Multimodal Discourse: In Search of Units

Andrej A. Kibrik (aakibrik@gmail.com)
Institute of Linguistics RAS and Lomonosov Moscow State University
B. Kislovskij per. 1, Moscow, 125009, Russia

Olga V. Fedorova (olga.fedorova@msu.ru)
Lomonosov Moscow State University, Russian Academy of National Economy and Public Administration, and Institute of Linguistics RAS
Leninskie Gory 1, Moscow, 119899, Russia

Julia V. Nikolaeva (julianikk@gmail.com)
Lomonosov Moscow State University and Institute of Linguistics RAS
Leninskie Gory 1, Moscow, 119899, Russia

Abstract

Human communication is inherently multimodal. In this study we focus on three channels of spoken discourse: the verbal component, prosody, and gesticulation. We address the question of units that can be identified within these components and in spoken multimodal discourse as a whole. The basic unit of the verbal channel is the clause, reporting an event or a state. A set of prosodic criteria help to define elementary discourse units, that is prosodic units serving as quanta of discourse production. The gestural channel consists of individual gestures, each defined by a set of features. Elementary discourse units are strongly coordinated with both clauses and gestures and can thus be considered basic units of multimodal discourse. Larger units can also be identified, such as prosodic sentences and series of gestures that again demonstrate coordination. By identifying units of natural discourse, coordinated across various channels, we make a step towards multimodal linguistics.

Keywords: discourse structure; multimodal discourse; clause; prosody; gesture; elementary discourse unit; sentence.

1. Introduction

In modern linguistics, as well as in other domains of cognitive science, there is a growing understanding that human communication is inherently multimodal. When we communicate orally, we not only produce chains of words, but also intonate, gesticulate, interact with eye gaze, etc. (Gibbon et al. eds., 2000; Kress, 2002; Hugot, 2007; So et al., 2009; Loehr, 2012; Ford, Fox, & Thompson, 2013; Goldin-Meadow, 2014, inter alia). A research program of multimodal linguistics is gradually evolving (Kibrik, 2010; Kress, 2010; Knight, 2011; Adolphs & Carter, 2013; Müller et al. eds., 2014) that treats the verbal structure on a par with non-verbal devices. Among non-verbal devices, sometimes only kinetic-visual behaviors are considered. But we find it very important to identify prosody (see e.g. Kodzasov, 2009), that is non-segmental aspects of the vocal signal, as a distinct communication channel.

Kibrik and Molchanova (2013) considered three communication channels employed in multimodal discourse: the verbal component, prosody, and kinetic-visual behavior. They found that all three channels play an important (and comparable) role in the overall process of conveying a message from a speaker to an addressee.

In this study we focus on three components of spoken discourse: the verbal component, prosody, and gesticulation. These components can be viewed separately to an extent but they are all interwoven in natural communication. As any human behavior, multimodal discourse has structure. If so, what are its *units*? We discuss the basic units found within the three channels considered separately (sections 2–4), and proceed with suggestions on coordinated basic units of multimodal discourse (section 5). In section 6 we discuss larger, more complex units of spoken discourse, and offer conclusions in section 7. This study is based on a corpus of Russian discourse, but some English examples are cited below for the ease of exposition.

2. The verbal channel

The verbal component of discourse largely consists of reporting events and states (Chafe, 1994). Languages have developed a universal syntactic structure for packaging events and states: *the clause*. Each clause reports an event or a state, along with their participants, or referents. For example, the minimal narrative *Veni, vidi, vici* consists of three events, each reported with a clause consisting of a single word: a verbal predicate, encoding in its inflection the subject participant. Consider a natural spoken example (from text SBC032 of the Santa Barbara corpus of spoken American English, see www.linguistics.ucsb.edu/research/santa-barbara-corpus), consisting of two clauses, each reporting an event:

And then I was forced out,
because I failed a promotion to commander!

Clauses may report events of various complexity and with various amount of detail, and they may include additional elements, especially connectors indicating the semantic relationships between clauses, such as *and then* or *because* in the example above. In various theories of discourse structure (e.g. Mann & Thompson, 1988; Carlson, Marcu & Okurowski, 2003; Wolf & Gibson, 2005) clauses are organized in a hierarchical network of nodes connected with discourse-semantic relations. Groups of clauses are often organized into syntactic units known as sentences, with the

links between clauses being tight to various degrees, see e.g. Givón, 2009; Laury & Ono, 2014.

3. The prosodic channel

Prosody directly encodes the dynamics of how thought unfolds during discourse production. There is a set of prosodic phenomena, including pausing, intonation contours, tempo patterns, loudness patterns, and accent placement, that converge in a unit of speech variously dubbed syntagm (Shcherba, 1955), intonation unit (Chafe, 1994), prosodic unit (e.g. Genetti & Slater, 2004), etc. We prefer the term *elementary discourse unit* (EDU), see Kibrik & Podlesskaya eds., 2009; Kibrik, 2011. EDUs are building blocks, or quanta, of spoken discourse. They are coordinated with breathing: one EDU is normally produced during an exhalation, and boundary pauses coincide with an inhalation. EDUs are linguistic representations of successive cognitive states, termed foci of consciousness in Chafe, 1994. EDU identification in speech is a procedure based on expert assessment. Well trained transcribers of spoken discourse strongly agree in EDU segmentation.

A remarkable fact about EDUs is their significant correlation with clauses. In a number of studies of various languages (Chafe, 1994 for English; Matsumoto, 2003 for Japanese; Genetti & Slater, 2004 for Newari; Wouk, 2008 for Sasak; Kibrik & Podlesskaya eds., 2009 for Russian, inter alia) the share of EDUs coinciding with clauses was found to vary between 50% and 70%. In the following example (from the same text; see spokencorpora.ru/showtranshelp.py for transcription conventions) lines #12 and #14 are clausal EDUs, while line #13 is a parcellated adjunct semantically belonging to the preceding clause but expressed with a subclausal EDU:

00:22.9	12	····(1.0) /**My** friend stood up /**behind his** **desk**,
00:26.0	13	··(0.2) in his /**fu-ull** \f-four \–stripes,
00:28.0	14	and \said:

Properties of EDUs have clear parallels in goal-directed behavior of non-human mammals. The exploratory movement of rodents in a new environment is organized in quanta (runs); runs are identified through initial acceleration and final deceleration, they are targeted at an informationally rich goal (analog of primary accent in discourse segments), they are separated by periods of freezing, etc. (see e.g. Kafkafi et al., 2001, Cherepov & Anokhin, 2008). These similarities suggest that the quantized structure of discourse and its specific prosodic aspects have deep behavioral, neurocognitive, and evolutionary roots.

4. The gestural channel

In the human kinetic-visual behavior, manual gesticulation plays a particularly important role. There are two widely accepted polar kinds of manual gestures. First, "emblems" (Efron, 1941/1972; Ekman & Friesen, 1969), also named "autonomous" (Kendon, 1983), or "quotable" gestures (Kendon, 1986), are manual signs with fixed form and relatively fixed meaning, widely shared by a given linguistic community. Second, "illustrative" or "spontaneous" (McNeill, 1992) gesticulation, also called "co-speech" or "speech-associated" gestures, (for an overview, see Kendon, 2004) consists of less conventional and more context-sensitive gestures. Illustrative gestures are incomparably more common in natural discourse (Nikolaeva, 2013). It is well established that illustrative gestures substantially participate in conveying a message from the speaker to the addressee (Cassell et al., 1999; Melinger & Levelt, 2004; Hostetter, 2011; Hall & Knapp eds., 2013). We posit the following major kinds of illustrative gestures: depictive ("iconic" + "metaphoric" in McNeill, 1992; "descriptive" in Kendon, 2004), metadiscursive ("pragmatic" in Payrató & Tessendorf, 2014), pointing ("deictic" in McNeill, 1992), and beats ("batons" in Efron, 1941/1972).

This study is primarily limited to depictive gestures, because they are particularly frequent in our corpus (59%) and contribute semantically (either in a redundant or in a complementary fashion) to the propositional content conveyed in the corresponding verbal component. Depictive gestures represent objects or act out events/states. Consider two initial EDUs from ex. 3 in the Appendix. EDU #17 *tam derevo* 'there is a tree' is accompanied by the following depictive gesture: the right hand palm faces down, fingers are half curled and widely spaced, the right hand moves up in front of the speaker's face, the left hand palm faces up at the chest level, with fingers half curled. EDU #18 *k derevu prižata lestnica* 'to the tree a ladder is pressed' is accompanied by two identical depictive gestures, the first of which cooccurs with the initial pause, and the second with the word *lestnica* 'ladder': the right hand faces the listener, fingers half curled, moves along a slanted line from the center right and down, the left hand remains at the chest level, faces up, with half curled fingers. Our dataset also includes metadiscursive gestures (see ex. 4) that demonstrate more recurrent properties compared to the depictive gestures, but still are a lot more variable than the emblems.

We use the term *gesture* to refer to the basic unit of co-speech gesticulation. Gesture is a communicatively significant manual movement, characterized by a unified pattern that includes trajectory, handshape and position, as well as other features. According to Kendon (1980, 2004) and McNeill (1992), the gestural structure includes units, phrases, and phases. The gesture unit (G-unit) "begins the moment the limb begins to move and ends when it has reached a rest position again" (McNeill, 1992: 83). A G-phrase consists of the following phases: a non-obligatory preparation, a non-obligatory pre-stroke hold, an obligatory stroke, a non-obligatory post-stroke hold, while a retraction (or recovery) is a part of G-unit (Kendon, 1980, 2004). There can be one or more G-phrases in a G-unit. Our understanding of "gesture" is close to G-phrase, but unlike the latter a gesture may include (though not obligatorily) a

retraction phase. In other words, a gesture ends either when the rest position is resumed or when another gesture begins.

5. Coordination of basic units

A key issue in the research program of multimodal linguistics is the question of coordination between the verbal, prosodic, and gestural channels. If we see discourse as a fundamentally multimodal process, we need to identify a unified basic unit of this process. A possible approach is to select one of the already established units as the basic one. As has been shown in section 3, EDU is a good candidate for this role, particularly because of its close connection with the quanta of non-linguistic behavior. Also note that prosody, serving as the source of criteria for EDU identification, is the ontogenetically earliest communication channel (see e.g. Crystal 1979, Blake 2000), preceding not only segmental speech but also gesticulation. We already know that EDUs strongly correlate with clauses. How do EDUs relate to gestures?

We explored this question on the basis of 14 Russian retellings of the Pear Film (Chafe, 1980), videorecorded and transcribed. Transcription, including temporal dynamics, pausing, annotation of EDUs, and other prosodic phenomena, was done with the help of the PRAAT program (www.fon.hum.uva.nl/praat). Gesture annotation was done in the ELAN program (www.lat-mpi.eu/tools/elan). A requirement observed in this work was independent annotation of clauses, EDUs, and gestures. The corpus consists of 37 minutes of videorecording, 1232 EDUs, and 705 gestures (414 of which are depictive).

We found that a prototypical EDU cooccurs with one depictive gesture, about 20% of EDUs cooccur with more than one gesture, see ex. 1: 9[1]; ex. 3: 18, 19 in the Appendix. This reminds of the well-known generalization: "A general rule is one gesture, one clause <...> some clauses have more than one gesture and some gestures cover more than one clause" (McNeill, 1992: 94).

Typically (approx. 90%), a depictive gesture falls within the temporal bounds of a single EDU. We also found that depictive gestures often (approx. 60%) cooccur with a whole EDU (ex. 3: 20, 21). When a gesture is shorter than the corresponding EDU, it is often temporally coordinated with the later part of the EDU, that is the typical locus of rhematic information (ex. 3: 17, 18). We can thus specify McNeill's claim, positing not just the relatedness of gestures to the vocal part of a message, but also a high degree of temporal coordination between gestures and EDUs.

6. Coordination of larger units

EDU being the basic unit of talk, there are higher order units, too. In particular, in various languages spoken correlates of written sentence have been found (Chafe, 1994, Genetti & Slater, 2004, Kibrik, 2008, 2011). Spoken sentence is established on the basis of prosodic criteria, such as target tone level (so-called period intonation), and functions as a structural unit larger than an EDU but shorter than an episode. Cognitively, in Chafe's (1994: 148) terms, a sentence is verbalization of a "superfocus of consciousness". Is there a correlate of prosodic sentence in the gestural channel?

By default, co-speech gestures are independent of each other. However, McNeill et al. (2001) discovered what can be called *gesture assimilation*. Some gestures are organized in series with repeated properties. McNeill et al. (2001) differentiate between the following two phenomena:

- in so-called *catchments*, formal properties of gestures (such as location in space, handshape and trajectory, etc.) may be repeated from one gesture to another, formal similarity conveying certain repeated semantic features;
- in gesture *inertia*, formal properties are shared in a series of gestures, but no semantic relatedness may be observed.

Fig. 1 illustrates four gestures, two of which accompany EDU #9 and two accompany EDUs #10–11 in example 1. These gestures depict:

- Fig. 1a — the abundance of pears;
- Fig. 1b — self-directed movement, putting pears into the apron;
- Fig. 1c — downward movement with the pears;
- Fig. 1d — outward movement of the pears, corresponding to the verb *vykladyval* 'was taking out'.

The uniform hand configuration with the slightly curled fingers depicts pears in the gardener's hands (Nikolaeva, 2013). This is an instance of catchment.

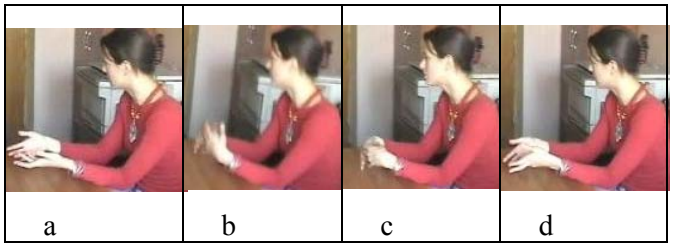

Figure 1. Catchment.

Catchments as series of gestures are possible candidates for gestural correlates of prosodic sentences. Out data includes about 150 instances of catchments. They split into two groups of equal size. In the first group, each gesture falls within the bounds of the corresponding EDU, and the boundaries of the gesture series coincide with the boundaries of the prosodic sentence, cf. ex. 3. These kinds of instances apparently support the coordination between the prosodic and gesture units. In the second group, a gesture series is coordinated with a certain part of a prosodic sentence (ex. 1; ex. 4). Looking into the second kind of instances more closely, it turns out that they mark the most informationally rich parts of sentences (ex. 4: 75, 76), whereas some other EDUs of the sentence are accompanied by independent gestures — ex. 4: 71 demonstrates two

[1] Here and below, the number after the colon refers to the EDU number within the given example. Examples are provided in the Appendix.

metadiscursive gestures "palm up, open hand" illustrating the process of information transfer (conduit metaphor). Overall, catchments are coordinated with prosodic sentences. Given that catchments are a special case of G-units (see section 4 above), we hypothesize that coordination with prosodic sentences can be extended to G-units in general. This latter point requires further investigation.

Turning to gesture inertia, consider Fig. 2 that illustrates three gestures, accompanying the three EDUs in example 2. These gestures depict:

- Fig. 2a — the sudden halt;
- Fig. 2b — the falling bicycle;
- Fig. 2c — the falling hat (a gesture similar in configuration and trajectory to the previous one but with a larger amplitude).

In this case gesture assimilation is only formal, in contrast to catchments, in which similar gestures contain shared semantic features.

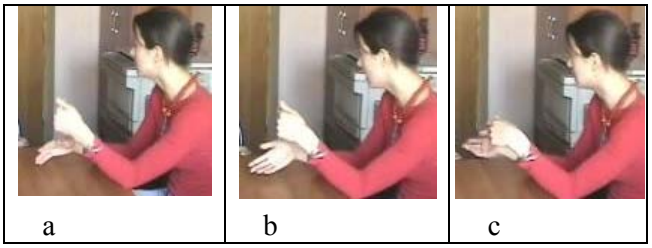

Figure 2. Gesture inertia.

In a first approximation, infrequent instances of gesture inertia appear to be coordinated with the unit of discourse known as *episode* (van Dijk, 1981). We are not aware of robust methods of episode identification, either semantic or prosodic, so we have identified episodes intuitively. Example 2 illustrates a typical situation, in which gesture inertia is a series of gestures bridging a sentence boundary and joining a group of EDUs that qualifies as a small episode.

7. Conclusion

We have found that the basic units of the three channels of multimodal discourse — verbal, prosodic, and gestural — are coordinated between each other. More specifically, the prosodically identified elementary discourse unit can be shown to be coordinated with the verbal channel and with the gestural channel. We have chosen the prosodic unit as the central one because it is established on the basis of general behavioral criteria. Unlike gesture, prosody is always present in talk. In the studies reported in Kibrik & Molchanova, 2013 it turned out difficult to individually separate the verbal channel, as talking inevitably involves prosody.

Apart from basic units, we have also discussed larger units of spoken discourse. It appears that prosodically identified sentences and episodes are coordinated with gesture series known as catchment and inertia.

Even though we are looking for structure and units in discourse, those should not be understood in the sense of absolute discreteness. Units, or quanta, do exist, but the boundaries between them are typically less than discrete. There are many instances of outliers and hybrids that complicate crisp and neat unit boundaries. As is shown by Kibrik (2015), this property of discourse structure is common with other levels of language, as well as cognition in general. Non-discrete effects abound both between syntagmatic units and between paradigmatic types. This resonates with McNeill's (2005) suggestion that gestures may be classified into dimensions rather than discrete categories, and a given gesture may, for instance, combine features of a depictive and a pointing gesture.

Acknowledgment

This study is supported by the Russian Science Foundation (grant #14-18-03819).

References

Adolphs, S., & Carter, R. (2013). *Spoken corpus linguistics: From monomodal to multimodal.* N.-Y.: Routledge.

Blake, J. (2000). *Routes to Child Language: Evolutionary and Developmental Precursors*. Cambridge: CUP.

Carlson, L., Marcu, D., & Okurowski, M. E. (2003). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. In J. van Kuppevelt & R. Smith (Eds.), *Current and new directions in discourse and dialogue*. Dordrecht: Kluwer.

Cassell, J., McNeill, D., & McCullough, K. E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and non-linguistic information. *Pragmatics and Cognition*, 7(1), 1–33.

Chafe, W. (1994). *Discourse, consciousness, and time. The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.

Chafe, W. (Ed.) (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, N.J.: Ablex.

Cherepov, A., & Anokhin, K. (2008). Development of automatic analysis and recognition of mouse behavior by segmentation and t-pattern method using video tracking. *Proceedings of Measuring Behavior 2008* (pp. 253–254). Maastricht, The Netherlands, August 26–29, 2008.

Crystal. D. (1979). Prosodic development. In P.J. Fletcher & M.A. Garman (Eds.), *Language acquisition* (pp. 33–48). Cambridge: CUP. (2nd edn., 1986, pp. 174–97.)

Efron, D. (1941/1972). *Gestures, race and culture*. The Hague: Mouton.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, 1, 49–98.

Ford, C. E., Thompson, S. A., & Drake, V. (2012). Bodily-visual practices and turn continuation. *Discourse Processes*, 49(3-4), 192–212.

Ford, C. E., Fox, B., & Thompson, S. A. (2013). Units and/or action trajectories? The language of grammatical categories and the language of social action. In B. Szczepek Reed & G. Raymond (Eds.), *Units of talk – Units of action*. Amsterdam: Benjamins.

Genetti, C., & Slater, K. (2004). An analysis of syntax and prosody interactions in a Dolakhā Newar: Rendition of the Mahābhārata (with appendices and sound files). *Himalayan Linguistics*, 3, 1–91.

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

Givón, T. (2009). Multiple routes to clause union: The diachrony of complex verb phrases. In T. Givón & M. Shibatani (Eds.), *Syntactic complexity: Diachrony, acquisition, neuro-cognition, evolution*. Amsterdam: Benjamins.

Goldin-Meadow, S. (2014). Widening the lens: What the manual modality reveals about language, learning, and cognition. *Philosophical Transactions of the Royal society*, 369.

Hall, J. A., & Knapp, M. L. (Eds.) (2013). *Handbooks of communication science: Nonverbal communication*. Berlin: De Gruyter Mouton.

Hostetter, A. B. (2011).When do gestures communicate? A meta-analysis. *Psychological Bulletin*, 137(2), 297–315.

Hugot, V. (2007). *Eye gaze analysis in human-human interactions.* Master of science thesis. Stockholm, Sweden.

Kafkafi, N., Mayo, C. L., Drai, D., Golani, D., & Elmer, G. I. (2001). Natural segmentation of the locomotor behavior of drug-induced rats in a photobeam cage. *Journal of Neuroscience Methods*, 109, 111–121.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relation between verbal and nonverbal communication*. The Hague: Mouton.

Kendon, A. (1983). Gesture and speech. How they interact. In J. M. Wiemann & R. P. Harrison (Eds.), *Nonverbal Interaction.* Beverly Hills: Sage.

Kendon, A. (1986). Some reasons for studying gesture. *Semiotica*, 62, 3–28.

Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press.

Kibrik, A. A. (2008). Est' li predloženie v ustnoj reči [Is there a sentence in spoken speech]. In A. V. Arxipov, L. V. Zaxarov, A. A. Kibrik et al. (Eds.), *Fonetika i nefonetika. K 70-letiju Sandro V. Kodzasova* [Phonetics and non-phonetics. Festschrift for 70 of Sandro V. Kodzasov]. Moscow: Jazyki slavjanskix kul'tur.

Kibrik, A. A. (2010). Mul'timodal'naja lingvistika [Multimodal linguistics]. In Yu. I. Aleksandrov, V. D. Solov'jev (Eds.), *Kognitivnyje issledovanija* [Cognitive studies], IV. Moscow: Institute of psychology.

Kibrik, A. A. (2011). Cognitive discourse analysis: Local discourse structure. In M. Grygiel and L. A. Janda (Eds.), *Slavic linguistics in a cognitive framework*. N.Y.: Peter Lang.

Kibrik, A. A. (2015). The problem of non-discreteness and spoken discourse structure. *Computational Linguistics and Intelligent Technologies*, 14, vol. 1, 225–233.

Kibrik, A. A., & Podlesskaja, V. I. (Eds.) (2009). Rasskazy o snovidenijax: Korpusnoe issledovanie ustnogo russkogo diskursa [Night Dream Stories: A corpus study of spoken Russian discourse]. Moscow: Jazyki slavjanskix kul'tur.

Kibrik, A. A., & Molchanova, N. B. (2013). Channels of multimodal communication: Relative contributions to discourse understanding. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 2704–2709). Austin, TX: Cognitive Science Society.

Knight, D. (2011). *Multimodality and active listenership: A corpus approach.* London: Bloomsbury.

Kodzasov, S. V. (2009). Issledovanija v oblasti russkoj prosodii [Studies in the field of Russian prosody]. Moscow: Jazyki slavjanskix kul'tur.

Kress, G. (2002). The multimodal landscape of communication. *Medien Journal*, 4, 4–19.

Kress, G. (2010). *Multimodality: A social semiotic approach to communication*. London: Routledge Falmer.

Laury, R., & Ono, T. (2014). The limits of grammar: Clause combining in Finnish and Japanese conversation. *Pragmatics*, 24(3), 561–592.

Loehr, D. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, 3(1), 71–89.

Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3), 243–281.

Matsumoto, K. (2003). *Intonation units in Japanese conversation.* Amsterdam: John Benjamins.

McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.

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

McNeill, D., Quek, F., McCullough, K.-E., Duncan, S., Furuyama, N., Bryll, R., Ma, X.-F., & Ansari, R. (2001). Catchments, prosody, and discourse. *Gesture*, 1, 9–33.

Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture*, 4, 119–141.

Müller, C., Fricke, E., Cienki, A., McNeill, D. (Eds.) (2014). *Body – Language – Communication*. Berlin: Mouton de Gruyter.

Nikolaeva, Ju. V. (2013). *Illustrativnyje žesty v russkom diskurse* [Gesticulation in Russian discourse]. Diss. cand. philol. science. Moscow, Russia.

Payrató, L., & Tessendorf, S. (2014). Pragmatic gestures. In Müller, C., Fricke, E., Cienki, A., McNeill, D. (Eds.) *Body – Language – Communication*. Berlin: Mouton de Gruyter.

Shcherba, L. V. (1955). *Fonetika francuzskogo jazyka* [French phonetics]. Moscow: Izdatel'stvo literatury na inostrannyx jazykax.

So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science*, 33, 115–125.

van Dijk, T. (1981). Episodes as units of discourse analysis. In D. Tannen (Ed.), *Analyzing discourse: Text and talk*. Georgetown: Georgetown University Press.

Wolf, F., & Gibson, E. (2005). Representing discourse coherence: A corpus-based study. *Computational Linguistics*, 31(2), 249–287.

Wouk, F. (2008). The syntax of intonation units in Sasak. *Studies in Language*, 32, 137–162.

Appendix. Examples[2]

1	**time, *s***	**EDU #**	**Transcript**	**gesture type**
	00:16	7	[···(0.5) u nego stojalo tri korziny] s grušami, ‘[he had three baskets] with pears,	depictive
	00:18	8	i on {[podnimalsja] na lestnicu, and he {[was climbing up] the ladder,	depictive
	00:20	9	[··(0.3) sobiral eti gruši] v [əə(0.3) fartuk], [was collecting these pears] into [the apron],	depictive, depictive
	00:22	10	[··(0.2) spu][skalsja [was climbing] [down	depictive
	00:23	11	i vykladyval]} eti gruši v korzinu. and was taking out]} these pears into the basket.’	depictive

2	**time, *s***	**EDU #**	**Transcript**	**gesture type**
	00:59	29	<[···(0.8) i ɯɯɯ(0.8) ego velosiped] vre= vrezalsja v kamen'. ‘<[and his bicycle] ran into a rock.	depictive
	01:02	30	··(0.4) [on] upal, [he] fell down,	depictive
	01:04	31	···(0.7) [s nego sletela] šljapa.> his hat fell off. (lit. [from him fell] the hat.>)’	depictive

3	**time, *s***	**EDU #**	**Transcript**	**gesture type**
	00:29	17	tam {[derevo], ‘there is {[a tree],	depictive
	00:30	18	[····(1.2)] k derevu prižata [lestnica], [ ] to the tree [a ladder] is pressed,	depictive, depictive
	00:32	19	[i vnizu lestnicy stojat] [tri korzinki], [and under the ladder there are] [three baskets],	depictive, depictive
	00:34	20	[dve iz kotoryx polnyje gruš], [two of which are full of pears],	depictive
	00:36	21	[a vtoraja pustaja].} [and the second one is empty].}’	depictive

4	**time, *s***	**EDU #**	**Transcript**	**gesture type**
	02:24	71	···(0.6) ə̈ə̈ə(0.6) [ə̈ə̈ə(0.7) ə̈ə̈ə(0.6)] [ə̈ə̈ə(0.8) i vdrug] pered nim ··(0.2) okazyvajutsja ··(0.1) neskol’ko ··(0.1) parnej, ‘[ ] [and suddenly] in front of him show up a few guys,	meta, meta
	02:28	72	···(0.6) troe, three of them,	
	02:29	73	···(0.5) niotkuda, from nowhere,	
	02:30	74	neponjatno otkuda vzjavšixsja, not clear where they are coming from,	
	02:31	75	i oni {[··(0.2) načinajut sobirat’ eti gruši], and they {[begin picking up these pears],	depictive
	02:33	76	i [pomogat’ emu skladyvat’]} v korzinu. and [helping him put them]} into the basket.’	depictive

[2] Notation in examples: Dots followed by decimal numbers — absolute pauses and their length in seconds; ə̈ə̈ə(0.3) and ɯɯɯ(0.8) — plain and nasal filled pauses; symbol = indicates a truncated word; comma indicates a non-sentence final EDU, period a sentence-final EDU; square brackets indicate the boundaries of individual gestures, curly brackets — catchments, angle brackets — gesture inertia.