Computational Linguistics and Intellectual Technologies:
Proceedings of the International Conference "Dialogue 2018"

Moscow, May 30—June 2, 2018

LANGUAGE PRODUCTION AND COMPREHENSION IN FACE-TO-FACE MULTICHANNEL COMMUNICATION[1]

Kibrik A. A. (aakibrik@gmail.com),
Fedorova O. V. (olga.fedorova@msu.ru)
Institute of Linguistics RAS and
Lomonosov Moscow State University, Moscow, Russia

Although language production and comprehension are parts of one and the same linguistic capacity, they have been studied separately for a long time. A key issue in the present day research is how the two processes are related, and whether transitions from thought to language and vice versa are accomplished by a single or two separate systems. Important progress in this area has been achieved in the field of psycho- and neurolinguistics; a brief review is provided in **Section 1**. In this paper we explore the production—comprehension relationship on the basis of our multichannel resource "Russian Pear Chats and Stories". In **Section 2** we describe this resource, including the stimulus material, data collection setup, participants and corpus size, and technical aspects. **Section 3** lays out two main theoretical notions: a model of face-to-face multichannel communication and a scheme of the production-comprehension interweaving in each interlocutor. In subsequent sections we discuss three case studies of production—comprehension relationships: relative contributions of kinetic channels to discourse understanding (**Section 4**), turn-taking and eye gaze (**Section 5**), and multichannel continuity (**Section 6**). The evidence of the multichannel corpus suggests a cognitive architecture that integrates language production and comprehension.

[1] This study is supported by Russian Science Foundation (grant #14-18-03819 "Language as is: Russian multimodal discourse").

1. Language production and comprehension in psycho- and neurolinguistic studies

A well-known integrated model of production and comprehension is proposed by [Pickering and Garrod 2013][2]. They critically represent the traditional view as a "cognitive sandwich" (a term from [Hurley 2008]), in which action (including production) and perception (including comprehension) are separate and distinct (see e.g. [Dell 1986] for language production and [MacDonald et al. 1994] for language comprehension). In contrast to this view, Pickering and Garrod argue that production and comprehension are interwoven, and that such interweaving is what enables people to make predictions regarding themselves and others. Pickering and Garrod use the notion of forward modeling in action[3], grounded in computational neuroscience [e.g. Wolpert et al. 2003]. According to this notion, speakers employ "forward production" in predicting their upcoming utterances. Listeners also use forward production models and covertly imitate speakers to predict their production. As Pickering and Garrod put it, "the account helps explain the rapidity of production and comprehension and the remarkable fluency of dialogue" [2013: 346].

The model introduced in Pickering and Garrod 2013 suggests that prediction plays a central role in language production and comprehension. The recent paper "Prediction is Production: The missing link between language production and comprehension" [Martin et al. 2018] just published in "Nature" is testing this hypothesis. According to the authors, except some indirect support showing that language production skills and prediction are related ([Hintz et al. 2016]; [Federmeier et al. 2010]; [Huettig 2015]), there has been no direct evidence so far that the production system is necessary for prediction during comprehension. To test this hypothesis, [Martin et al. 2018] explore whether availability of the production system is indeed necessary for prediction during sentence comprehension. They compared three groups of participants reading Spanish sentences containing an expected vs. an unexpected NP such as *El rey llevaba en la cabeza* ***una corona / un sombrero*** *antigua/antiguo* 'The king wore on his head an old **crown** [Fem] / **hat** [Masc]'. Lexical prediction can be measured through event-related potential responses derived from electrophysiological recording during sentence reading: the less predictable a word is, the more negative is the N400 component. Participants of the first group were assigned a secondary verbal task preventing them from using their inner speech, while participants from two other groups had other secondary tasks, presumably language-unrelated. The expectation effect was reduced in the first group compared to two others groups. This finding demonstrates that preventing subvocal rehearsal of the verbal input during sentence reading hinders prediction in sentence comprehension.

Neuroimaging studies of language have typically focused on either production or comprehension of speech material such as syllables, words, or sentences. A study

---

[2] For other integrated approaches see the CAPPUCCINO model by [McCauley and Christiansen 2011] and the "P-chain" framework by [Dell and Chang 2013], unifying the processes of comprehension, production and acquisition.

[3] Cf. a similar set of ideas in the early theory of motor control proposed by [Nikolai A. Bernstein 1967].

reported in Silbert et al. 2014 challenges this common practice, as well as the traditional assumption that the linguistic processes are primarily lateralized in the left hemisphere (see e.g. [Indefrey and Levelt 2004]; [Indefrey 2011]; but cf. [Jung-Beeman 2005]; [Hickok and Poeppel 2007] on language comprehension). [Silbert et al. 2014], looking at the production and comprehension of spontaneous narratives, identified all areas that are reliably activated in the brains of speakers telling a 15 minutes long narrative. Next, they identified areas that are reliably activated in the brains of listeners as they comprehended the same narrative. The results indicate that narrative production is not localized in the left hemisphere but recruits an extensive bilateral network, which overlaps extensively with the comprehension system. This study provides strong evidence for a close link between production and comprehension processes. Silbert et al. argue that "a shared neural mechanism supporting both production and comprehension facilitates communication and underline the importance of studying comprehension and production within unified frameworks" [2014: E4687].

To sum up, a number of current experimentally oriented students of communication assume that production and comprehension are interwoven, that prediction plays a central role in language production and comprehension, and that brain networks involved in language production and comprehension strongly overlap. The reviewed studies are limited to the unimodal perspective, according to which language is a purely vocal phenomenon. In this study we put the discussion of the production—comprehension relationship in a broader context of multimodal/multichannel communication.

2. Russian Pear Chats and Stories

We explore the production—comprehension relationships on the basis of our new resource "Russian Pear Chats and Stories" [Kibrik 2018b; Kibrik and Fedorova 2018]. We have used the well known Pear Film [Chafe ed. 1980] as the stimulus material for collecting recordings and the state of the art equipment including three individual industrial cameras (100 fps) and two eyetrackers Tobii Glasses II (50 Hz).

We have developed a new procedure of data collection. Each session lasted for about one hour and involved four participants with fixed roles—the Narrator (N), the Commentator (C), the Reteller (R), and the Listener (L). Before recording began, the Narrator and the Commentator each watched the film on a personal computer, trying to memorize the plot as precisely as possible. Then the Narrator told the Reteller about the plot of the film; this is a monologic stage—*first telling*. During the subsequent, interactive, stage—*conversation*—the Commentator added details and corrected the Narrator's story where necessary, and the Reteller checked his/her understanding of the plot, asking questions to both interlocutors. Then the Listener joined the group and another monologic stage—*retelling*—followed, during which the Reteller was retelling the film to the Listener. Finally, the Listener wrote down the content of the film.

The resource includes 40 sessions, with 160 Russian native speakers aged 18–36, including 60 men and 100 women. The overall volume of the resource is 15 hours of recording and about 170 K words. Vocal (auditory) data are annotated using the Praat program (fon.hum.uva.nl/praat), in accordance with a scheme including temporal

dynamics, segmentation into elementary discourse units (EDUs), absolute and filled pauses, accents, accelerated tempo, reduced pronunciation, lowered tonal register, etc. [Kibrik and Podlesskaya eds. 2009]. For the transcription of the kinetic (visual) data we used the annotation software ELAN (lat-mpi.eu/tools/elan). We annotated the following layers for facial/cephalic/manual/torso gestures: movements, movement chains, gestures, gesture chains, gesture phases, self-adaptors, postures, and posture changes. (See Litvinenko et al. 2017 for a more detailed description.) Gaze targets are coded as "surroundings" or "interlocutor", the latter further subdivided into "face", "hands", "torso", and "other". The minimal fixation duration is 100 ms, i.e. a participant' fixation on a target must last for at least 100 ms to be recognized as a gaze event [Fedorova 2017].

3. Face-to-face multichannel communication: Theoretical schemes

In face-to-face communication, interlocutors combine verbal structure, prosody, eye gaze, as well as facial, head, hand and torso gestures to produce integrated multichannel discourse. All of these communication channels are employed simultaneously and in conjunction with each other [Kress 2002; Kibrik 2010, 2018a,b; Müller et al. eds. 2014], see **Fig. 1**.

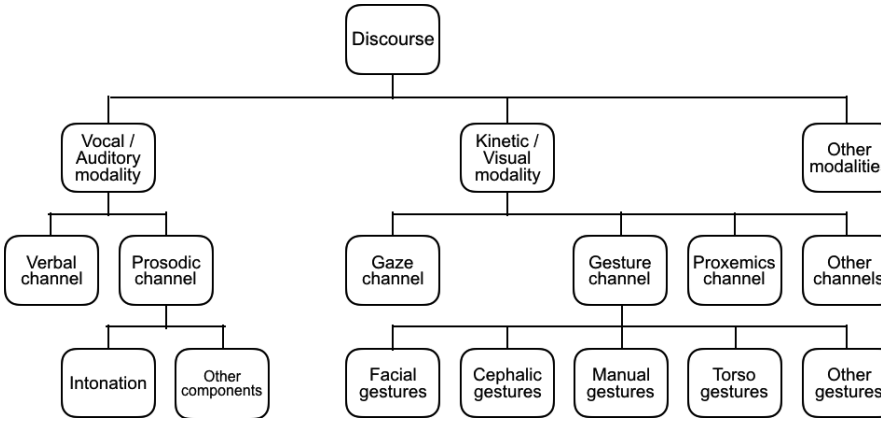

Fig. 1. Model of multichannel discourse

From the perspective of comprehension, one distinguishes the auditory and visual modalities; in terms of production, the same modalities can be dubbed vocal and kinetic.

The major change to the traditional notion of unimodal linguistic communication, necessitated by the bimodal (and multichannel) approach such as shown in **Fig. 1**, is the following. During the process of face-to-face communication each interlocutor performs the roles of addresser and addressee simultaneously. For example, a speaker, while producing speech at a given moment, simultaneously monitors

the listener's kinetic behavior (nodding, gaze, and manual gesticulation). **Figure 2** depicts the production-comprehension ensemble.

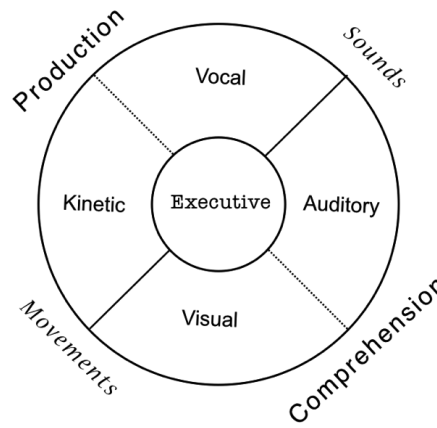

Fig. 2. Scheme of the production-comprehension interweaving in each interlocutor, taking part in face-to-face multichannel communication

The Executive is the central controlling component of the system (cf. similar executive components in theoretical models such as in [Baddeley 2007]; [Levelt 1989]; [McNeill 1992]). As in other models of cognitive processing, the Executive controls attentional processes and enables the system to selectively attend to some stimuli and ignore others.

Relying on the data of our resource, we discuss below three case studies of how processes of production and comprehension are interwoven in natural communication.

4. Case study 1. Kinetic channels: Relative contributions to discourse comprehension

In previous work we addressed the question of the relative contribution of various communication channels to the overall comprehension of spoken discourse. [Kibrik and Molchanova 2013] considered three communication channels employed in multichannel discourse in isolation: the verbal component, prosody, and kinetic-visual behavior. They found that all three channels play an important role in the overall process of conveying a message from a speaker to an addressee. They also found that participants had difficulties integrating the information from the visual and the prosodic channels, in the absence of the verbal channel. This suggests that in normal communication the verbal channel plays the role of an anchor to which the information from other channels is attached.

In [Kibrik and Molchanova 2013], the choice of a certain isolated communication channel was imposed upon experiment participants. In this study we observe how participants dynamically choose themselves which channel is most relevant at the

given time from the point of view of their current communicative goals. Such choice is registered with the help of eyetrackers identifying attention allocation.

In Section 2 we mentioned three stages of communication events, analysed in the "Russian Pear Chats and Stories" corpus. The second of these stages, *conversation*, is interactive: all the interlocutors actively contribute vocal and kinetic material. The first and the third stages (*first telling* and *retelling*) are monologic: in each of them only one participant is actively talking, that is producing the vocal signal. But the roles are distributed differently: in the *first telling*, the main speaker is the Narrator, while in *retelling* the Reteller. This variation across the three stages allows us to compare the participants' oculomotor behavior. Specifically, we compare the visual attention[4] distribution in one participant: the Narrator. The analysis is based on three recordings: 04, 06 and 23.

Table 1. The distribution of Narrator's visual attention, *first telling* (summary duration, s and in %)

	Gaze target		
Recording #	face R	surroundings	Total
04	143.767 (61.2%)	91.205 (38.8%)	234.972 (100%)
06	47.908 (46.0%)	56.646 (54.0%)	104.554 (100%)
23	63.849 (38.3%)	103.032 (61.7%)	166.881 (100%)

As the data in Table 1 suggest, a primary speaker divides his/her visual attention exclusively between the primary listener's face (in this case, the Reteller) and the surroundings.

Table 2. The distribution of Narrator's visual attention, *conversation* (summary duration, s and in %)

	Gaze target					
Recording #	face R	face C	hands R	hands C	surroundings	Total
04	386.078 (65.5%)	40.077 (6.8%)	4.840 (0.8%)	0.320 (0.1%)	158.390 (26.8%)	589.705 (100%)
06	236.841 (46.7%)	95.216 (18.8%)	29.114 (5.7%)	29.074 (5.7%)	116.844 (23.1%)	507.089 (100%)
23	65.819 (20.9%)	178.228 (56.5%)	0.000 (0%)	1.380 (0.4%)	69.983 (22.2%)	315.41 (100%)

[4] It is generally recognized that attention and eye movements are closely related, even though the nature of this relationship is not yet fully understood; see e.g. [Smith and Schenk 2012].

As is clear from Table 2, when the Narrator is involved in multi-party discourse, watching the surroundings takes significantly[5] less time; the results for the Narrators in all three recordings are statistically indistinguishable. Apart from that, his/her attention is distributed between the interlocutors' faces and, to some extent, their hands. This apparently depends on the level of their activity in conversation.

Table 3. The distribution of Narrator's visual attention, *retelling* (summary duration, s and in %)

	Gaze target			
Recording #	face R	hands R	surroundings	Total
04	246.541 (74.2%)	54.826 (16.5%)	31.123 (9.3%)	332.490 (100%)
06	324.598 (75.5%)	57.914 (13.5%)	47.626 (11.0%)	430.138 (100%)
23	184.728 (77.2%)	12.697 (5.3%)	41.945 (17.5%)	239.370 (100%)

The evidence in Table 3 demonstrates that, while listening to the Reteller, the Narrator directs his/her gaze at the surroundings to a still lesser extent (the difference is significant for recordings 04 and 06). The vast majority of fixations are on one interlocutor: the Reteller, who is the primary speaker; the results in all three recordings are statistically indistinguishable. The Narrator's attention is distributed between the Reteller's face and, to a lesser extent, his/her hands.

We thus can conclude that a participants's visual attention is distributed in a systematic way, depending on his/her role as the primary speaker vs. an equal interlocutor vs. a listener. This distribution varies from one discourse stage to another and is sensitive to the participants' communicative activity.

5. Case study 2. Turn-taking and eye gaze

One of the chapters of the monograph [Grishina 2017] is titled "A grammar of gaze". In that chapter Elena A. Grishina explored the gaze direction of interlocutors at turn boundaries. Her analysis is based on the Russian movies ["Brilliantovaja ruka" The Diamond Arm (1968)] and ["Den' vyborov" Election Day (2007)]. Grishina found substantial differences in the speaker's gaze distribution at turn boundaries (her sample includes 527 instances). In particular, when a speaker directly addresses the listener (a "provoking" speech act), at the end of his/her turn s/he watches the listener, thus controlling the process of turn handover. If a speaker performs a neutral speech act, not demanding an immediate response from the listener, his/her gaze is usually directed at the surroundings.

[5] Here and below the statistical significance was analysed with the help of the chi-square test (Holm correction for multiple comparisons, $p < 0.01$).

We have tested this generalization against our data. In our pilot study, we looked at the conversation stage (8.5 min.) of recording 22 of the "Russian Pear Chats and Stories" corpus. Our material suggests two important differences from the Grishina's material. First, we analyse the talk and gaze of three interlocutors. Second, the behavior of two interlocutors was registered with the help of eyetrackers, which ensures a high precision of annotation. In the analysis we use the scores vocal transcript, prepared by Nikolay A. Korotaev. In the analysed conversation we identified 107 instances of turn boundaries, among these 32 were preceded by provoking turns and 75 by neutral turns. Table 4 illustrates an excerpt from the conversation's vocal transcript.

Table 4. An example of multichannel turn-taking

TimeS	TimeE	Narrator		Commentator		Reteller	
701,73						R-v076	Но не \очень симпатичная девочка¿
702,89	703,08	N-v329	\Нет,				
703,08		N-v330	она не \очень /симпатичная, ≈≈				
703,24				C-v261	(' 0.61) Не \знаю.		
	703,30						
703,30						R-v077	{laugh 8.32}
	704,53						
	704,59						
704,59	706,15			C-v262	/Вполне себе симпатичный		
706,48	707,15			C-v263	\Круглощёкий.		

In the excerpt shown in Table 4 the Narrator first watches the Reteller. At the beginning of her turn (N-v329) she moves her gaze to the surroundings and then, cued by the Commentator's voice, shifts her gaze at him. As for the Commentator, at the beginning of his turn (C-v261) he watches the Reteller, and then shifts his provoking gaze at the Reteller, waiting for her response.

The obtained results generally accord with Grishina's conclusions, but some important differences have been noted. First, the interlocutors' gaze is rarely directed at the surroundings; 75% of the time it is distributed between two partners in communication. Still, during those intervals when interlocutors' gaze is actually directed at the surroundings, 95% of the time this is in line with the Grishina's generalization stating that this happens in neutral speech acts. Second, the basic principle "Watch the interlocutor who has just started talking" is violated in certain instances (accounting for 12% of all instances); according to our current interpretation, this happens when a speaker assesses his/her own vocal contribution as being of low significance.

6. Case study 3. Multichannel continuity

The division of communication into three stages, including the *first telling*, the *conversation*, and the *retelling*, was originally seen as a technical procedure but later developed into a research issue in its own right. Multichannel communication is so organized that identifying boundaries between stages is rarely an easy task. Various channels suggest their own boundaries that do not have to coincide. Whereas the verbal, the prosodic, and the manual-gestural components are relative well coordinated (cf. Fedorova et al. 2016 on the degree of such coordination), the cephalic gestures, the facial gestures and eye movements regularly disturb coordination. In particular,

in the course of a vocal-manual pause, typical of a stage boundary, interlocutors usually produce signals that convey "turn handover". Such a signal is often a head turn, accompanied by particular facial movements, especially smile. Furthermore, the interlocutor's gaze frequently lags behind, remaining fixed on the other participant.

Consider an example from recording 35, specifically the boundary between the *first telling* and the *conversation*. Table 5 illustrates a five seconds excerpt that embraces a whole gamut of various vocal and kinetic actions performed by the interlocutors. A vocal expression is shown in line 1 of the Table. The behaviors listed in line 9 and further are clearly separable from that vocal expression, and we posit a boundary at the end of line 1, that is at 0.8 s from the beginning. In contrast, the behaviors listed in lines 2 to 8 all intersect that boundary.

Table 5. Recording #35, around the boundary between the *first telling* and the *conversation*

1.	N: vocal	Я закончила.	00.000–00.800
2.	R: manual	adaptor II[6]	00.000–05.100
3.	R: gaze	fixation	00.000–01.010
4.	N: gaze	fixation	00.000–01.620
5.	C: gaze	fixation	00.000–02.240
6.	N: cephalic	nod	00.150–01.120
7.	N: manual	adaptor II	00.250–05.100
8.	R: facial	smile	00.550–05.100
9.	R: cephalic	turn	00.960–01.800
10.	N: cephalic	turn	01.580–02.360
11.	C: cephalic	turn	02.280–02.570
12.	R: cephalic	nod	02.380–02.830
13.	C: cephalic	turn	03.380–04.010
14.	N: vocal	Ну?..	03.600–03.800
15.	C: vocal	Я не помню…	04.160–04.980
16.	C: manual	adaptor II	04.240–05.100

We can thus generalize that communication between interlocutors is not interrupted even for a fraction of a second. It is being supported by a network of channels. Work load is being swiftly and dynamically carried over from one channel to another. An interlocutor simultaneously functions as an addresser and an addressee.

7. Conclusion

Traditionally, production and comprehension are regarded as distinct processes. Some modern approaches, however, amend this dichotomy, proposing that production and comprehension are interwoven, and such interweaving is possible on the

[6] So-called "adaptors II" are minor movements without a clearly identifiable communicative function.

basis of prediction. Some recent studies on the functional neuroanatomy of language suggest that the brain networks involved in speaking and listening strongly overlap.

The evidence of multichannel communication also suggests a cognitive architecture that integrates language production and comprehension. As Pickering and Garrod said, "interlocutors must simultaneously produce their own contributions and comprehend the other's contribution. Clearly, an approach to language processing that assumes a temporal separation between production and comprehension cannot explain such behavior" [2013: 330]. Communicative actions of the interlocutors thus form a complex and heterogeneous network that must be credited a capability to involve simultaneous and multidirectional thought exchange.

References

1. *Baddeley A. D.* (2007), Working memory, thought, and action, Oxford: Oxford University Press.
2. *Bernstein N. A.* (1967), The co-ordination and regulation of movements, Oxford: Pergamon Press.
3. *Chafe W.* (ed.) (1980), The pear stories: Cognitive, cultural, and linguistic aspects of narrative production, Norwood: Ablex.
4. *Dell G. S.* (1986), A spreading-activation theory of retrieval in sentence production, Psychological Review, Vol. 93, pp. 283–321.
5. *Dell G. S., Chang F.* (2013), The P-chain: relating sentence production and its disorders to comprehension and acquisition, Philosophical Transactions of the Royal Society B: Biological Sciences, Vol. 369 (1634).
6. *Federmeier K. D., Kutas M., Schul R.* (2010), Age-related and individual differences in the use of prediction during language comprehension, Brain Language, Vol. 115 (3), pp. 149–161.
7. *Fedorova O. V.* (2017), Distribution of the interlocutors' visual attention in natural communication: 50 years later [Raspredeleniye zritel'nogo vnimaniya sobesednikov v estestvennoy kommunikatsii: 50 let spustya], E. V. Pechenkova, M. V. Falikman (eds.) Cognitive science in Moscow: new research. Proceedings of the conference [Kognitivnaya nauka v Moskve: novye issledovaniya. Materialy konferentsii]. Moscow: BukiVedi, IPPiP, pp. 370–375.
8. *Fedorova O. V., Kibrik A. A., Korotaev N. A., Litvinenko A. O., Nikolaeva Ju. V.* (2016), Temporal coordination between gestural and speech units in multimodal communication [Vremennaya koordinatsiya mezhdu zhestovymi i rechevymi edinitsami v mul'timodal'noy kommunikatsii], Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2016" [Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii], RGGU, Moscow, pp. 159–170.
9. *Grishina E. A.* (2017), Russian gestures from a linguistic perspective [Russkaya zhestikulyatsiya s lingvisticheskoy tochki zreniya], Moscow: Jazyki slavyanskoy kul'tury.

10. *Hickok G., Poeppel D.* (2007), The cortical organization of speech processing, Nature Reviews Neurosciences, Vol. 8 (5), pp. 393–402.
11. *Hintz F., Meyer A. S., Huettig F.* (2016), Encouraging prediction during production facilitates subsequent comprehension: Evidence from interleaved object naming in sentence context and sentence reading, The Quarterly Journal of Experimental Psychology, Vol. 69 (6), pp. 1056–1063.
12. *Huettig F.* (2015), Four central questions about prediction in language processing, Brain Researcg, Vol. 1626, pp. 118–135.
13. *Hurley S.* (2008), The shared circuits model (SCM): How control, mirroring, and simulation can enable imitation, deliberation, and mindreading, Behavioral and Brain Sciences, Vol. 31 (01), pp. 1–22.
14. *Indefrey P.* (2011), The spatial and temporal signatures of word production components: a critical update, Frontiers in Psychology, Vol. 2, p. 255.
15. *Indefrey P., Levelt W. J.* (2004), The spatial and temporal signatures of word production components, Cognition, Vol. 92 (1–2), pp. 101–144.
16. *Jung-Beeman M.* (2005), Bilateral brain processes for comprehending natural language, Trends in Cognitive Science, Vol. 9 (11), pp. 512–518.
17. *Kibrik A. A.* (2010), Multimodal linguistics [Mul'timodal'naya lingvistika], Yu. I. Aleksandrov, V. D. Solov'yev (eds.), Cognitive studies [Kognitivnyye issledovaniya], Vol. IV, Institute of psychology, Moscow, pp. 134–152.
18. *Kibrik A. A.* (2018a), Russian multichannel discourse. Part I. Setting up the problem [Russkiy mul'tikanal'nyy diskurs. Chast' I. Postanovka problemy], Psikhologicheskiy zhurnal, Vol. 39 (1), pp. 70–80.
19. *Kibrik A. A.* (2018b), Russian multichannel discourse. Part II. Corpus development and avenues of research [Russkiy mul'tikanal'nyy diskurs. Chast' II. Razrabotka korpusa i napravleniya issledovaniy], Psikhologicheskiy zhurnal, Vol. 39 (2), pp. 78–89.
20. *Kibrik A. A., Fedorova O. V.* (2018), A "portrait" approach to multichannel discourse, Eleventh International Conference on Language Resources and Evaluation (LREC), Japan, 5–12 May 2018.
21. *Kibrik A. A., Molchanova N. B.* (2013), Channels of multimodal communication: Relative contributions to discourse understanding, M. Knauff, M. Pauen, N. Sebanz, I. Wachsmuth (Eds.), Proceedings of the 35th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society, pp. 2704–2709.
22. *Kibrik A. A., Podlesskaja V. I.* (eds.), (2009), Night Dream Stories: A corpus study of spoken Russian discourse [Rasskazy o snovideniyakh: korpusnoye issledovaniye russkogo ustnogo diskursa]. Moscow: Jazyki slavyanskikh kul'tur.
23. *Kress G.* (2002), The multimodal landscape of communication, Medien Journal, Vol. 4, pp. 4–19.
24. *Levelt W. J. M.* (1989), Speaking: From intention to articulation, MIT Press.
25. *Litvinenko A. O., Nikolaeva Ju. V., Kibrik A. A.* (2017), Annotation of Russian manual gestures: Theoretical and practical issues [Annotirovaniye russkikh manual'nykh zhestov: teoreticheskiye i prakticheskiye voprosy], Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2017" [Komp'yuternaya Lingvistika i Intellektual'nye

Tekhnologii: Trudy Mezhdunarodnoy Konferentsii “Dialog 2017”], Moscow: RGGU, pp. 255–268.
26. *MacDonald M.C., Pearlmutter N. J., Seidenberg M. S.* (1994), The lexical nature of syntactic ambiguity resolution, Psychological Review, Vol. 101, pp. 676–703.
27. *Martin C. D., Branzi F. M., Bar M.* (2018), Prediction is production: The missing link between language production and comprehension, Nature, Vol. 8 (1079).
28. *McCauley S., Christiansen M. H.* (2011), Learning simple statistics for language comprehension and production: The CAPPUCCINO model, L. Carlson, C. Hölscher, T. Shipley (eds.), Proceedings of the 33rd annual conference of the cognitive science society, Austin, TX: Cognitive Science Society, pp. 1619–1624.
29. *McNeill D.* (1992), Hand and mind: What gestures reveal about thought, Chicago: University of Chicago Press.
30. *Müller C., Fricke E., Cienki A., McNeill D.* (eds.) (2014), Body—Language—Communication: An international handbook on multimodality in human interaction, Berlin: Mouton de Gruyter.
31. *Pickering M. J., Garrod S.* (2013), An integrated theory of language production and comprehension, Behavioral and Brain Sciences, Vol. 36 (04), pp. 329–347.
32. *Silbert L. J., Honey C. J., Simony E., Poeppel D., Hasson U.* (2014), Coupled neural systems underlie the production and comprehension of naturalistic narrative speech, Proceedings of the National Academy of Sciences of the United States, Vol. 111, pp. E4687–E4696.
33. *Smith D. T., Schenk T.* (2012), The premotor theory of attention: time to move on?, Neuropsychologia, Vol. 50, pp. 1104–1114.