Examining the Historical Development of Techno-Scientific Biomedical Communication in Russia

Alexander N. Marchenko
Science department
St Petersburg State University
St. Petersburg, Russia
a.marchenko@spbu.ru

Vera V. Danilova
Institute of Applied Economic Studies
Russian Presidential Academy of National Economy and Public Administration
Moscow, Russia
maolve@gmail.com

Svetlana V. Popova
Science department
Technological University Dublin, Ireland
St Petersburg State University
St. Petersburg, Russia
svp@list.ru

Wedencley S. Alves
Communication Department
Federal University of Juiz de Fora (UFJF)
Juiz de Fora, Brazil
wedencley@gmail.com

Vera M. Karpova
Faculty of Sociology
Lomonosov Moscow State University
Moscow, Russia
wmkarpova@yandex.ru

Sergey V. Kurushkin
St Petersburg State University
St. Petersburg, Russia
s.kurushkin@spbu.ru

2021 Communication Strategies in Digital Society Seminar (ComSDS) | 978-1-6654-4096-7/20/$31.00 ©2021 IEEE | DOI: 10.1109/ComSDS52473.2021.9422848

***Abstract*—The article focuses on the shift from traditional to techno-scientific biomedical discourse in health-related media communication. The Russian case should be put in context of successes and failures of governmental policies and campaigns, as well as diverse changes in social life that took place in the last decade. To this end, we developed a method for the reconstruction of the timeline that triggered the ongoing techno-scientific progress in Russia based on the decade-long history of healthcare media reports. The dataset includes news reports collected using health and medicine-related keywords from two prominent Russian newspapers (”Rossiyskaya Gazeta” and ”Kommersant”, from 2010 to 2019). The method employs text-mining tools including topic modelling (Latent Dirichlet Allocation-based), term statistics and graphs of keyphrase relationships to analyze the dynamics of media communication in healthcare during the study period. The results of the study confirm that before 2015 the coverage of health-associated topics includes criminal incidents and the “modernization” of the state healthcare system. After 2016, we detected the spread of a new discourse related to drugs e-labeling, new drugs, networks and databases development.**

Keywords*—Mass media, biomedia, biopolitics, government relations, natural language processing

I. INTRODUCTION

The growth rate of mediatization related to digital healthcare solutions including patient-specific data analytics using artificial intelligence and machine learning, applications for selftracking, contact tracing, etc engages science, state, business and the media itself in reflection on the impact of technoscientific changes in culture and society. Health mediatization trends received much attention in literature [1] [2] [3] while their historical prerequisites are underexplored. By historical prerequisites we mean the timeline (preceding events and processes) that tends to be overlooked. Instead, the technoscientific progress is seen as an immediate improvement resulting mainly from advanced democracies and markets. In this paper, we build upon the assumption that the observed shift from traditional to techno-scientific biomedical discourse in health-related media communication occurred for different reasons across countries. Here, communication means the infrastructure (news, social media) for social integration through public discourse [4]. Health communication is defined by A. Endaltseva as a process shaping health knowledge, practices and meanings production [5].

The Russian case is not a simple mirroring of Western practices. It should be put in context of successes and failures of governmental policies and campaigns, as well as diverse changes in social life that took place in the last decade. The present study concerns the reconstruction of the historical timeline (sequence of events and processes) that triggered the current techno-scientific progress in Russia as it is captured by the media today based on media covering of the health agenda considered in the temporal dimension. The timeline's modelling determines the novelty of the study. Our goal is to test the capacity of the timeline to explain the advent of techno-scientific biomedia discourse in the Russian mass media.

The newspapers (”Rossiyskaya gazeta”(”RG”) and ”Kommersant”) are selected by criteria including size, scope, ownership patterns and political leaning. ”RG” is a newspaper published by the Government of Russia as the official gazette, while Kommersant is a private newspaper, one of the five biggest and most reliable media houses in Russia.

The hypothesis is that, despite the prevalent assumptions, the reasons for the shift to techno-scientific biomedical discourse in health journalism differ across countries and it is not exclusively the consequence of advanced democracies and markets [6]. Russia with its developing political system does not simply imitate the Western practices. Successes and failures of governmental policies and campaigns, as well as diverse changes in social life predetermine the historical changes in discourse. Hallin, Figenschou and Thorbjørnsrud [7] also observed that social context shapes health news.

In this context, the relevance of the study is defined by representativeness and usefulness of the historical dynamics of the traditional and digital health discourses in the Russian “hybrid” media system for grasping the BRICS and other emerging states' health reporting patterns.

II. Related work

Examining the historical changes of the Russian health communication in social perspective requires an overview of similar frames used for structuring the global and Russian health communication. As D. Hallin points out, news coverage is intimately related to wider historical transformations in the social organization of health and medicine [3]. The consideration of the proliferation of health news within a larger process of biomedicalization [8] highlights the first (since 1940-50s) and second (since 1985) "social transformations" of medicine following the US example. They are related to the institutionalization and "expansion of medical jurisdiction", as well as to the "integration of techno-scientific innovations". As discussed in [7], the features that positioned health news as an important factor in shaping these broader transformations can be found in the 1960s and 1970s. On the global scale, as examples of these news, the authors consider the reports dating back to the early 60s (public debates over Thalidomide, a drug whose treatment resulted in severe birth defects) and ending with the coverage of the scandals similar to the Vioxx scandal in the U.S. press: a drug that increased the risk of heart attack and stroke. This process includes social movements of women in the 1970s and gays in the 1980s with HIV/AIDS debates and so on [3].

In Russia, there are works that explore, on the one hand, the traditions of science communication including biomedicalization [9], as well as the patterns of soviet-based biotechnology governance in Post-Soviet Russia [10]. Also, Russian researchers investigate the evidences of the development of health communication after the dissolution of the Soviet Union into a new intuitive ecosystem of health communication that is currently emerging in Russia [5]. On the other hand, there are studies related only to the previous decade where the medical significance of Russian population censuses held in the 2002 and 2010 is analyzed, as well as the role of the state regulation in the healthcare field [11], the initiatives of the World Health Organization (WHO), for example the 2011 campaign against obesity (in 2011 in Russia), alcohol abuse prevention programs in 2000-2010s and anti-smoking legislation implementation since 2013 [12]. The self-tracking devices' consumption in Russia is investigated by E. Nim [13] who links the start of the penetration of this phenomenon to the "Quantified self" movement that started in 2007. The above studies are conducted to the need of the investigation of health discourse in Russia in the last decade.

The overview of the previous works emphasizes the growth of interest in the social context of health reporting. However, there is a lack of studies examining the history of health reporting. The particular aspects that are covered in literature and the highlighted general trends including business and digital orientation of health reporting both do not reveal the conditions and social context that make these trends evolve into an emerging media system like Russian.

To fill the gap, in this paper, we propose finding the historical terms that point to the presence of health discourse in Russian news. Based on the decade-long history of healthcare media reports we reconstruct and analyze the timeline that triggered the ongoing techno-scientific progress in Russia. A dataset is collected that includes news reports gathered using health and medicine-related keywords from two prominent Russian newspapers (”RG” and ”Kommersant”, from 2010 to 2019), and text-mining tools are applied including topic modelling, term frequency statistics, keyphrase extraction and the construction of graphs of keyphrase relationships to analyze the dynamics of media communication in healthcare during the study period.

III. The dataset and experimental prerequisites

The dataset was collected automatically for the period from 2010 to 2019 from two prominent Russian newspapers, ”RG” and ”Kommersant” based on the mentions of medicine and health-related terms. In total, it comprised 2,262 texts of news stories published in each month of December. This month was chosen because it represents a "summary" of the year, and it was also in December until 2016 that the annual Presidential Address to the Federal Assembly took place where national healthcare matters were covered. Since texts are collected using a predefined set of keywords, the final sample contains two types of texts. The first type includes texts that discuss medicine and healthcare-related issues. In the second type, the mentions provide an indirect association with the topic: these texts contain spans related to work, criminal incidents, holidays but the main body is about unrelated topics. To capture the relevant topics we extract paragraphs that include keywords for the further processing and analysis. For a similar reason, we do not consider the text of the headings whose role is to attract readers and it tends to contain metaphoric tropes that do not reflect the content and its social significance.

The first manual monitoring of the data taking into account the transformation of the Russian economy with the sanctions' external influence and import substitution directly predicts that two periods should be considered: the first one between 2010 and 2015 and the second from 2016 to 2019. Below we compare the mass media texts produced in 2010-2015 and 20162019. Text mining is performed using several tools, which allows increasing the confidence in the obtained results: topic modelling using Latent Dirichlet Allocation (LDA) with Gibbs sampling [14], word frequency statistics, and an algorithm for keyphrase extraction and the study of keyphrase relationships on graphs. We draw a distinction between techno-scientific and traditional medical discourse indicators.

Traditional medicine is understood in this work as a set of services related to time, place and population. By technoscientific medical discourse we mean the coverage of hightech, digital and commercial components of healthcare. According to Foucault [15], discourse transformation occurs, in the first place, beyond the communication (in forms of production and social relations). In the first transformation we expect a shift from localization (binding to physical location) to universalization of services. To trace this shift, the following keywords are introduced: "country", ”citizen”, ”population”, ”region”, ”man”, ”child”, ”hospital”, ”time”, ”place”, ”center”; and, the markers of the shift are ”budget”, ”ruble”, ”price”, ”number”, ”income”, ”procurement”.

Secondly, the transformation can occur within communication. Here, we expect reorientation from the concept of "modernization" of the state healthcare system (indicated by keywords ”program healthcare”, ”sphere healthcare”, ”healthcare modernization”, ”lifestyle”, ”health protection”) to taking part in the international competition in hi-tech medicine (sample keywords: ”drug”, ”market”,

”rubles”, ”cost”, ”law”, ”project”). Finally, the transformation can occur alongside communication. Here, we expect the disappearance of the topic ”criminal incidents” and a shift to the regulation of the relationship between the state and businesses. However, this transformation will be considered in the future work as resulting from two previous measures.

Below we consider three groups of experiments including word statistics analysis, topic modelling and the construction of keyphrase graphs based on the data sample for the period from 2010 to 2019 collected from the Russian newspapers ”Kommersant” and ”RG”.

IV. Experiment Description and Results

All texts from the dataset were lemmatized and tagged with parts of speech using PyMystem3 1.

A. *Word Statistics*

The frequencies in the first group of words related to traditional medicine such as ”modernization”, ”insurance” and ”fund” drop in both newspapers (see Fig. 1a). At the same time, the mentions of territory names and political authorities also decline including keywords like ”province”, ”area”, ”state”, ”republic”, ”district”, ”region”, ”governor”. In the second group, the growth in the number of keywords, such as ”manufacturer”, ”research”, ”technology”, ”networks” and ”data” is observed (See Fig. 1 b). The same dynamics is characteristic of other keywords in this group including ”procurement”, ”product”, ”Internet” and others.

The frequency statistics analysis grounds the hypothesis that the communication switches from the coverage of political institutions to techno-scientific medicine agenda.

The complementary decrease and increase of frequencies corresponding to these two groups of keywords in so typologically different newspapers as ”Kommersant” and ”RG” are associated with the coverage of the reform of the public healthcare system in Russia, the so-called government’s “modernization” of medicine. The reform started in 2010 when the Law on Compulsory Medical Insurance was adopted. Before the medical industry was able to obtain funds from various sources and in 2010 it was decided to completely switch to singlechannel financing through the Federal Compulsory Medical Insurance Fund [16]. In 2015, the reform’s coverage was finished. The state reform of medicine was one of the historical milestones in the discursive transformation. The Russian mass media turns from the coverage of the state-headed reform of traditional medicine with its localized services toward technoscientific components in healthcare.

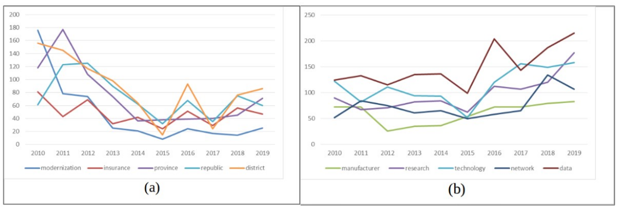

Fig. 1. The dynamics of the keywords ”modernization”, ”insurance”, ”province”, ”republic”, ”district” (a); and ”manufacturer”, ”research”, ”technology”, ”networks” and ”data” (b) in ”RG” and ”Kommersant” per year (in total). Source: Compiled by the authors.

[1] https://pypi.org/project/pymystem3/

B. *LDA Modelling*

This study employs a conventional topic modelling scheme based on LDA [17]. As indicated by the authors in [18], properly trained and evaluated LDA-based topic models are a powerful tool for content analysis in social science that help discover themes overlooked by human coders and are less biasprone. LDA is a particularly popular unsupervised parametric approach that models documents as mixtures of topics and topics as mixtures of words (probabilistic distributions over words). For a given time span (before or after 2015), each document D from the collection (news articles related to healthcare) is assumed to have a distribution over k topics where discrete topic distributions are drawn from a symmetric Dirichlet distribution.

Upon overviewing the recent papers reporting the application of two widely used LDA-based packages, namely Javabased Mallet[2] [19] [20] [21] and Python-based Gensim3 [22] [23] [24], as well as extensive experimentation with both packages on the same data followed by the visualization in LDAvis [25], we settled on the use of the Mallet package [14] [26] compare both tools and point out that both have their strengths and weaknesses. Mallet’s underlying approach relies on Gibbs sampling, which has well-known implications for the runtime complexity [27], because the training process requires to keep the entire dataset in memory. On the other hand, as shown by [19], Mallet performs better than Gensim from the perspective of the coherence value. Roughly, coherence reflects the degree of mutual support between subsets (word sets) within each topic in a topic model. C v coherence [28] used in this paper combines the indirect cosine measure with the NPMI (Normalized Pointwise Mutual Information) and the boolean sliding window and it is reported to be the best measure in terms of runtime and correlation to human ratings [28]. Dataset pre-processing for LDA includes lemmatization with PyMystem34, Russian stopwords removal using NLTK5 and bag-of-words representation using Gensim libraries. The best LDA setup is found as follows. Following [29] we use asymmetric alpha (prior for topic proportions within documents), which combined with symmetric beta (prior for word weights in topic distributions) proved to enhance the quality of topic models. In Mallet, alpha can be optimized each N iterations using the optimize interval parameter equal to N. In the field, it was observed that although frequent alpha optimization in Mallet increases coherence, it affects topic quality due to the growing prevalence of topics with small coverage (topics that are present in few documents)6. Since we aim to capture the most prominent topics, less frequent optimization is given a priority.

The optimal number of topics for each time span is searched for in the interval from 2 to 50 with varying optimize interval (10, 50, 100, 500, 1000). A wider search interval is avoided, because, as observed in [22], selecting too many topics leads to overfitting. Also, a low number of topics ensures the explicability and efficient analysis of each topic. We consider the best number of topics as corresponding to the best c v coherence value as it is done in

[2] http://mallet.cs.umass.edu/

[3] https://radimrehurek.com/gensim/

[4] https://pypi.org/project/pymystem3/

[5] https://www.nltk.org/

[6] https://dragonfly.hypotheses.org/1051

a number of works including [19] and [20]. Also, in case a coherence peak that includes the highest coherence value is shared by the LDA runs with different optimize interval values, it is considered to indicate the best number of topics. [26] and [20] exploit the same idea by comparing the results with varying seed. The resulting optimal number of topics in this study is 30. LDA topics are represented by a distribution of all tokens (words) in the vocabulary and LDAvis shows the percentage of tokens covered by each topic (topic proportions in the dataset).

For the first period, between 2010 and 2015, the model reflects traditional medical discourse in topics 1-4 and 6 in Fig. 2 represented by the keywords ”system”, ”help”, ”treatment”, ”population” in topic 1; ”work”, ”education”, ”possibilities” in topic 2; ”Russia”, ”country”, ”citizen” in topic 3; ”person”, ”child”, ”hospital” in topic 4; ”budget”, ”fund”, ”insurance” in topic 6.

The terms relevant to this discourse cover a significant part of the dataset (Fig. 2): topics 1 - 14.8% of tokens, topic 2 8.8%, topic 3 - 7.5%, topic 4 - 6.7%, topic 6 - 4.5. Only topic 5, whose relevant terms include ”drug”, ”quality”, ”program”, and ”price”, covers 6.4% of the dataset and can be defined as both public and business-oriented but not techno-scientific.

The second period, between 2016 and 2019, is presented in the Fig. 3. The traditional medicine discourse terms are present in topic 2. This topic covers 14.6% of dataset tokens (Fig. 3). The keywords are ”person”, ”hospital”, ”help” and others. Other topics associated with traditional medicine were replaced with the topics of social insurance and benefits programs (topic 1) and private finance and economics-centric discourse with the keywords ”drug”, ”percentage”, ”procurement” and so on (topic 3).

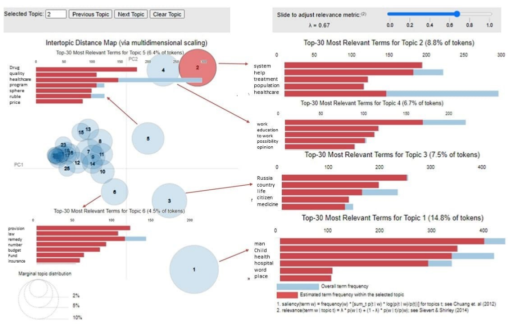

Fig. 2. Terms with the highest frequencies in LDA-produced clusters for both newspapers for the period from 2010 to 2015.

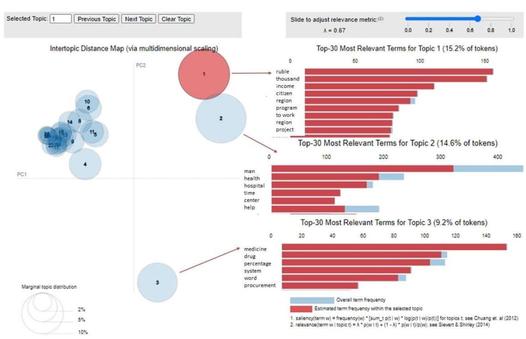

Fig. 3. Terms with the highest frequencies in LDA-produced clusters for both newspapers for the period from 2016 to 2019

C. Keyphrases Modelling

Keyphrases are extracted from texts as multi-word continuous sequences of nouns and adjectives separated by punctuation and words with other parts of speech than mentioned. For each text, the corresponding list of keyphrases is then extended by word sequences of length two or more. These word sequences should be presented in their exact form in any keyphrase extracted from a given processed text and should occur more than once as sub-sequence within other keyphrases extracted from dataset’s texts. When referring to keyphrases, we mean the combination of keyphrases and extracted word sequences.

The proposed algorithm derives from the ideas of the fast and effective keyphrase extraction algorithm RAKE [30] combined with the idea of extracting the most frequent word sub-sequences that roots from category-based search systems that automatically generate named categories using search results [31] [32] [33]. RAKE extracts keyphrases from the text as contiguous word sequences separated by punctuation and terms from the extended stopwords list. Along with the standard list of stopwords, the extended list includes terms that are more often found as adjacent to a keyphrase than inside it and whose frequency is above a given threshold. This list is assumed to be universal for a domain of knowledge: in [30], it is applied to keyphrase extraction from academic papers. However, its construction requires a domain-specific collection of documents with keyphrases labelled by domain experts. Since there is no such corpus available for our study and its construction is out of the scope of this paper, an alternative is found that exploits an important observation on part-ofspeech restrictions made in the keyphrase extraction field. According to this observation, keyphrase extraction quality increases significantly when only nouns and adjectives are allowed within keyphrases [34] [35] [36]. Therefore, all words with parts of speech other than nouns and adjectives are considered as additional separators. Another key observation we exploit is that keyphrases should be multi-word due to the following: out of all keyphrases, single-word keyphrases are only a small group [37], and 2) in single-word keyphrases, the percentage of false positives is much higher than in multiword keyphrases [38]. We exploited and evaluated very similar algorithm ideas in our previous researches [39] [40] [41].

When keyphrases are extracted than we build the graphs of keyphrases co-occurrences for each period respectively. In the present graph, keyphrases are vertices and edges connect phrases that co-occur in more than 3 texts. Vertex size is proportional to the number of keyphrase occurrences in a texts, edge thickness is proportional to the number of texts where the connected keyphrases co-occur. When visualized, only those vertices are displayed that correspond to keyphrases with occurrence frequency no less than the threshold p. The vertices with no edges are not visualized.

The following results were obtained. The graph for the period between 2010 and 2015 is built with $p > 39$, $p > 49$. The following topics can be identified (top-down):

1) prevention (”healthy”, ”lifestyle”) and therapy (”health status”) (p>49);

2) politics and state (”Vladimir Putin”) (p>39);

3) economics ("billion", "million", "thousand rubles") and threats to health resulting from criminal incidents ("harm to health", "the Criminal Code") (p>29),

4) different organizational issues related to the healthcare system's reform ("modernization", "program", "sphere", "system of healthcare") (p>19). This topic is displayed in Fig. 4.

After 2016, a less established discourse is reflected for *p* >29. Here, the prevention topic is not present. The politicization and economization topics remain active and grow when displayed with lower thresholds (>19 and >9). In the third level of the vertex frequency (p>9), a sub-discourse emerges related to the Russian state program of e-labeling of drugs and medicines. It is localized in keyphrases such as "labeling system", "drug labeling", "cost of medicines", "drug market participants" depicted in Fig. 5.

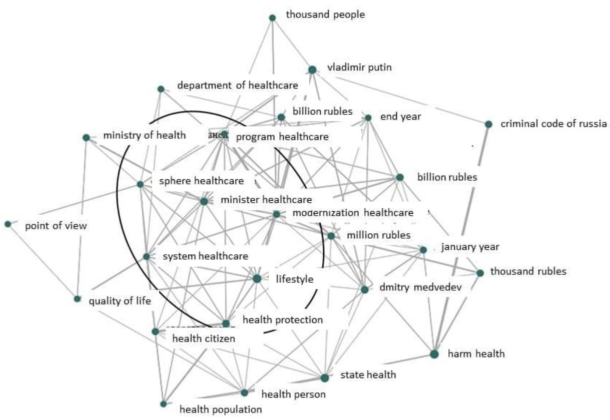

Fig. 4. Visualization for the vertex frequency >19 in "RG" and "Kommersant" between 2010 and 2015.

Fig. 5. Visualization of vertices for p>9 ("RG" and "Kommersant", 20162019).

V. DISCUSSION

Both newspapers share a communication stream that dynamically changes the coverage direction from the discourse on the state reform of the public healthcare (2010-2015) that represents the traditional medicine to techno-scientific, commercial and digital dimensions of healthcare. There are two main trends in the temporal dynamics of term statistics: a decrease in the occurrence of terms related to the medicine "modernization" reform, and an increase in the number of terms related to private initiatives and their regulation in the second half of the decade. This trend could mean not only the centralization of the healthcare system, as one of the negative consequences of the state reform, but also a shift from physical localization of medical services to online healthcare.

The LDA modelling allows identifying two periods in the discourse formation. The first one is well-established, related to state-headed initiatives. The corresponding topics cover a large portion of the dataset. Moreover, we should take into account the generality and heterogeneity of the sample and typological differences between the newspapers. The 2015's discourse switches to the coverage of the advent of the private sector, which replaces the localization of services and introduces universal Internet-driven decisions. The second period is less-established and diverse, its three main topics do not mention the state healthcare reform. The traditional medicine discourse is replaced by reports on social insurance and benefits programs, and it is also significant that the word 'system' moved to the context of the private finance and economics-centric sub-field of communication.

The graph representation of relations between the extracted items allows grasping key features of the discursive space. So, digital and techno-scientific issues are less stable, which is demonstrated by less strong vertex connections. New private finance and economics-centric sub-field of communication focuses not only on power, the state and business get involved in the communication. In state-business communication, the public finance issues prevail over the public health itself. This is evidenced by the example of the Russian state program of e-labeling of drugs. Started in the 2017, this program, as well as the import substitution process, exerted massive influence on the communicative environment.

In summary, the results suggest that the historical terms related to healthcare system's reform and the program of drug e-labeling correspond to the traditional and techno-scientific medicine discourses in the Russian mass media. The healthcare system reform's coverage can be related to influence of the state and propaganda but the coverage of the program of drug e-labeling is characteristic of state-business communication.

VI. CONCLUSION

The aim of this study was to test the timeline of the historical terms related to the advent of techno-scientific health communication in the Russian mass media. We have combined a content analysis approach with an automatic processing of a wide sample of news stories published between 2010 and 2019 in two Russian newspapers, "Kommersant" and "RG". 2,262 news stories were processed using three techniques including basic term statistics analysis, topic modelling and the analysis of keyphrase relationships on graphs.

Our results reveal that although techno-scientific biomedia seems to be a global phenomenon, in each country it is conditioned by different historical contexts. In Russia before 2015, this discourse includes the topic of criminal incidents and “modernization” of the state healthcare system. After 2016, we detected the spread of a new discourse, related to drugs e-labeling, new drugs, networks and databases development.

According to the standard assumptions stemming from the digital research, techno-scientific biomedia arises with the technological progress, as well as with advanced democracies and solid markets. In a country with consolidated economy and an emerging political system as Russia, the adoption of foreign experience can be expected [6]. We contrast the timeline in which the media logic changed with that of the associated events and conclude that the current advances in techno-science are the consequence of the processes in politics (national healthcare reform since 2010, competitive recovery of Russian pharmaceutical enterprises, import substitution), economics (establishment of innovation development institutes, redistribution of public funds, governmental control and drug e-labelling), as well as of the endeavours of media professionals. The discussions in media and digital humanities research are important to perceive the production conditions of the press coverage, and the discourse analysis enables us to understand the discursive changes.

Acknowledgment

We would like to thank next researchers that take part of this project for their effort, commitment and all their work: Nataliya R. Yusupzyanova, St Petersburg State University, St Petersburg, Russia, and Stephanie Lyanie de Melo e Costa, Fundacao Oswaldo Cruz (Fiocruz), Brazil.

The study was funded by Grant of President of Russia, project number MK-968.2020.6.

References

[1] D. C. Hallin. (2020) Media, politics and health: An interview with daniel hallin. [Online]. Available: http://www.revistahcsm.coc.fiocruz. br/english/media-politics-and-health/

[2] C. L. Briggs and D. C. Hallin, *Making health public: How news coverage is remaking media, medicine, and contemporary life*. London, UK; New York, NY, US: Routledge, 2016.

[3] D. C. Hallin, M. Brandt, and C. Briggs, “Biomedicalization and the public sphere: Newspaper coverage of health and medicine, 1960s2000s,” *Social Science and Medicine*, vol. 96, pp. 121–128, 2013.

[4] S. Splichal, “Public sphere and the media,” *International Encyclopedia of the Social Behavioral Sciences (Second Edition)*, pp. 603–609, 2015.

[5] A. Endaltseva, “Communication and health knowledge production in contemporary russia from institutional structures to intuitive ecosystems,” *Strategic Communications in Russia (1st Edition)*, 2020.

[6] I. Kang and A. Marchenko, “The internet is plural: The sociopolitical shaping of the russian and korean “internets”,” *Lecture Notes in Computer Science*, vol. 11551, pp. 272–281, nov 2018.

[7] D. C. Hallin, T. U. Figenschou, and K. Thorbjornsrud, “Biomedicalization and media in comparative perspective: Audiences, frames, and actors in norwegian, spanish, u.k. and u.s.” *The International Journal of Press/Politics*, vol. 13 (3), p. 228–246, Sep. 2020.

[8] A. E. Clarke, J. K. Shim, L. Mamo, J. R. Fosket, and J. R. Fishman, “Biomedicalization: Technoscientific transformations of health, illness, and u.s.” *American Sociological Review*, vol. 68, pp. 161–194, Apr. 2003.

[9] Y. B. Balashova, *Traditions of Science Mediatization in Russia in a Global Context*. Newcastle upon Tyne (UK): Cambridge Scholars Publishing, 2018.

[10] T. Novossiolova, *Governance of Biotechnology in Post-Soviet Russia. Palgrave Macmillan*. Newcastle upon Tyne (UK): Palgrave Macmillan, 2017.

[11] T. I. Frolova, *Zhurnalistika dlja zdorov’ja nacii. Mediakommunikacii i chelovecheskoe razvitie [Journalism for the health of the nation. Media communications and human development] (In Russian)*. Moscow (Rus.): Lomonosov MSU Publ., 2013.

[12] A. Endaltseva and M. Mordvinova, “Health communication and rhetorical challenges for the anti-smoking policy in russia,” *Journal of Communication in Healthcare*, vol. 8:4, pp. 325–334, Oct. 2015.

[13] E. G. Nim, “Student discourse on digital self-tracking: rhetorics and practices,” *Monitoring of Public Opinion: Economic and Social Changes*, vol. 2, pp. 191–211, Feb. 2020.

[14] D. Mimno, H. Wallach, E. Talley, M. Leenders, and A. McCallum., “Optimizing semantic coherence in topic models,” *Proc. of the 2011 Conference on Emperical Methods in Natural Language Processing*, pp. 262–272, Jul. 2011.

[15] M. Foucault, “The birth of biopolitics,” *Ethics: Subjectivity and truth. P. Rabinow (Ed.)*, pp. 73–80, 1997.

[16] E. Kudryavtseva and A. Manuylova. (2015) Reforma za zdravie [reform as prayer service for health] (in russian). [Online]. Available: https://www.kommersant.ru/doc/2841652

[17] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[18] L. Hagen, “Content analysis of e-petitions with topic modeling: How to train and evaluate lda models?” *Inf. Process. Manag.*, vol. 54:6, pp. 1292–1307, Nov. 2018.

[19] R. Zhou, A. Awasthi, and J. S.-L. Cardinal, “The main trends for multitier supply chain in industry 4.0 based on natural language processing,” *Comput. Ind.*, vol. 125, pp. 1033–1069, Dec. 2020.

[20] J. Fang and F. Partovi, “Criteria determination of analytic hierarchy process using a topic model,” *Expert Systems with Applications*, vol. 169, p. 114306, 2020.

[21] S. Cho, C. ung Park, and M. Song, “The evolution of social health research topics: A data-driven analysis,” *Social Science Medicine*, vol. 265, p. 113299, Nov. 2020.

[22] K. Porter, “Analyzing the darknetmarkets subreddit for evolutions of tools and trends using lda topic modeling,” *Digital Investigation*, vol. 26, Supplement, pp. S87–S97, Jul. 2018.

[23] Z. Kastrati, A. Kurti, and A. S. Imran, “Wet: Word embedding-topic distribution vectors for mooc video lectures dataset,” *Data in Brief*, vol. 28, p. 105090, Feb. 2020.

[24] M. Riesener, C. Dolle, G. Schuh, and C. T¨ onnes, “Framework for¨ defining information quality based on data attributes within the digital shadow using lda,” *Procedia CIRP*, vol. 83, pp. 304–310, May 2019.

[25] C. Sievert and K. Shirley, “Ldavis: A method for visualizing and interpreting topics,” *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pp. 63–70, Jun. 2014.

[26] I. Ebeid and J. Arango. (2016) Mallet vs gensim: Topic modeling evaluation report. [Online]. Available: 10.13140/RG.2.2.19179.39205/1

[27] H. Jelodar, Y. Wang, R. Orji, and S. Huang, “Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach,” *IEEE J Biomed Health Inform*, vol. 24, pp. 2733–2742, Oct. 2020.

[37] Z. Liu, P. Li, Y. Zheng, and M. Sun, “Clustering to find exemplar terms for keyphrase extraction,” *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 257–266, 2009.

[28] M. Roder, A. Both, and A. Hinneburg, “Exploring the space of topic coherence measures,” *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM ’15*, pp. 399–408, Feb. 2015.

[29] H. Wallach, D. Mimno, and A. McCallum, “Rethinking lda: Why priors matter,” *Proceedings of Advances in Neural Information Processing Systems (NIPS 2009)*, pp. 1973–1981, Dec. 2009.

[30] S. Rose, D. Engel, N. Cramer, and W. Cowley, “Automatic keyword extraction from individual documents,” *Text Mining: Applications and Theory*, pp. 1 – 20, 03 2010.

[31] D. Zhang and Y. Dong, “Semantic, hierarchical, online clustering of web search results,” *Advanced Web Technologies and Applications, Lecture Notes in Computer Science*, vol. 3007, pp. 69–78, 04 2004.

[32] A. Bernardini, C. Carpineto, and M. D’Amico, “Full-subtopic retrieval with keyphrase-based search results clustering,” *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web*

Intelligence and Intelligent Agent Technology, volume 1, pp. 206–213, 2009.

[33] H.-J. Zeng, Q. C. He, Z. Chen, W.-Y. Ma, and J. Ma, "Learning to cluster web search results," *Proceeding SIGIR'04*, pp. 210–217, 2004.

[34] R. Mihalcea and P. Tarau, "Textrank: Bringing order into texts," *Conference on Empirical Methods in Natural Language Processing*, p. 404–411, 2004.

[35] T. Zesch and I. Gurevych, "Approximate matching for evaluating keyphrase extraction," *International Conference RANLP*, p. 484–489, 2009.

[36] A. Hulth, "Improved automatic keyword extraction given more linguistic knowledge," *Conference on Empirical Methods in Natural Language Processing*, p. 216–223, 2003.

[38] S. Popova and I. Khodyrev, "Ranking in keyphrase extraction problem: is it suitable to use statistics of words occurrences?" *Proceedings of the Institute for System Programming. Volume 26, issue 4*, pp. 123–136, 2014.

[39] S. Popova, V. Karpova, V. Danilova, A. Marchenko, and J. Cardiff, "Topic dynamics study with keyphrases: Livejournal and covid-19 case," *Proceedings of the FRUCT29 - Conference of Open Innovations Association FRUCT*, submitted for publication.

[40] S. Popova, I. Khodyrev, F. Egorov, and V. Danilova, "Academic search. methods of displaying the output to the end user," *Lecture Notes in Computer Science. Proceedings OTM Workshops*, vol. 9416, pp. 476– 480, 2015.

[41] S. Popova, G. Skitalinskaya, and I. Khodyrev, "Estimating keyphrases popularity in sampling collections," *Lecture Notes in Computer Science. Proceedings OTM Workshops*, vol. 9416, pp. 481–491, 2015.