Perception and Effectiveness of Content in Social Networks: Russian Case[1]

Alexander A. Kharlamov[1][0000-0003-2942-5101],
Maria A. Pilgun[2] [0000-0002-8948-7075]

[1] Institute of Higher Nervous Activity and Neurophysiology of RAS, Moscow, RF
Moscow State Linguistic University, Moscow, RF
Higher School of Economic, RF
kharlamov@analyst.ru
[2] Russian Academy of Sciences, Institute of Linguistics (Russia), RF
pilgunm@yandex.ru

Abstract

The paper considers studying the perception and effectiveness of media content in the Russian-language social networks, analyzing the causes that affect the perception and distribution of network content. The material used was the database of the Russian social network VKontakte. A cross-disciplinary multimodal approach and psycholinguistic analysis were used. The study revealed that the perception of content, the success of posts (the maximum number of likes, reposts, comments, the achievement of the author's communicative goal – transferring users' actions from online to offline mode) ensure a source of distribution (the symbolic capital of the author increases the significance of the text when it is perceived by the network community), the use of features of the current communicative situation and the accuracy of the imperative strategy.

Keywords: social media, speech perception, digital content, semantic role.

1. Introduction

The importance of language learning and communication in Russian-language social networks, and problems of content effectiveness is determined by the growth in the number and activity of Runet users. Thus, in September 2018, the audience of the Runet amounted to 1 773 385 284 users (according to LiveInternet). Data illustrating the statistics for October and November are presented in Table 1.

1 The work was carried out within the framework of the RFBR project «The structure and content of ideas about a person, state, power in the linguistic consciousness of Russian speakers (Republic of Crimea)» (19-012-00295 A)

Table 1

Main indicators of Runet sites traffic

Social network	Authors	Authors	Messages
Browsing	2 277 233 933	2 334 978 477	2 334 523 814
Sessions	424 100 001	411 191 791	407 634 892
Visitors	149 451 614	149 936 458	147 830 604
Average online	1 194 984	881 392	1 194 555
Average active online	573 665	633 563	710 352
Average duration (min.)	5.6	5.9	5.9
Views per visitor	15	16	16

The most popular Russian-language resource VKontakte (see Table 2) views itself as a platform designed to bring people, services and companies together by creating simple and convenient communication tools. Technical and communication features of the resource indicate it is in high demand among users: 97 million users per month; 6.5 billion messages per day; 1 billion «likes» per day; 77% users of mobile platforms; 500 million video views per day; 86 language versions (see https://vk.com/about).

Table 2

Statistics for social networks (January 2018)

Social network	Authors	Messages
VKontakte	26 633 778	410 168 754
Instagram	10 396 161	89 995 624
Facebook	1 824 256	53 266 232
Twitter	1 010 690	41 184 629
My world (my.mail.ru)	66 957	4 336 042
LiveJournal	59 523	3 387 579

(https://br-nalyt-ics.ru/statistics/author?hub_id=16048anddate=201708andcountry_id=20andperiod_type=mont)

With the expansion of the virtual space, transformations began to occur in the functioning of speech structures that are caused by the need to adapt to new communicative conditions. M. Martin believes that the «new language» was formed in the Internet in order to compensate for the lack of non-verbal and paraverbal means of communication [1]. Indeed, web-communication formats that replace spoken language communication to some extent, even if there are auditory components, should compensate for the lack of information that is transmitted using paraverbal, mimic, pantomimic, proxemic means. In addition, the virtual space imposes additional specific modes on a specific communicative situation. Convergence and multidimensionality of online communications require a cross-disciplinary approach, mixed methods that are used in many studies that represent various aspects of interaction in the web environment [2-6].

There is already an extensive research tradition revealed in analysis of content and network interactions. In particular, the specifics of communicative processes in social networks, the impact on political and social processes were studied in [7-10].

Recently, analysis of network communications has been presented by research in the field of multimodal dynamic networks [11], as well as bimodal and three-modal networks [12-14]. One-modal «friend-friend» communications are of little interest for researchers.

Research in a multimodal perspective has become widespread in recent decades [15–17]. It is a multimodal approach that seems to be the most adequate in analyzing network content, since it allows conversion of data and information coming through various channels.

Analysis of the media space as a multimodal sphere is also becoming increasingly widespread in various studies [18-19]. Meanwhile, it is difficult to disagree with the opinion of D. Crystal, that, whatever the culture of the Internet, it is still largely based on texts [20].

The problem of content evaluation remains the most important task, both for educational, and scientific, social, public and other purposes. Today, there are two approaches to solving this problem: using technical resources and direct perception by a human and expert evaluation. In particular, automatic text evaluation becomes quite popular in evaluation of L2 written works when teaching a foreign language and testing a large number of students [21-24]. Thus, the Coh-Metrix program was widely spread and positively rated by experts, although researchers note a number of other problems associated with its use [24–25].

An important criterion of adequacy and objectivity is the comparison of the results obtained using software and direct human participation, for example, from professional experts (see, for example, [26]. It is the contamination of these two approaches, the comparison of the results obtained by automatic means and with the direct participation of the human, using analytical procedures that seems to be the most correct solution to the problem of identifying the accuracy and adequacy of the content evaluation.

One should focus on the issue of content effectiveness. Today, a sufficient amount of research confirms that computer models using digital footprints of people are capable of diagnosing a person's character with a high degree of accuracy. Thus, predictive analytics based on the construction of psychological portraits and behavior models of actors are presented in [27]. Researchers consider the «like» to be a universal digital footprint [27-29]. In addition, reposts and comments are also bright markers of high demand for particular content in a network environment.

Thus, within the framework of this study, the following indicators will be considered as indicators of the effectiveness and relevance of (demand for) network content: the number of likes, reposts, and comments as well as the achievement of a communicative goal (in this case: switching from online to offline communication, organizing an opposition meeting).

2. Identification of causes affecting the perception and distribution of network content

The goal of this study is to identify the causes that affect the perception and distribution of network content.

Research questions are:

- What determines the perception of network content by actors?
- What influences the effectiveness of network content?
- What content is most popular in a network environment and attracts the attention of users?

The material for the study was the database of the social network VKontakte (March 2017) associated with opposition meetings that took place on March 26 under the conventional name «On vam ne Dimon»/«He's not Dimon for you» (#DimonOtvetit/#Dimon will take responsibility). The protests were caused by the distribution of the investigative film by A. Navalny in the network about the activities of Prime Minister Dmitry Medvedev.

- content of involved actors – n 43 712,
- content of active actors – n 15 021,
- relevant posts – n 23 602,
- number of words – n 470 893,
- number of characters – n 3 569 442.

The content is used on the basis of the rules of VKontakte[2] specified in p. 7.1.3., as well as art. 1274 of the civil code.

2.1. Research method

The study involved a cross-disciplinary multimodal approach. Content analysis, semantic and psycholinguistic analysis were also used to correctly interpret the perception of the content.

Formal analysis was performed in stages, combining quantitative and qualitative methods of analysis.

[2] P. 7.1.3. By posting on the Site his/her lawfully owned Content, the User grants the other Users a non-exclusive right to use it in the framework of the functional provided by the Site by viewing, reproducing (including copying) and other rights exclusively for the purpose of personal non-commercial use, except where such use infringes or may infringe the right owner's interests protected by law.

2.1.1. Procedures

The sequence of actions during the analysis corresponded to the following algorithm:

I. Selection of users involved in the semantic field «On vam ne Dimon»/«He's not Dimon for you» (#DimonOtvetit/#Dimon will take responsibility).

  I.1. Selection of active actors that generate content within the framework of this semantic field.

  I.2 Analysis of communicative network actions and digital footprints of active actors.

II. Content analysis.

  II.1. Identification of a common network of concepts characterizing the entire collected database, and the thematic structure of the analyzed content (using TextAnalyst 2.0 and Automap).

  II.2. Allocation of two types of posts: database 1 with the maximum number of digital footprints; and database 2 with zero indicators of digital footprints. Both groups were analyzed using TextAnalyst2.0.

  II.3. Analysis of the two groups using the method of semantic differential.

At the stage of content research, the semantic network common for the whole corpus and the thematic structure of the analyzed content extracted from the network were formed using the TextAnalyst program.

The list of the most significant notions and concepts (having the highest rank) bearing the main meaning (semantic load) obtained from the semantic network, made it possible to determine the core of information, the semantic accents most important for actors.

The identification of the thematic structure of the selected network content made it possible to describe the content in the form of a hierarchy of related topics and sub-topics that reflect the basic concepts and correspond to the nodes of the network notions. The tools used allowed creation of a hierarchical thematic structure, identification of the basic topic and revelation of the relationships with sub-topics that form multi-level semantic networks.

In addition, the content was analyzed regarding the connectivity of the subject tree. Topic clusters that emerged as a result of changing the threshold for the weight of relationships in the network of notions (the break of more or less strong relationships) were identified. Such analysis of the collected dataset allowed analyzing the structure of the consolidated media text in various sections and at various levels of semantic depth.

2.1.2. Tools

TextAnalyst 2.0 is system for automatic semantic text analysis. The TextAnalyst technology is intended for the automatic statistical formation of a homogeneous (associative) semantic network of text (which is a semantic portrait of the text) describing a certain situation. The semantic network is a graph, where the set of vertices corresponds to the concepts of the text, and the arcs correspond to the relationships of these concepts in the text. Both the vertices and the arcs of the network have weight characteristics reflecting the ranks of the concepts and their relationships in the text.

Such a semantic portrait makes it possible to reveal the semantic core of the text, which characterizes the situation described in the text to the fullest extent.

The formed semantic network of the text can be used to identify the thematic structure of the text. The minimal tree-like sub-graph of the semantic network obtained from it by breaking weak relationships and revealing the most significant vertex is the table of contents of the text characterizing its sections with regard to correlating interconnected parents and children of various levels.

The semantic network of the text is formed in several steps. (1) At the first step, information that generates information noise is removed from the text: stop words, service words and common words are removed; lemmatization is performed, which makes the subsequent analysis sustainable. (2) At the second step, a frequency network of the text is formed by combining the word pairs found in the sentences of the text into chains. The frequency of occurrence of words in the text in this network characterizes the weights of the vertices, and the frequency of occurrence of word pairs in the sentences of the text characterizes the weights of the arcs. (3) At the third step, the weights characteristics of the vertices are recalculated using an iterative procedure, taking into account the connectivity of the vertices in the network. The vertices associated with a large number of vertices with large weights receive a greater weight compared to other vertices.

Thus, this re-ranking of vertices takes into account the connectivity of words in the text to a depth of n steps, where n corresponds to the number of iterations in the re-ranking procedure (in the TextAnalyst program, n = 10). Accounting for word relationships in the text to a depth of n steps distinguishes the algorithm for constructing a subject tree from algorithms for thematic modeling, which uses the so-called mono-gram model of text (bag-of-words), in which words are considered as non-related to each other in the text.

Automap is a text mining tool that allows extraction of information using network analysis methods; it supports the extraction of several types of data from unstructured documents.

Gephi (algorithm Force Atlas 2) is software for visualizing network structures [30].

Tableau is a platform for creating visual analytics, or interactive data visualization.

2.2. Results and discussions

I. At the first stage of the study, all users who showed interest in the opposition meeting were identified. Those involved in the linguistic mode were analyzed by verbal content collected by the hashtags #dimonotvetit containing the original author's content and user content about active actors within the specified semantic field.

Verbal content analysis scheme

1. Identification and analysis of the semantic field.

1.1. Explicit means of expressing intentions, evaluations, opinions.

1.2. Direct means of influence (effect, persuasion).

1.2.1. Spelling and graphic tools.

1.2.2. Speech tools:

1.2.2.1. Phonetic.

1.2.2.2. Lexical.

1.2.2.3. Morphological.

1.2.2.4. Syntactic.

1.2.2.5. Stylistic.

1.3. Implicit expression of intention, evaluation, opinion:

1.3.1. Spelling and graphic tools.

1.3.2. Speech tools:

1.3.2.1. Lexical.

1.3.2.2. Morphological.

1.3.2.3. Syntactic.

1.3.2.4. Stylistic.

1.3.3. Rhetorical means.

1.3.4. Speech means of influence.

1.3.5. Associative links.

The results of data collection, analysis and clustering of the involved users are presented as visualization (fig. 1).

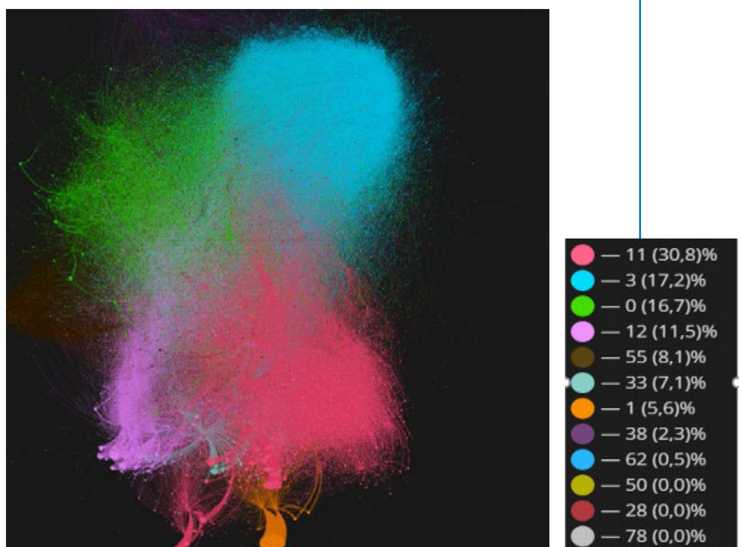

Fig. 1. Involved users who reacted to the March 26 meeting

The algorithm of clustering the graph vertices was performed according to [31]. The analysis used only the content generated by real users. The artificial entities were revealed using a profile analysis: no friends, no or a small amount of visual data (photo without a particular user's face), no posts from other users with links to the actor, activity only during certain periods corresponding to significant time periods associated with political events, as well as belonging to already known botnets. In the course of the study, the SocialDataHub technology was used, which made it possible to automatically analyze the profile in the social network.

I.1. Further, an analysis was performed that allowed identification of active actors who generated some content within this semantic field and who took direct part in protest events. The analysis was performed using a verbal mode (author's content) and a visual mode – according to the data of static visual materials using the SocialDataHub algorithm, which provides an accuracy of up to 85%.

Analysis of the main characteristics of the actors and the three clusters of participants by objectives.

I.2. The analysis of communicative network actions made it possible to identify and rank the digital footprints of active actors for further research.

The communicative network behavior was analyzed according to the following scheme:

1. The number of «likes».

2. The number of posts on a relevant topic.

3. Communicative resources:

    3.2. Preferred strategies (ritual, persuasive, imperative, provocative).

    3.3. Tactics of direct communicative impact.

    3.4. Tactics of indirect communicative impact.

4. Number of reposts on the relevant topic.

5. Number of comments.

6. Number of links.

7. Number and characteristics of groups.

8. Number of messages.

9. Integrated communication resources of indirect influence.

The results are presented in fig. 2.

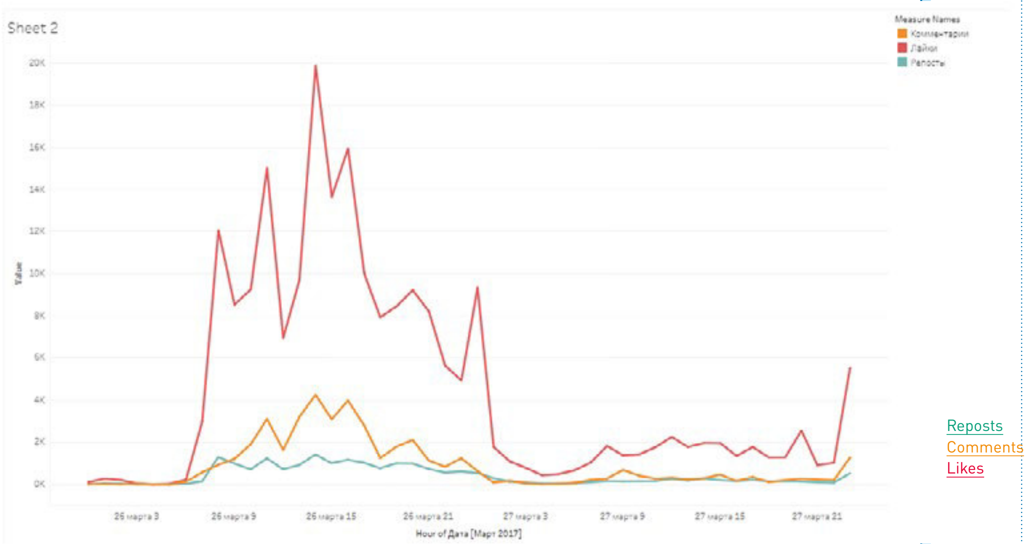

Fig. 2. Ratio of digital footprints (likes, reposts, comments)

II. Content analysis.

II.1. At the initial step of the content research, frequency concepts, a common network of motions that characterize the entire collected database, and the thematic structure of the analyzed content were identified (see Fig. 3).

It is natural that the entire information hierarchy in this content belongs to a single topic, has the appearance of a tree with a single root: «DimonOtvetit». The maximum weight of the notion (100) was found for the following nominations: DimonOtvetit/Medvedev/Dimon/OnVamNeDimon/OnNamNeDimon (Dimon will take responsibility/Medvedev/

Dimon/He's not Dimon for you/He's not Dimon for us), meeting / protest against corruption, Russia/country, Tverskaya, cities, detained.

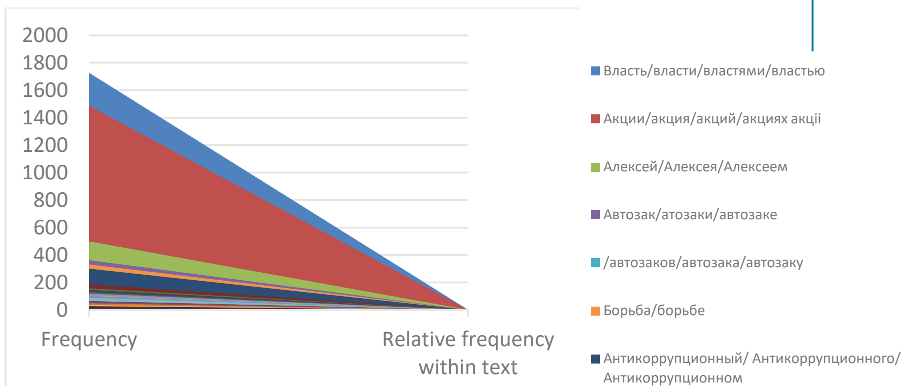

Fig. 3. Frequency concepts

The maximum relationship weight for the concept of corruption in the common database was assigned to the following nominations: meetings against corruption (74), cities (45), Medvedev (42)/Dimon (38), country (41).

The analysis of the thematic structure of the collected content reflects the main content of the integrated media texts associated with meetings against corruption throughout the country, and the persona of an official from the top government echelons who provoked public outrage and served as a catalyst for the protest movement.

II.2. In accordance with the selected criteria for the effectiveness and demand for the content, the posts were ranked by the number of digital footprints.

II.2.1. Further, media texts were identified, which caused the greatest resonance in the network environment (dataset 1).

Features of dataset 1:
memory size – 98 KB;
number of posts – 40;
number of words – 2 226;
number of characters – 16 031;
number of likes – 52 040;
number of comments – 7 295;
number of reposts – 3 813.

As well as media texts that did not cause any reaction at all from the network community (dataset 2).

Features of dataset 2:
memory size – 97 KB;
number of posts – 119;
number of words – 1 830;
number of characters – 14 763;
number of likes –0;
number of comments – 0;
number of reposts – 0.

II.2.2. The construction of the semantic networks for both datasets made it possible to reveal the most significant information emphasized by the authors of the messages.

In semantic network 1, the notions referring to the initiator of protest movements have the greatest weight. The center of the semantic network, the relevant information are personalized; the attention of users is concentrated on the political leader who initiated the meetings, who becomes the undisputed leader of opinions, greatly expanding the virtual electorate.

Curiously, semantic network 2 represents a wider set of information clusters. Notions characterizing the causes for the protest meetings have the maximum weight. The center of the semantic network, the relevant information are not only connected with the opposition leader, but also represent the organization (ACF, Anti-Corruption Foundation), which conducted the anti-corruption investigation; the semantic focus is also on the reaction of the authorities, negative attitude to the media information policy, etc.

The thematic structure of dataset 1 (meeting – 100) is not to be inferior regarding dataset 2 (meetings – 100; fire extinguisher – 99).

The associative networks with a nuclear nomination «Meeting» identified in the materials of the 1st and 2nd datasets are generally identical (see fig. 4, 5). Meanwhile, it should be noted that the associative search against the base of dataset 1 revealed a great significance of the effect of covering the meetings held, their territorial distribution, detentions, etc.

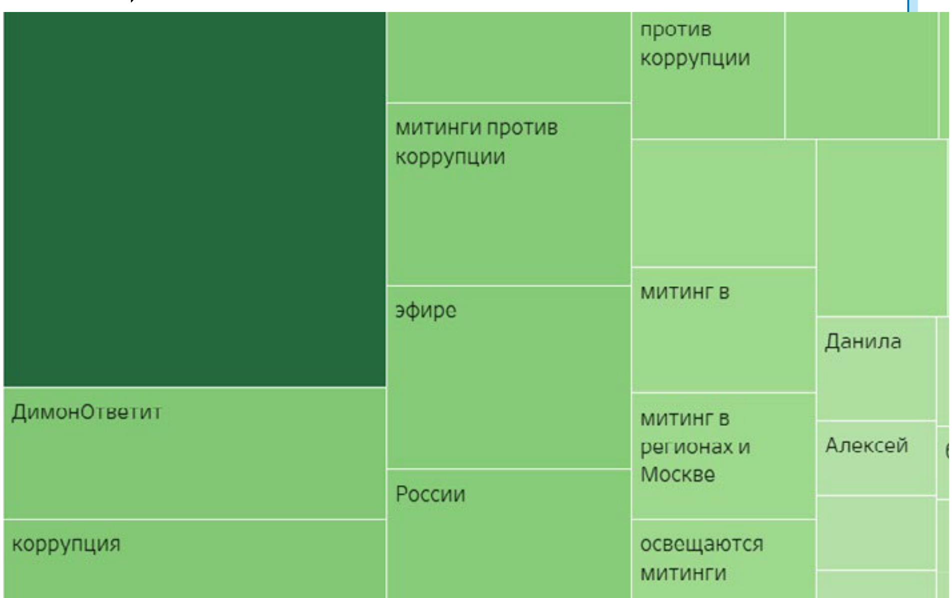

Fig. 4. Associative search against the base of dataset 1

II.3. At the final stage of the study, the perception of content was tested with the direct participation of the subjects using the semantic differential method. The focus group consisted of 20 participants (aged 20-25, 60% females, 40% males). The results also showed no significant discrepancy in the perception of database 1 and database 2 (see Fig. 6).

Fig. 5. Associative search against the base of dataset 2

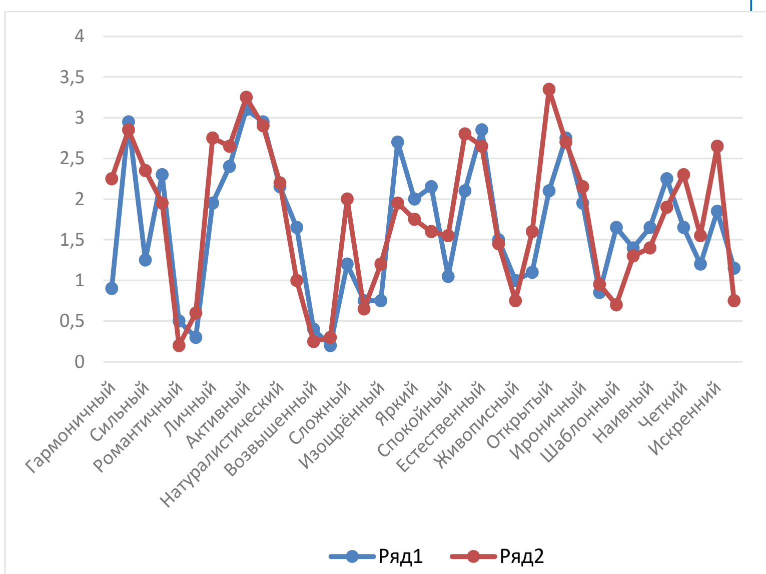

Fig. 6. Results obtained using the semantic differential method

Thus, the semantic and linguostylistic characteristics of database 1 and database 2 are almost the same, in contrast to the characteristics of network effectiveness, which oppose the two data clusters as the most effective and the less effective ones.

3. Conclusion

The results of the study suggest that the perception and effectiveness of network content is determined by a complex of causes.

The effectiveness of the content, the degree of reaction of the network community depends on the level of symbolic capital of the source of content distribution, taking into account the characteristics of the communicative situation and the success of the imperative strategy used.

The significance of network content is determined, first of all, by the actor, who is recognized as the author of a particular message. The charisma of the opposition leader, as well as the circle of his/her companions who distributed the content that constitutes database 1, played a large role in the intensity of information waves. The demand for content generated by the leader of opinions (including a close circle of actors) and the success of posts (the maximum number of likes, reposts, comments) provide in many respects the actor's symbolic capital, which increases the significance of the text when it is perceived by the network community. Database 2 was generated by ordinary users of the network with a small symbolic capital.

The content generated by the actor with the greatest symbolic capital is the most popular in the network environment, attracting the attention of users. Of course, it should be noted that this provision is true to a greater extent in a certain segment of the network environment and the circle of actors who are influenced by a certain opinion leader.

The selection of adequate communication tools, the successful solution of topical tasks makes it possible for the opinion leader to actively influence public opinion in the network and destroy the boundaries of online and offline communications, translate virtual intentions into real actions. Protests against corruption on March 26, 2018, according to official data, involved 70 000 people (unofficial data suggest a much higher figure) in 100 cities of Russia.

Content identified and analyzed in the course of the research can be characterized as implementing an imperative strategy aimed at the irrational sphere and has a strong emotional charge.

This communication strategy corresponded to the acute need in the Russian society (especially for young people) of social justice, a great protest potential of a destructive type. In particular, despite the clear message of the action (#Dimon will take responsibility), the protest was mostly of an undifferentiated nature («Enough!»).

The main characteristics of the addressee (active actors) were also taken into account when implemented in the content: the social inactivity of the actors is combined with pragmatism, the desire for comfort and sustainable well-being, which must be provided by external forces and not by the participants themselves. In addition, one should highlight the craving for personal fulfillment, hypertrophied self-esteem, and egocentrism against the background of general dissatisfaction. It is indicative that the priorities of the participants of this protest action are in the sphere of hobby (music, cinema, technical innovations, etc.) and sports (cybersport).

Another important feature is the desire for gamification, which is becoming the leading communicative trend. One can talk about gamification as a type of linguistic consciousness (shoe tossing – a symbol that Alexey Navalny took from the film «Wag the dog», toy ducks, behavior with special police forces). In a media text, this is expressed in sarcastic tonality and postmodern discourse.

It can be assumed that the growth of the protest potential in this situation is associated more with the effective work of the Navalny team with a certain group of actors, and not with sustained political activity.

Regarding the motivation of the protest media content actors, the following can be concluded: analysis of the data indicates a lack of stable political preferences and a wide range of political views of the actors. And the explicit support of Navalny's position was expressed only by the youngest actors who make up the cluster «neophytes» (newcomers). «Ideological activists» and «hedonists» expressed a wide range of assessments from indifference to disapproval, disappointment and sarcasm towards the opposition leader.

To a certain degree, this situation can be explained by the lack of a common agenda. In addition to the anti-corruption pathos, the meeting participants were united by a negative attitude towards the federal media, an ethical interpretation of corruption, and a reflection on the formed opposition «youth vs power».

It should be concluded that the perception of content in the network largely depends on the actor, who is identified as the author of the message, in proportion to the volume of his/her symbolic capital and the degree of influence as a leader of opinions, on the relevance of the current communicative situation presented in the network environment and on the effectiveness of the imperative strategy.

References

1. *Martin M* (2018) Dictionary of Digital Pictograms and Glossary for Internet Use and Portable Telephones. Cambridge Scholars Publishing, Lady Stephenson Library, Newcastle upon Tyne.
2. *Sauter T* (2014) «What's on your mind?» Writing on Facebook as a tool for self-formation. J New Media & Society 16: 823–839. doi:10.1177/1461444813495160.
3. *Lipschultz JH* (2014) Social Media Communication: Concepts, Practices, Data, Law and Ethics. Routledge, New York and London.
4. *Verboord M* (2014) The impact of peer-produced criticism on cultural evaluation: A multilevel analysis of discourse employment in online and offline film reviews. J New Media & Society, 16: 921–940. doi:10.15405/epsbs.2018.09.02.61.
5. *Dunbar RIM, Arnaboldi V, Conti M, Passarella A* (2015) The structure of online social networks mirrors those in the offline world. J Social Networks 43: 39–47. doi:10.1016/j.socnet.2015.04.005.
6. *Ryan L.D'Angelo A* (2018) Changing times: Migrants' social network analysis and the challenges of longitudinal research. J Social Networks 53: 148-158. doi.10.1016/j.socnet.2017.03.003.
7. *Tufekci Z, Wilson Ch* (2012) Social Media and the Decision to Participate in Political Protest: Observations From Tahrir Square. J Communication 62(2): 363–379. doi:10.1111/J.1460-2466.2012.01629.X.

8. *Dunbar RIM, Arnaboldi V, Conti M, Passarella A* (2015) The structure of online social networks mirrors those in the offline world. J Social Networks 43: 39-47. doi:. 10.1016/j.socnet.2015.04.005.
9. *González-Bailón S, Wang N* (2016) Networked discontent: The anatomy of protest campaigns in social media. J Social Networks 44: 95–104. doi:10.1016/j.socnet.2015.07.003.
10. *Brusco MJ, Doreian P* (2019) Partitioning signed networks using relocation heuristics, tabu search, and variable neighborhood search. J Social Networks 56:70-80. doi:10.1016/j.socnet.2018.08.007.
11. *Roth C, Cointet J-P* (2010) Social and semantic coevolution in knowledge networks. J Social Networks 32: 16–29. doi.10.1016/j.socnet.2009.04.005.
12. *Latapy M, Magnien C, Vecchio ND* (2008) Basic notions for the analysis of large two-mode networks. J Social Networks 30(1): 31–48. doi:10.1016/j.socnet.2007.04.006.
13. *Murata T* (2010) Detecting communities from tripartite networks. In: Rappa M, Jones P, Freire J, Chakrabarti S. WWW, ACM, pp1159–1160. doi:10.1145/1772690.1772853.
14. *Opsahl T* (2013) Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. J Social Networks 35. doi: 10.1016/j.socnet.2011.07.001.
15. *Kress G* (2010) Multimodality. A Social Semiotic Approach to Contemporary Communication. Routledge, Londres.
16. *Lutkewitte C* (2013) Multimodal Composition: A Critical Sourcebook. Bedford/St. Martin's, Boston - New York.
17. *Wacewicz S, Zywiczynski P* (2017) The multimodal origins of linguistic communication. J Language & Communication 54: 1–8. doi:10.1016/j.langcom.2016.10.001.
18. *Alexander J, Rhodes J* (2014) On multimodality: New media in composition studies. Conference on College Composition and Communication/National Council of Teachers of English, Urbana.
19. *Velkova J* (2018) Studying Emerging Data Practices: Creating a Cultural Biography of Objects Through Using the Web as an Ethnographic Resource. Sage research methods, London. http://methods.sagepub.com/case/emerging-data-practices-creating-cultural-biography-ethnographic-resource.
20. *Crystal D* (2004) Language and the Internet. Cambridge University Press, Cambridge.
21. *Ericsson PF, Haswell RH* (2006) Machine scoring of student essays: Truth and consequence. Utah State University Press, Logan.
22. *Deane P* (2013) On the relation between automated essay scoring and modern views of the writing construct. J Assessing Writing 18(1): 7–24. doi:10.1016/j.asw.2012.10.002.
23. *Weigle SC* (2013) English language learners and automated scoring of essays: Critical considerations. J Assessing Writing 18(1): 85–99. doi:10.24093/awej/vol9no2.11.
24. *Matthews J, Wijeyewardene I* (2018) Exploring relationships between automated and human evaluations of L2 texts. J Language Learning & Technology 22(3): 143–158. doi:10.1177/0265532216676851.
25. *McNamara DS, Graesser, AC, McCarthy P, Cai, Z* (2014) Automated evaluation of text and discourse with Coh-Metrix. Cambridge University Press, New York.
26. *Crossley SA, McNamara DS* (2012) Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication and their relations to judgments of essay quality. J Research in Reading 35(2): 115–135.
27. *Kosinski M, Stillwell D, Graepel T* (2013) Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci USA 110(15): 5802–5805.
28. *Bachrach Y, Kohli P, Stillwell D, Graepel T* (2013) Manifestations of user personality in website choice and behaviour on online social networks. J Mach Learn 95(3): 357–380.

29. *Youyou Wu, Kosinski M, Stillwell D* (2015) Computer-based personality judgments are more accurate than those made by humans. J PNAS 112 (4): 1036-1040. doi.10.1073/pnas.1418680112.
30. *Jacomy M, Venturini T, Heymann S, Bastian M* (2014) ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. J PLoS ONE 9(6): e98679.
31. *Girvan M, Newman MEJ* (2002) Community structure in social and biological networks. Proceedings of the National Academy of Sciences 99(12): 7821–7826. doi: 10.1073/pnas.122653799.