpus of spoken Russian "Night Dream Stories"* (pp. 478–487). Moscow: LRC Publishers. (In Russian).

Kibrik, A.A. (2010). Multimodalnaya lingvistika [Multimodal linguistics]. *Cognitive Studies, IV*, 134–152. (In Russian).

Kibrik, A.A. (2016). Jazyk kak on est [The language as it is]. In Yu.I. Aleksandrov, & K.V. Anokhin (Eds.), *The Seventh International Conference on Cognitive Science. Svetlogorsk, June 20–24. Abstracts* (pp. 674–676). Moscow: Institute of psychology RAS. (In Russian).

Kibrik, A.A. (2018a). Russkyi multikanalnyi diskurs. Chast I. Postanovka problemy [Russian multichannel discourse. Part I. Setting up the problem]. *Psikhologicheskii zhurnal, 39*(1), 70–80. (In Russian).

Kibrik, A.A. (2018b). Russkyi multikanalnyi diskurs. Chast II. Razrabotka korpusa i napravleniya issledovanii [Russian multichannel discourse. Part II. Corpus development and avenues of research]. *Psikhologicheskii zhurnal, 39*(2), 79–90. (In Russian).

Kibrik, A.A., & Fedorova, O.V. (2018). A "portrait" approach to multichannel discourse. In *Eleventh International Conference on Language Resources and Evaluation (LREC)* (pp. 1908–1912). Miyazaki, Japan.

Kibrik, A.A., & Podlesskaya, V.I. (Eds.). (2009). *Corpus of spoken Russian "Night Dream Stories".* Moscow: LRC Publishers. (In Russian).

Kipp, M. (2004). *Gesture generation by imitation: From human behavior to computer character animation (Doctoral dissertation).* Boca Raton, FL: Dissertations.com. Retrieved from https://pdfs.semanticscholar.org/ec7c/0f3f6eacc9acb6709b7d03094ef23984216a.pdf.

Kipp, M., Neff, M., & Albrecht, I. (2007). An annotation scheme for conversational gestures: How to economically capture timing and form. *Language Resources and Evaluation — Special Issue on Multimodal Corpora, 41*(3–4), 325–339. doi:10.1007/s10579-007-9053-5

Kita, S. (2003). *Pointing where language, culture, and cognition meet.* Mahwah, NJ: Lawrence Erlbaum.

Kita, S., van Gijn, I., & van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth, & M. Fröhlich (Eds.), *Gesture and sign language in human-computer interaction. International gesture workshop, Bielefeld, Germany, September 17–19, 1997, Proceedings* (pp. 23–35). Berlin: Springer. doi:10.1007/bfb0052986

Kopp, S., Bergmann, K., & Wachsmuth, I. (2008). Multimodal communication from multimodal thinking — towards an integrated model of speech and gesture production. *International Journal of Semantic Computing, 2*(1), 115–136. doi:10.1142/s1793351x08000361

Kress, G. (2002). The multimodal landscape of communication. *MedienJournal, 26*(4), 4–18. doi:10.24989/medienjournal.v26i4.411

Kreydlin, G.E. (2002). *Neverbalnaya semiotika: Yazyk tela i estestvennyi yazyk [Nonverbal semiotics. Body language and natural language].* Moscow: Novoe literaturnoe obozrenie. (In Russian).

Kreydlin, G.E. (2010). Mekhanizmy vzaimodeistviya verbalnykh i neverbalnykh edinits v dialoge [Mechanisms of interaction between verbal and non-verbal units in a dialogue]. In N.D. Arutunova (Ed.), *Logicheskii analiz yazyka. Mono-, dia-, polilog v raznykh yazykakh i kulturakh [Logical analysis of language. Monolog, dialog and polilog in various languages and cultures]* (pp. 173–183). Moscow: Indrik. (In Russian).

Ladewig, S.H., & Bressem, J. (2013). A linguistic perspective on the notation of gesture phases. In C. Müller, A. Cienki, E. Fricke, S.H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body — Language — Communication: An international handbook on multimodality in human interaction* (Vol. 2, pp. 1060–1078). Berlin: Mouton.

Lausberg, H., & Sloetjes, H. (2015). The revised NEUROGES-ELAN system. An objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture. *Behavior Research Methods, 48*(3), 973–993. doi:10.3758/s13428-015-0622-z

Litvinenko, A.O., Nikolaeva, J.V., & Kibrik, A.A. (2017a). Annotirovanie multikanalnogo diskursa: manualnye zhesty [Multichannel discourse annotation: manual gestures]. In E.V. Pechenkova, & M.V. Falikman (Eds.), *Kognitivnaya nauka v Moskve: novye issledovaniya. Materialy konferentsii [Cognitive science in Moscow: New research. Conference proceedings]*. Moscow: BukiVedi, IPPiP. (In Russian).

Litvinenko, A.O., Nikolaeva, J.V., & Kibrik, A.A. (2017b). Annotirovanie russkikh manualnykh zhestov: teoreticheskie i prakticheskie voprosy [Annotation of Russian manual gestures: Theoretical and practical issues]. In *Computational linguistics and intellectual technologies. Papers from the Annual International Conference "Dialogue"* (Vol. 16, pp. 255–268). Moscow: RGGU. (In Russian).

McNeill, D. (1979). *The conceptual basis of language.* Hillsdale, NJ: Erlbaum.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92*, 350–371.

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

McNeill, D. (Ed.). (2000). *Language and gesture.* Cambridge: Cambridge University Press.

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

McNeill, D. (2013). Gesture as a window onto mind and brain, and the relationship to the linguistic relativity and ontogenesis. In C. Müller, A. Cienki, E. Fricke, S.H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body — Language — Communication: An international handbook on multimodality in human interaction* (Vol. 1, pp. 28–54). Berlin: Mouton.

McNeill, D., & Duncan, S.D. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge: Cambridge University Press.

Müller, C. (1998). *Redebegleitende Gesten: Kulturgeschichte, Theorie, Sprachvergleich.* Berlin: Arno Spitz. (In German).

Müller, C., Bressem, J., & Ladewig, S.H. (2013). Towards a grammar of gestures: A form-based view. In C. Müller, A. Cienki, E. Fricke, S.H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body — Language — Communication: An international handbook on multimodality in human interaction* (Vol.1, pp.707–733). Berlin: Mouton.

Müller, C., Cienki, A., Fricke, E., Ladewig, S.H., McNeill, D., & Teßendorf, S. (Eds.). (2013). *Body — Language — Communication: An international handbook on multimodality in human interaction.* Berlin: Mouton. Müller, C. (2004). Forms and uses of the palm up open hand: A case of a gesture family? In C. Müller (Ed.), *The semantics and the pragmatics of everyday gestures* (pp.233–256). Berlin: Weidler.

Müller, C., Ladewig, S.H., & Bressem, J. (2013). Gestures and speech from a linguistic perspective: A new field and its history. In C. Müller, A. Cienki, E. Fricke, S.H. Ladewig, D. McNeill, & S. Teßendorf (Eds.), *Body — Language — Communication: An international handbook on multimodality in human interaction* (Vol.1, pp.55–82). Berlin: Mouton.

Müller, C., & Tag, S. (2010). The dynamics of metaphor: Foregrounding and activating metaphoricity in conversational interaction. *Cognitive Semiotics, 6*(Supplement), 85–120. doi:10.1515/cogsem.2010.6.spring2010.85

Nikolaeva, Y.V. (2004). Funktsionalnye i semanticheskie osobennosti illyustrativnykh zhestov v ustnoy rechi (na materiale russkogo yazyka) [Functional and semantic aspects of illustrative gestures in Russian spoken discourse]. *Voprosy Jazykoznanija (Topics in the study of language), 4*, 48–64. (In Russian).

Nikolaeva, Y.V. (2009). Segmentatsiya ustnogo narrativa i izobrazitelnye zhesty: kineticheskie priznaki granits i svyazej mezhdu segmentami diskursa [Segmentation of oral narrative discourse and illustrative gestures: visual clues as segment markers]. In *Computational linguistics and intellectual technologies. Papers from the Annual International Conference "Dialogue"* (Vol.8, pp.340–345). Moscow: RGGU. (In Russian).

Pike, K.L. (1967). *Language in relation to a unified theory of the structure of human behavior.* The Hague: Mouton.

Pogodin, A.L. (1913). *Yazyk kak tvorchestvo (psikhologicheskie i sotsialnye osnovy tvorchestva rechi): proiskhozhdenie yazyka [Language as a creative activity (psychological and social foundations of speech creativity): The origin of language].* Kharkov, Russia: V.A. Lezin. (In Russian).

Scherer, K.R., & Ekman, P. (1982). *Handbook of methods in nonverbal behavior research.* Cambridge: Cambridge University Press.

Scollon, R. (2006). Multimodality and the language of politics. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (Vol.9, pp.386–387). Oxford: Elsevier. doi:10.1016/b0-08-044854-2/00729-x

Trippel, T., Gibbon, D., Thies, A., Milde, J.T., Looks, K., Hell, B., & Gut, U. (2004). Cogest: A formal transcription system for conversational gesture. In *Proceedings of LREC 2004* (pp.2215–2218). Lisbon, Portugal.

Wagner, P., Kopp, S., & Malisz, Z. (2014). Gesture and speech in interaction: An overview. *Speech Communication, 57*, 209–232. doi:10.1016/j.specom.2013.09.008

Yakubinsky, L. (1986). O dialogicheskoj rechi [On dialogic speech]. In *Yakubinsky, L.P. Izbrannye raboty. Yazyk i ego funktsionirovanie [Selected works. Language and its functioning]* (pp.17–58). Moscow: Nauka. (In Russian) (Original work published 1923).

Żywiczyński, P., Wacewicz, S., & Orzechowski, S. (2017). Adaptors and the turn-taking mechanism: The distribution of adaptors relative to turn borders in dyadic conversation. *Interaction Studies, 18*(2), 276–298. doi:10.1075/is.18.2.07zyw

методы

Аннотирование движений рук в мультиканальном дискурсе: жесты, адапторы и мануальные позы

Алла Олеговна Литвиненко
Институт языкознания РАН, Москва, Россия

Андрей Александрович Кибрик
Институт языкознания РАН, Москва, Россия;
МГУ имени М.В. Ломоносова, Москва, Россия

Ольга Викторовна Федорова
МГУ имени М.В. Ломоносова, Москва, Россия;
Институт языкознания РАН, Москва, Россия

Юлия Владимировна Николаева
МГУ имени М.В. Ломоносова, Москва, Россия;
Институт языкознания РАН, Москва, Россия

Аннотация. Статья посвящена процессу аннотирования движений рук в контексте мультиканального дискурса (на материале русского мультиканального корпуса «Рассказы и разговоры о грушах»). В работе представлена как подробная аннотационная схема, разработанная в рамках проекта, так и сопутствующая процедура разметки. Движение рук в целом рассматривается как сложный поток действий и бездействия, который следует сегментировать и описывать в виде последовательности простых кинетических единиц, основываясь на их кинетических характеристиках (усилие, скорость, траектория, направление и т.д.). После завершения процесса сегментации полученные единицы или комбинации единиц могут быть интерпретированы как жесты или другие виды мануального поведения (периоды различного рода неподвижности, смены мануальной позы, адапторы и т.д.) на основании не только их кинетической формы, но и функционирования в дискурсе. Полученная в результате многоуровневая аннотация в дальнейшем может быть использована для изучения взаимозависимостей между мануальным поведением и другими каналами мультиканального дискурса.

Контактная информация: Алла Олеговна Литвиненко, allal1978@gmail.com; 125009, г. Москва, Б. Кисловский пер., д. 1 стр. 1, Институт языкознания РАН; Андрей Александрович Кибрик, aakibrik@gmail.com; Ольга Викторовна Федорова, olga.fedorova@msu.ru; Юлия Владимировна Николаева, julianikk@gmail.com.

Ключевые слова: мультимодальность, мультиканальный дискурс, мультиканальный корпус, жесты рук, движения рук, мануальные жесты, мануальная поза, аннотация жестов, аннотационная схема

Благодарности. Проект поддержан Российским научным фондом, грант № 14-18-03819.

Статья поступила в редакцию 15 октября 2017 г. Принята в печать 25 июня 2018 г.

Литература

Гришина Е. А. Русская жестикуляция с лингвистической точки зрения. Корпусные исследования. М.: ЯСК, 2017.

Доброгаев С. М. Учение о рефлексе в проблемах языковедения // Языковедение и материализм. Вып. 2 / Под ред. Н. Я. Марра. Ленинград: Соцэкгиз, 1931. С. 105–173.

Кибрик А. А. Просодические портреты говорящих // Рассказы о сновидениях: корпусное исследование устного русского дискурса / Под ред. А. А. Кибрика, В. И. Подлесской. М.: ЯСК, 2009. С. 478–487.

Кибрик А. А. Мультимодальная лингвистика // Когнитивные исследования. 2010. Т. IV. С. 134–152.

Кибрик А. А. Язык как он есть // Седьмая международная конференция по когнитивной науке. Тезисы докладов. Светлогорск, 20–24 июня 2016 г. / Под ред. Ю. И. Александрова, К. В. Анохина. М.: ИП РАН, 2016. С. 674–676.

Кибрик А. А. Русский мультиканальный дискурс. Часть I. Постановка проблемы // Психологический журнал. 2018a. Т. 39. № 1. С. 70–80.

Кибрик А. А. Русский мультиканальный дискурс. Часть II. Разработка корпуса и направления исследований // Психологический журнал. 2018b. Т. 39. № 2. С. 79–90.

Крейдлин Г. Е. Невербальная семиотика: язык тела и естественный язык. М.: Новое литературное обозрение, 2002.

Крейдлин Г. Е. Механизмы взаимодействия вербальных и невербальных единиц в диалоге // Логический анализ языка. Моно-, диа-, полилог в разных языках и культурах / Под ред. Н. Д. Арутюновой. М.: Индрик, 2010. С. 173–183.

Литвиненко А. О., Николаева Ю. В., Кибрик А. А. Аннотирование мультиканального дискурса: мануальные жесты // Когнитивная наука в Москве: новые исследования. Материалы конференции / Под ред. Е. В. Печенковой, М. В. Фаликман. М.: БукиВеди, ИППиП, 2017a.

Литвиненко А. О., Николаева Ю. В., Кибрик А. А. Аннотирование русских мануальных жестов: теоретические и практические вопросы // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной международной конференции «Диалог». М.: РГГУ, 2017b. Т. 16. С. 255–268.

Николаева Ю. В. Функциональные и семантические особенности иллюстративных жестов в устной речи (на материале русского языка) // Вопросы языкознания. 2004. Т. 4. С. 48–64.

Николаева Ю. В. Сегментация устного нарратива и изобразительные жесты: кинетические признаки границ и связей между сегментами дискурса // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной международной конференции «Диалог». М.: РГГУ, 2009. Т. 8. С. 340–345.

Погодин А. Л. Язык как творчество (психологические и социальные основы творчества речи): происхождение языка. Харьков: В. А. Лезин, 1913.

Рассказы о сновидениях: корпусное исследование устного русского дискурса / Под ред. А. А. Кибрика, В. И. Подлесской. М.: Языки славянских культур, 2009.

Федорова О. В., Кибрик А. А., Коротаев Н. А., Литвиненко А. О., Николаева Ю. В. Временная координация между жестовыми и речевыми единицами в мультимодальной коммуникации // Компьютерная лингвистика и интеллектуальные технологии. По материалам ежегодной международной конференции «Диалог». М.: РГГУ, 2016. Т. 15. С. 159–170.

Якубинский Л. П. О диалогической речи // Якубинский Л. П. Избранные работы. Язык и его функционирование. М.: Наука, 1923/1986. С. 17–58.

Abner N., Cooperrider K., Goldin-Meadow S. Gesture for linguists: A handy primer // Language and Linguistics Compass. 2015. Vol. 9. No. 11. P. 437–451. doi:10.1111/lnc3.12168

Allwood J., Loredana C., Jokinen K., Navarretta C., Paggio P. The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena // Language Resources and Evaluation. 2007. Vol. 41. No. 3-4. P. 273–287. doi:10.1007/s10579-007-9061-5

Argyle M. Bodily communication. New York: International Universities Press, 1975.

Birdwhistell R. L. Kinesics and context. Philadelphia: University of Pennsylvania Press, 1970. doi:10.9783/9780812201284

Body — Language — Communication: An international handbook on multimodality in human interaction / C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, S. Teßendorf (Eds.). Berlin: Mouton, 2013.

Bressem J., Ladewig S. H., Müller C. Linguistic Annotation System for Gestures (LASG) // Body — Language — Communication: An international handbook on multimodality in human interaction / C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, S. Teßendorf (Eds.). Berlin: Mouton, 2013. Vol. 2. P. 1098–1124.

Caldognetto E. M., Poggi I., Cosi P., Cavicchio F., Merola G. Multimodal score: An ANVIL based annotation scheme for multimodal audio-video analysis // Proceedings of the LREC-Workshop on multimodal corpora / J.-C. Martin, E. D. Os, P. Kühnlein, L. Boves, P. Paggio, R. Catizone (Eds.). Lisbon, Portugal: 2004. P. 29–33. URL: http://www.lrec-conf.org/proceedings/lrec2004/ws/ws10.pdf.

Cienki A. Metaphoric gestures and some of their relations to verbal metaphorical expressions // Discourse and cognition: Bridging the gap / J.-P. König (Ed.). Stanford, CA: Center for the Study of Language and Information, 1998a. P. 189–204.

Cienki A. Straight: An image schema and its metaphorical extensions // Cognitive Linguistics. 1998b. Vol. 9. No. 2. P. 107–150. doi:10.1515/cogl.1998.9.2.107

Cienki A., Müller C. Metaphor, gesture and thought // Cambridge handbook of metaphor and thought / R. W. Gibbs (Ed.). Cambridge: Cambridge University Press, 2008. P. 483–501.

Duncan S., Rohlfing K., Loehr D. Multimodal annotation tools // Body — language — communication: An international handbook on multimodality in human interaction / C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, S. Teßendorf (Eds.). Berlin: Mouton, 2013. Vol. 1. P. 1015–1022.

Dutsch D. The body in rhetorical delivery and in theatre — An overview of classical works // Body — language — communication: An international handbook on multimodality in human interaction / C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, S. Teßendorf (Eds.). Berlin: Mouton, 2013. Vol. 1. P. 329–343.

Efron D. Gesture, race and culture. Paris: Mouton, 1972.

Ekman F., Friesen W. V. The repertoire of nonverbal behavior: Categories, origins, usage, and coding // Semiotica. 1969. Vol. 1. No. 1. P. 49–98. doi:10.1515/semi.1969.1.1.49

Fundamentals of nonverbal behavior / R. S. Feldmann, B. Rimé (Eds.). Cambridge: Cambridge University Press, 1991.

Graf F. Gestures and conventions: The gestures of Roman actors and orators // A cultural history of gesture / J. Bremmer, H. Roodenburg (Eds.). Cambridge: Polity Press, 1994. P. 36–58.

Handbook of linguistic annotation / N. Ide, J. Pustejovsky (Eds.). Netherlands: Springer, 2017. doi:10.1007/978-94-024-0881-2

Handbook of multimodal and spoken dialogue systems: Resources, terminology and product evaluation / D. Gibbon, I. Mertins, R. K. Moore (Eds.). Berlin: Springer, 2000.

Karpiński M., Jarmołowicz-Nowikow E., Czoska A. Gesture annotation scheme development and application for entrainment analysis in task-oriented dialogues in diverse cultures // Proceedings of GESPIN 2015 Conference / G. Ferré, M. Tutton (Eds.). Nantes, France: 2015. P. 161–166.

Kendon A. Some relationships between body motion and speech: An analysis of an example // Studies in dyadic communication / A. W. Siegman, B. Pope (Eds.). New York: Elsevier, 1972. P. 177–210.

Kendon A. Gesticulation and speech: Two aspects of the process of utterance // The relationship of verbal and nonverbal communication / M. R. Key (Ed.). The Hague: De Gruyter Mouton, 1980. P. 207–227.

Kendon A. Gesture: Visible action as utterance. New York: Cambridge University Press, 2004. doi:10.1017/cbo9780511807572

Kibrik A. A., Fedorova O. V. A "portrait" approach to multichannel discourse // Eleventh International Conference on Language Resources and Evaluation (LREC). Miyazaki, Japan: 2018. P. 1908–1912.

Kipp M. Gesture generation by imitation: From human behavior to computer character animation (Doctoral dissertation). Boca Raton, FL: Dissertations.com, 2004. URL: https://pdfs.semanticscholar.org/ec7c/0f3f6eacc9acb6709b7d03094ef23984216a.pdf.

Kipp M., Neff M., Albrecht I. An annotation scheme for conversational gestures: How to economically capture timing and form // Language Resources and Evaluation — Special Issue on Multimodal Corpora. 2007. Vol. 41. No. 3-4. P. 325–339. doi:10.1007/s10579-007-9053-5

Kita S. Pointing where language, culture, and cognition meet. Mahwah, NJ: Lawrence Erlbaum, 2003.

Kita S., van Gijn I., van der Hulst H. Movement phases in signs and co-speech gestures, and their transcription by human coders // Gesture and sign language in human-computer interaction. International gesture workshop, Bielefeld, Germany, September 17–19, 1997, Proceedings / I. Wachsmuth, M. Fröhlich (Eds.). Berlin: Springer, 1998. P. 23–35. doi:10.1007/bfb0052986

Kopp S., Bergmann K., Wachsmuth I. Multimodal communication from multimodal thinking — towards an integrated model of speech and gesture production // International Journal of Semantic Computing. 2008. Vol. 2. No. 1. P. 115–136. doi:10.1142/s1793351x080 00361

Kress G. The multimodal landscape of communication // MedienJournal. 2002. Vol. 26. No. 4. P. 4–18. doi:10.24989/medienjournal.v26i4.411

Ladewig S. H., Bressem J. A linguistic perspective on the notation of gesture phases // Body — Language — Communication: An international handbook on multimodality in human interaction / C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, S. Teßendorf (Eds.). Berlin: Mouton, 2013. Vol. 2. P. 1060–1078.

Lausberg H., Sloetjes H. The revised NEUROGES-ELAN system. An objective and reliable interdisciplinary analysis tool for nonverbal behavior and gesture // Behavior Research Methods. 2015. Vol. 48. No. 3. P. 973–993. doi:10.3758/s13428-015-0622-z

Language and gesture / D. McNeill (Ed.). Cambridge: Cambridge University Press, 2000.

McNeill D. The conceptual basis of language. Hillsdale, NJ: Erlbaum, 1979.

McNeill D. So you think gestures are nonverbal? // Psychological Review. 1985. Vol. 92. P. 350–371.

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

McNeill D. Gesture and thought. Chicago: University of Chicago Press, 2005.

McNeill D. Gesture as a window onto mind and brain, and the relationship to the linguistic relativity and ontogenesis // Body — Language — Communication: An international handbook on multimodality in human interaction / C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, S. Teßendorf (Eds.). Berlin: Mouton, 2013. Vol. 1. P. 28–54.

McNeill D., Duncan S. D. Growth points in thinking-for-speaking // Language and gesture / D. McNeill (Ed.). Cambridge: Cambridge University Press, 2000. P. 141–161.

Metaphor and gesture / A. Cienki, C. Müller (Eds.). Amsterdam: John Benjamins Publishing Company, 2008. doi:10.1075/gs.3

Müller C. Redebegleitende Gesten: Kulturgeschichte, Theorie, Sprachvergleich. Berlin: Arno Spitz, 1998.

Müller C. Forms and uses of the palm up open hand: A case of a gesture family? // The semantics and the pragmatics of everyday gestures / C. Müller (Ed.). Berlin: Weidler, 2004. P. 233–256.

Müller C., Bressem J., Ladewig S. H. Towards a grammar of gestures: A form-based view // Body — Language — Communication: An international handbook on multimodality in human interaction / C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, S. Teßendorf (Eds.). Berlin: Mouton, 2013. Vol. 1. P. 707–733.

Müller C., Ladewig S. H., Bressem J. Gestures and speech from a linguistic perspective: A new field and its history // Body — Language — Communication: An international handbook on multimodality in human interaction / C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, S. Teßendorf (Eds.). Berlin: Mouton, 2013. Vol. 1. P. 55–82.

Müller C., Tag S. The dynamics of metaphor: Foregrounding and activating metaphoricity in conversational interaction // Cognitive Semiotics. 2010. Vol. 6. No. Supplement. P. 85–120. doi:10.1515/cogsem.2010.6.spring2010.85

Multimodality in language and speech systems / B. Granström, D. House, I. Karlsson (Eds.). Dordrecht: Kluwer, 2002.

Nonverbal communication / R. A. Hinde (Ed.). Cambridge: Cambridge University Press, 1972.

Pike K. L. Language in relation to a unified theory of the structure of human behavior. The Hague: Mouton, 1967.

Scherer K. R., Ekman P. Handbook of methods in nonverbal behavior research. Cambridge: Cambridge University Press, 1982.

Scollon R. Multimodality and the language of politics // Encyclopedia of language & linguistics / K. Brown (Ed.). Oxford: Elsevier, 2006. Vol. 9. P. 386–387. doi:10.1016/b0-08-044854-2/00729-x

Trippel T., Gibbon D., Thies A., Milde J. T., Looks K., Hell B., Gut U. Cogest: A formal transcription system for conversational gesture // Proceedings of LREC 2004. Lisbon, Portugal: 2004. P. 2215–2218.

The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production / W. Chafe (Ed.). Norwood, NJ: Ablex, 1980.

Wagner P., Kopp S., Malisz Z. Gesture and speech in interaction: An overview // Speech Communication. 2014. Vol. 57. P. 209–232. doi:10.1016/j.specom.2013.09.008

Żywiczyński P., Wacewicz S., Orzechowski S. Adaptors and the turn-taking mechanism: The distribution of adaptors relative to turn borders in dyadic conversation // Interaction Studies. 2017. Vol. 18. No. 2. P. 276–298. doi:10.1075/is.18.2.07zyw