Sentiments in Russian Medical Professional Discourse during the Covid-19 Pandemic

Irina Ovchinnikova
Sechenov First Moscow State
Medical University
Moscow, Russia
ovchinnikova.ig@1msmu.ru

Liana Ermakova
HCTI - EA 4249
Université de Bretagne
Occidentale
Brest, France
liana.ermakova@univ-brest.fr

Diana Nurbakova
LIRIS UMR 5205 CNRS
INSA Lyon - University of Lyon
Villeurbanne, France
diana.nurbakova@insa-lyon.fr

Abstract

Medical discourse within the professional community has undeservingly received very sparse researchers' attention. Medical professional discourse exists offline and online. We carried out sentiment analysis on titles and text descriptions of materials published on the Russian portal Mir Vracha (90,000 word forms approximately). The texts were generated by and for physicians. The materials include personal narratives describing participants' professional experience, participants' opinions about pandemic news and events in the professional sphere, and Russian reviews and discussion of papers published in international journals in English. We present the first results and discussion of the sentiment analysis of Russian online medical discourse. Based on the results of sentiment analysis and discourse analysis, we described the emotions expressed in the forum and the linguistic means the forum participants used to verbalise their attitudes and emotions while discussing the Covid-19 pandemic. The results showed prevalence of neutral texts in the publications since the medical professionals are interested in research materials and outcomes. In the discussions and personal narratives, the forum participants expressed negative sentiments by colloquial words and figurative language.

1 Introduction

Medical discourse exists in communication within the professional community and in doctor-patient communication. In communicative acts, a dominant participant is a healthcare professional who influences patient's social behaviour (Waitzkin, 1989). Researchers of doctor-patient communication described its institutional character, high level of bureaucratic and legislative regulation, peculiar professional ethics (Kuipers, 1989), and sensitivity to cultural diversity (Ferguson and Candib, 2002). Meanwhile, medical online discourse within the professional community has undeservingly received sparse researchers' attention. Healthcare professionals arrange communication to discuss studies and practical issues, to provide colleagues with an opinion summarizing a unique practical experience or a case study (Kuipers, 1989). The professionals communicate through online conferences, professional portals and professional CRM platforms. In professional medical communication, a free discussion and prevalence of facts over sentiments match principles of evidence-based medicine. Meanwhile, professionals are unable to avoid emotions while discussing medical subjects (Franz and Murphy, 2018). The outburst of the Covid-19 pandemics slightly changed the norms in medical professional communication motivating researchers and practitioners to publish raw data and promote a certain treatment (Bavel, 2020). Researchers' emotional involvement into scientific discussion jeopardises objectivity and independence of results and conclusions. The pandemic evokes public interest in medical professional discussions since the media spread controversial news about vaccine development and disease treatment. Due to the interest, a description of medical discourse peculiarities is of importance for communication researchers and social media.

In this paper, we characterise medical discourse within the professional community on a Russian medical portal bringing into focus sentiments and linguistic means to express emotions in medical texts

generated by and for professional physicians. **The objective of our research** is to discover sentiments in texts published on the medical professional forum and to describe a set of linguistic means to express the sentiments and emotions in the professional communication, especially in the context of current pandemics. To the best of our knowledge, sentiments and emotions in the Russian medical professional discourse during the Covid-19 pandemics have not been described yet.

The remainder of the paper is organised as follows. We discuss the related work in Section 2. We present our data collection and the methodology in Section 3. Next, we report our results in Section 4 and provide detailed discussion in Section 5. Finally, we conclude the paper with Section 6.

2 Related Work

In this Section, we describe two groups of works closely related to our study, namely: medical discourse features and sentiment analysis of medical discourse.

2.1 Medical Discourse Features

Medical discourse is a highly regulated institutional professional discourse. It includes various genres for communication among competent professionals and conversations with healthcare customers where a biomedical model of practice co-exists with patient centeredness (Khan, 2019). The structure of the discourse provides a frame to express authority, status, social control in appropriate ways according to the professional ethics, bureaucratic procedures and institutional processes in the society (Kuipers, 1989). A set of medical discourse constituents contains concepts of scientific ideology, restrictions due to professional legal and ethical norms, peculiar genres of professional communication (surgery reports, collective intelligence case reports, medical reviews, etc.), medical semiotics (including terminology), highly structured procedures and sign systems (dress code, observation schedules, etc.), peculiar terminology and patient centeredness (Khan, 2019; Shuravina, 2013; Staiano, 2016).

The structure of texts in the discourse presupposes prerequisites to introduce a problem, a review of familiar approaches to solve it, an explanation of a new basic idea to contribute to the problem solution, and clear outcome to apply into practice (Wilce, 2009; Waitzkin, 1989). The choice of lexical formulas in professional communication depend on the objective of a message (to inform, to clarify, to discuss) and a medical domain the text belongs to (scientific research in particular area, introduction of a new drug, instruction to employ new equipment, etc.). The texts addressed to the professionals require a peculiar structure that facilitates comprehension. Nye with co-authors (Nye et al., 2018) showed the abstract semantic structure of articles published on PubMed. According to the study, the text structure of the medical research articles contains parts for data description (patients, diagnoses), for discussion of procedures (intervention) and outcome (intervention consequences), for deliberation of advantages of a new approach. The discussion part often contains explanations of diagnosis in terms of medical semiotics (Staiano, 2016). The topics of the medical professional discourse are connected to the texts semantic structure and the discourse constituents. The subjects of the texts represent categories corresponding to diagnosis, intervention, treatment, ethics, legislative and bureaucratic regulation, and research.

Sentiments and evaluation of social conflicts are avoided in communication within the professional community. Medical language generally excludes a critical appraisal of the social context (Waitzkin, 1989). The medical text structure in the professional community does not include the content of this type since the texts are produced by and for competent professional physicians (Kuipers, 1989).

A basic idea of a message in medical discourse belongs to scientific ideology because it provides a 'technical' solution to a problem (see a definition of the scientific ideology in (Habermas, 1970). The scientific ideology presupposes objectivism, analytic cognitive style, impersonalisation, generalisation. In medical professional communication, texts transmit the scientific ideology due to discussing 'technical' solutions to problems and enforcing professional domination (Habermas, 1970). While communicating with patients, health professionals follow the structure to get objective data and to solve a problem. Meanwhile, sentiments are required in medical discourse in doctor-patient communication due to its patient centeredness (Ferguson and Candib, 2002; Khan, 2019). However, it is still unclear whether sentiments penetrate into communication within the professional community.

2.2 Sentiment Analysis of Medical Discourse

The analysis of sentiments expressed by experts in medical discourse on professional portals and forums can provide valuable information about predominant subjects and the position of healthcare workers on the important topics. The analysis of sentiments allows researchers to clarify ways and means that are in use to express emotions in the professional community. The professional portal monitoring enables detection of the community attitudes to prominent events and medical abnormalities during the current pandemic. Nevertheless, the results of sentiment analysis in medical discourse are much behind those in other domains (Zunic et al., 2020). There are few domain-specific corpora and lexicons publicly available for sentiment analysis in the healthcare area (Zunic et al., 2020). However, linguistic and content features of medical discourse require a domain-specific sentiment source (Denecke and Deng, 2015). Xu et al. (Xu et al., 2015) carried out citation sentiment analysis in biomedical clinical trial papers in order to decide whether a study can be reproduced or not. Abdaoui (Abdaoui, 2016) studied forums for healthcare professionals to learn classification models used after to identify messages posted by experts in other health forums. He analysed healthcare workers' expertise and reputation and the sentiments they express (emotions and polarity). The work resulted in the construction of French lexicon of sentiments and emotions called FEEL (French Expanded Emotions Lexicon). The results of sentiment analysis are of high importance for healthcare practitioners who would like to improve doctor-patients communication according to the patients' attitudes to their diseases. Rajput (Rajput, 2020) tackles the challenge of reflecting the real mental state of a patient by applying sentiment analysis of texts gathered via various sources including social media. Alnashwan et al. (Alnashwan et al., 2019) identified a set of categories related to the Lyme disease and classified medical forum posts into those categories. Salas-Zárateand et al. (Menasalvas et al., 2017) analysed the positive or negative polarity of document aspects (a review, a piece of news, and a tweet, among others) based on ontologies in the diabetes domain.

Several works were carried out on the classification of patient-authored content into positive, negative and neutral (Ali et al., 2013; Bobicev et al., 2012; Carrillo-de Albornoz et al., 2018; Goeuriot et al., 2012)). Yadav et al. (Yadav et al., 2018) analysed the sentiment based on available information from 'patient.info' which contains opinions about medical conditions self narrated by the users. Multiple forms of medical sentiments can be inferred from users' medical condition, treatment, and medication (Denecke and Deng, 2015; Yadav et al., 2018). However, the implicit sentiment should be also considered (Denecke and Deng, 2015).

Zunic, Corcoran, and Spasic (Zunic et al., 2020) performed a systematic review of the literature about sentiment analysis in medical discourse by studying PubMed and MEDLINE. They found that the majority of data were gathered from social support platforms and social networks serving as online information exchange. In many cases, communities were created around severe and chronic health conditions and discuss treatments (medications, surgery, orthodontic services), prevention, practitioners, or a health care system in general.

To the best of our knowledge, the sentiment analysis has not been applied yet to the discussions held by Russian medical online communities. We carry out the analysis to learn the peculiarities of medical professional communication in Russian and ways to express emotions in medical professional discourse.

3 Materials and method

3.1 Data Collection

We carried out our research based on the RSS data from 29/06/2020 to 11/08/2020 from the professional portal Mir Vracha ('МирВрача'[1]: The doctor's world) which contains various materials useful for healthcare workers. Created in 2011 by a medical team, now this web resource counts more than 100,000 users. Famous national professional resources guide the portal following the principles of evidence-based medicine. The access to the portal content is restricted; however, certified professionals and medical universities students have free access to the portal and are able to publish their materials on its forum. We collected 185 entries in Russian in XML format. Each entry contains title, link, description, and

---

[1]`https://mirvracha.ru`

content. Texts from the fields `title`, `description`, and `content` were concatenated, lower-cased and tokenised with `nltk.word_tokenize`. The `content` includes more than 90,000 word forms. Then we manually analysed the top 10,000 frequent words w.r.t. the aspect of their semantics and correspondence to the medical discourse categories and topics described by Ney et al. (Nye et al., 2018). We excluded stop words and examined the high frequency words (50+ occurrences) to obtain reliable results. While considering categories and topics, we analysed contexts of a frequent word and then attributed the most relevant category to the word. Semantic connections within categories were examined based on the WordNet[2]. For instance, пациент 'patient' is mostly mentioned in discussions of studies including researches of drugs to treat Covid-19; therefore, it belongs to '*Research*'. Thus, the topics were selected based on the distributive analysis of the frequent words. The final number of topics was settled on according to the material represented on the portal.

3.2 Methodology

We performed sentiment analysis of titles and description with the `Dostoevsky`[3] sentiment analysis library for Python. The analysis is based on evaluation of words with emotional connotations in an input text. It takes a text as input and returns a sentiment classification of positive, negative or neutral. Dostoevsky's model is trained on the RuSentiment dataset consisting of more than 30,000 comments from the Russian social network VKontakte[4].

Thus, in this paper we analyse the material collected from the forum Mir Vracha, which contains (a) personal narratives describing participants' professional experience; (b) participants' opinions about pandemic news and events in the professional sphere; (c) Russian reviews and discussion of papers published in international journals in English. We analysed contexts of words: '*Covid-19*, *the coronavirus*, *Covid*, *SARS-COV-2*' attempting to reveal emotional attitude and linguistic means to express emotions. All the authors of the present paper are Russian native speakers and have a linguistic background.

4 Results

4.1 Topics of Medical Discourse Materials Published in the Forum

We discuss the semantic groups of words from the frequency list considering the top 10,000 word forms. Their frequency varies from 465 (это 'this') to 2 occurrences (2,826 words). The most frequent word in the list is пациент 'patient' (582 occurrences). Semantic analysis allows us to distribute the words (frequency 50+) among the topics of medical discourse (see Table 1). We distinguish six main topics of online medical discourse based on the discourse description in (Nye et al., 2018) and (Kuipers, 1989).

Since the portal is aimed at informing the community and providing the healthcare professionals with a platform to discuss news and issues of the current interest, '*Research*' contains more frequent words than any other topic. The '*Diagnosis*' includes names of diseases that had been discussed in the participants' narratives and reviews due to various reasons. The category '*Bureaucracy*' contains words that denote the regulation of professional communication and practice. The high frequency words from the general lexicon correspond to the outcomes of intervention. Nevertheless, Covid-19 appeared to be the most discussed issue because the topic '*Pandemic*' had been referred to by 672 frequent inflexion forms. The Covid-19 pandemic is characterised in 55 texts from 185 entries.

Thus, the topics correspond to the medical discourse constituents and ideology. The most frequent of the top 10,000 words do not refer to sentiments or emotions. Professional communication appears to discuss the prominent topics mostly avoiding emotional expressions.

4.2 Results of Sentiment Analysis

4.2.1 General Data Analysis

We carried out sentiment analysis on text titles and descriptions. The results of the analysis are shown in Table 2. A **title** was classified as *positive* or *negative* when it contains a word with a positive (as

---

[2]`https://wordnet.princeton.edu/`
[3]`https://pypi.org/project/dostoevsky/`
[4]`https://vk.com/`

Table 1: Representativeness of the medical discourse topics in the list of the frequent words

Topics	words belonging to the topic	Word frequency
Research	Author (автор), clinical (клинический), data (данные), patients (пациент) result (результат), study (исследование)	1,428
Diagnosis	Disease (заболевание), illness (болезнь), infection (инфекция), pain (боль), pregnancy (беременность), stroke (инсульт), symptom (симптом), syphilis (сифилис), virus (вирус)	682
Pandemic	Covid-19, coronavirus (коронавирус), sars-cov-2	672
Intervention	Antibodies (антитела), help (помощь врача), MRI (МРТ), therapy (терапия), treatment (лечение)	532
Bureaucracy	Doctor (врач), healthcare (здравоохранение), medical (медицинский), Ministry of Health (Министерство здравоохранения), organization (организация)	522
General lexicon	Death (смерть), health (здоровье), life (жизнь), risk (риск)	496

Table 2: Sentiment classification of the texts from the "MirVracha" portal

Classification	Positive	Negative	Neutral	Skip
Text titles	9	13	156	7
Text descriptions	1	21	160	3

улучшить 'to improve' or улыбка 'a smile') or negative emotional connotation (as опасность 'danger', смерть 'death', боль 'pain'), respectively, or as *neutral*, otherwise. The titles present the approach to the text content, which is of importance for its author. The titles are short. Due to ambiguous associations evoked by the title, it is hard to characterise the content of the text. Thus, the result of sentiment analysis on the text description does not match the classification of its title. The title 'Вымрем, как мамонты и динозавры?' ('Will we die out like mammoths and dinosaurs?') is *neutral*; nevertheless, the text description is classified as *negative* since the author doubted the vaccination efficacy and a chance to overcome the Covid-19 pandemic next year, using words with negative emotional connotations (рок 'doom', умирание 'dying', злой 'evil', etc.). Since the text descriptions include more information about the text content and therefore, represent a better candidate for the sentiment analysis algorithm, we will characterise sentiments in the descriptions.

Sentiment analysis on **text descriptions** examined initial sentences of texts and narratives published on the portal. The algorithm failed to classify 3 texts, while 160 descriptions were classified as neutral. Nevertheless, 21 texts showed negative content; one description was estimated as a text with positive content (see Table 2).

4.2.2 Analysis of Words with Emotional Connotations

Since the sentiment analysis is based mostly on lexical semantics, the words with emotional connotations contributed to classification of the text description sentiment, positive or negative. In Table 3, there is a set of words with emotional connotations that occurred in the initial sentences of the input texts; the words are arranged in alphabetical order, their frequencies are shown in brackets.

The set contains four positive words belonging to the general Russian lexicon. Among words with negative connotations, there are adjectives and nouns denoting negative feelings and emotions ('bad, grim, guilty, offended, sadness', etc.), adjectives and nouns referring to unpleasant experience ('agonizing, complaint, deception, idiotic, pain', etc.), verbs denoting sad and tragic events ('to beat, to die', etc.). Several words belong to legal discourse ('criminal, accused' ets.) since the forum participants discussed legislative regulation of medical practice; the words reveal the topic '*Bureaucracy*' where the discussions of legal issue belong. Nevertheless, in the set of words with negative connotations, words from medical

Table 3: Words with emotional connotations in the text descriptions

Positive class, # descriptions = 1		
Beautiful (красивый 3), a smile (улыбка 6), sweet (милый 4), young (молодой 40)		

Negative class, # descriptions = 21

Accused (обвиняемый 5), acne (угри 2), adiposity (ожирение 30), agonizing (мучительный 2), bad (плохой 11), to beat (бить 1), bleeding (кровотечение 27), complaint (жалоба 49), coronavirus (коронавирус 155, коронавирусный 93) criminal (уголовный 13), danger (опасность 32), death (смерть 100), deception (ложь 2), deceiver (обманщик 2), to die (умирать 8, умереть 44), to die out (вымирать 1), disabled teenager (подросток-инвалид 1), to disappear forever (кануть в бездну 1), disgrace (немилость 2), disease (заболевание 245), doom (рок 2), dying (умирание 1), evil (злой 1), to fall (падать 11), to get sick (болеть 2, заболеть 25), grim (мрачный 2), guilty (виноватый 10, виновен 5), horrible (жуткий 6), horror (ужас 7), idiotic (идиотский 2), illness (болезнь 93), infection (инфекция 109), infertile (бесплодный 2), lamentation (стенание 1), to lie (врать 1), loss (потеря 34), miscarriage (невынашивание 5), moribund (умирающий 2), nightmare (кошмар 3), offended (обиженный 4), pain (боль 100), prison (тюрьма 16), prohibited (запретить 4, запрещать 4), resignation (отставка 4), sadness (печаль 2), scary (страшный 24), threat (угроза 26), trouble (проблема 83), useless (бесполезный 2), to vomit (стошнить 2), worse (хуже 13)

discourse appear to be the most representative group. The group includes names of diseases and negative outcomes of treatment that may be considered as a part of the lexicon of the 'Diagnosis' theme. According to the results of sentiment analysis, the participants who shared their negative experience in personal narratives produced the descriptions with negative content. Since Covid-19 represents a new disease, the healthcare professionals have been mostly discussing research results and publications in medical journals since the pandemic outbreak.

4.3 Analysis of Texts about Covid-19

Texts about Covid-19 were classified as negative (7 texts) and neutral (48 texts). The studies did not evoke any emotional conversations on the forum since the texts devoted to consideration of the Covid-19 origin and treatment were mostly classified as neutral.

In the *neutral* texts, the forum participants used terminology (рецептор ACE2 'ACE2 receptor', гипертензивный 'hypertensive' etc.), while considering the effect of drugs on the coronavirus (азитромицин 'azithromycin', etc.). They referred to publications discussing other pandemics ('испанка' "Spanish" influenza') and the most recent articles reported a research of the coronavirus (статья в Nature Microbiology 'the article in Nature Microbiology', etc.). Thus, the content of the texts about Covid-19 reflects the basic topics of the professional medical discourse: '*Research*', '*Diagnosis*', '*Intervention*', '*Outcomes*', '*Contagion*', '*Bureaucracy*' (including legal and ethic regulation).

In the *negative* texts, words with negative connotations denote predictions as to the outcome of the pandemic, the evaluation of the social behaviour and the bureaucratic regulation, names of disease (see Table 3). Thus, the negative content mostly covers the '*Diagnosis*' topic.

Moreover, a text with negative content includes a critical appraisal of the social context due to '*Black lives matter*' protest during the outburst of Covid-19 in New York: "An article summarising editorial articles about the link between COVID-19 and racism from Nature, Science and three leading medical journals was published on June 13. As they are all very similar in terms of meaning and content, I'll cite the most illustrative example: "*A number of people who see the protests as possible drivers of more covid-19 cases have suggested that it is hypocritical of doctors to support anti-racism protestors now when weeks earlier they denounced large crowds due to the risk of spreading covid-19. Framing anti-racism*

protests against the control of covid-19 is a false dichotomy." [5] [6]. The author explores boundaries of the medical ethics when social protests endangers people health. Healthcare professionals must follow the ethical norms regardless of their social sympathy and political opinions. Large crowds on the streets of American cities facilitate spreading the virus therefore doctors and practitioners have to condemn the protests; however, condemnation would show sympathy to racism. Words лицемерие 'hypocrisy', ругать 'to scold' manifest negative attitude to professionals who support the protesters; nevertheless, the author keeps distance from evaluating the events emotionally. However, the text shows importance of including critical appraisal of the social context during the current pandemic.

Contexts of words referring to the current Covid-19 pandemic reveal a set of linguistic means to express an attitude to the event. Nevertheless, the means rarely occurred in the publications. Direct evaluation is represented by words that denote feelings and emotions (печаль 'sadness'; пренебрежение 'disdain' etc.) or refer to a negative status, consequences, events (бездумный 'harebrained'; угроза 'threat'; немилость 'disgrace', etc.). The forum participants preferred to express irony while discussing the coronavirus (В современной медицине не бывает новостей не про COVID-19. Самые нековидные вести чуток да соприкасаются с коронавирусной заразой 'In the current medicine, all news are associated with Covid-19. Even non-Covid-19 news touches on the infection a little bit). The ironic attitude is manifested by Russian colloquial words (чуток 'just a bit', зараза 'infection' and 'bad person'). Discussing drugs and vaccine development, physicians used figurative language. Comparisons and metaphors allow diminishing the threat by emphasizing its temporal character (коронавирусная напасть 'coronavirus scourge') and associating the contagion with a fire (спичка при пожаре 'playing with fire', 'add fuel to the fire'). The idioms and tropes influence the perception of information about the virus generating cultural associations with sentiments expressed in the Russian literature masterpieces (русская рулетка 'Russian roulette'; сколько патронов в барабане 'how many cartridges are there in the cylinder').

As we have mentioned earlier, the forum participants did not intend to express emotions; they appealed to emotions and used figurative language when they were interested in attracting readers' attention to current problems in fighting the pandemic. The forum participants cited literary texts and used colloquial expressions, figurative language, and idioms. The typical attitude to the subject of the discussion, which the forum participants expressed in their texts, is ironical evaluation of the professional experience.

5 Discussion

The materials published on the Mir Vracha portal represent the medical professional discourse as it reveals itself in communication among healthcare professionals. To the best of our knowledge, we present the first results and discussion of the sentiment analysis of Russian medical discourse online. The texts published on the forum reflect the intention to consider an individual practice against the backdrop of the current research materials. Due to the interest in the research, the most frequent words belong to '*Research*' and '*Diagnosis*' discourse categories.

The topics of the physicians' publications on the professional portal differ from the popular topics in the communities on social media. Zunic et al. (Zunic et al., 2020) showed that members of communities were interested in consideration of 'Intervention' (medications, operations, etc.), while the Russian healthcare professionals are focused on the '*Research*' and '*Diagnosis*' topics. The professionals did not discuss the subjects that were the most popular themes on Twitter. Abd-Alrazaq et al. (Abd-Alrazaq et al., 2020) found out four main themes of the tweets devoted to the Covid-19 pandemic: origin of the virus; its sources; its impact on people, countries, and the economy; and ways of mitigating the risk of infection. In

---

[5]The original: "Тринадцатого июня вышла заметка, собравшая воедино редакторские статьи о связи COVID-19 и расизма, появившиеся в журналах Nature, Science и трех ведущих медицинских журналах [3]. Они довольно похожи по смыслу и содержанию, поэтому я приведу лишь самый яркий пример: "Группы людей, которые рассматривают протесты как один из возможных двигателей распространения случаев COVID-19, отмечали, что это лицемерие со стороны врачей, поддерживающих протестующих сейчас, когда лишь несколькими неделями раньше они ругали большие скопления людей за распространение COVID-19. Противопоставление между протестами против расизма и контролированием COVID-19 - ложная дихотомия"

[6]The referred article was published on BioEdge website: https://www.bioedge.org/bioethics/covid-19-driven-by-racism-say-major-journals/13469

July-August 2020, the Russian healthcare professionals had been concentrating on the virus origin and the ways to recognise and to treat the disease. The difference in the set of themes between the professional community and ordinary users of social media reveals boundaries of the medical professional discourse. Within the discourse, its semantic categories dominate in communication among competent professionals who are motivated to obtain more medical information and to offer 'technical' solutions.

The topics of the publications (see Table 1) represent the discourse categories of medical professional discourse. The discourse category contains the semantics relevant to medicine and healthcare associated with the medical professional language and the texts structure to arrange the professional communication (Wilce, 2009; Nye et al., 2018; Staiano, 2016).

Peculiar terminology including Latin terms is an essential constituent of medical discourse. Complicated connections of terms reflect interactions of concepts and references to a subject, which are to be examined and described in various conditions. Names of the diseases are the high frequency words in our materials. Other terms were also in use on the portal; however, the forum participants used different terms according to the subject they discussed, so each term occurred twice or once. Nevertheless, the terminology caused the algorithm of sentiment analysis to classify the publications on the Mir Vracha as neutral. Meanwhile, in the communities on social media, texts patients suffering from cancer received essential differences in the sentiments (emotions and polarity: see (Abdaoui, 2016)). The difference in the emotional attitude to the content between the professional community and the community of the patients reveals different communicative intentions and objectives as well as the medical discourse boundaries.

Along with the recent development of narrative medicine, the importance of professional social control brings doctor-patient communication into focus in medical discourse studies (Franz and Murphy, 2018). Meanwhile, online communication within the professional community shows patient centeredness through the frequency of words referring to patients and their experience ('patient, pain, therapy' etc.). The patient centeredness is not obvious due to the scientific ideology of medical professional discourse, which brings into focus the '*Research*' category. Since the scientific ideology supports finding a solution to current and forthcoming problems, the participants of the forum on the Mir Vracha portal discussed facts and case studies preferring to avoid emotional outbursts.

Sentiment analysis results show reasonable evaluation of the texts published on the medical professional portal. Absence of emotional words in the list of high frequency words confirms the unessential contribution of the emotional expressions to the content of the medical professional discourse texts. The texts contain emotions and attitudes expressed by means of figurative language. Meanwhile, doctor-patient communication on the Internet shows instances of metaphors, metonyms, and idioms in written messages on forums (Kharitonova et al., 2019). The professionals use a similar set of linguistic means to express emotions and attitudes in their communication within the community. The physicians expressed negative emotions caused by their professional experience during the current Covid-19 pandemics appealing to irony as a tool to diminish the threat and the seriousness of the problem they must find a solution to.

6   Conclusion

We carried out sentiment analysis on titles and text descriptions of materials published on the Russian professional medical portal Mir Vracha (90,000 word forms approximately). To the best of our knowledge, we present the first results and discussion of the sentiment analysis of Russian professional online medical discourse. The texts were generated by and for physicians and reflect the peculiarities of medical professional discourse. The medical professional discourse revealed in communication within the professional community is focused on the semantic categories '*Research*', '*Diagnosis*', '*Intervention*', '*Bureaucracy*'. The categories provide linguistic means to verbalise the professional activity and bureaucratic regulation of medical practice. Based on the results of sentiment analysis and discourse analysis, we described the emotions expressed in the forum and linguistic means the forum participants used to verbalise their attitudes and emotions while discussing the Covid-19 pandemic. The attitudes and emotions show negative sentiments manifested by colloquial words and figurative language. In further research, it is necessary to widen online medical professional communication materials for sentiment analysis in order to describe a thesaurus of sentiment linguistic means.

References

Alaa Abd-Alrazaq, Dari Alhuwail, Mowafa Househ, Mounir Hamdi, and Zubair Shah. 2020. Top concerns of tweeters during the covid-19 pandemic: Infoveillance study. *J Med Internet Res*, 22(4):e19016, Apr.

Amine Abdaoui. 2016. *French Social Media Mining: Expertise and Sentiment. (Fouille des Médias Sociaux Français: Expertise et Sentiment)*. Ph.D. thesis, University of Montpellier, France.

Tanveer Ali, David Schramm, Marina Sokolova, and Diana Inkpen. 2013. Can I hear you? sentiment analysis on medical forums. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 667–673, Nagoya, Japan, October. Asian Federation of Natural Language Processing.

Rana Alnashwan, Humphrey Sorensen, Adrian O'Riordan, and Cathal Hoare. 2019. Accurate classification of socially generated medical discourse. *Int. J. Data Sci. Anal.*, 8(4):353–365.

J.J. Van et al. Bavel. 2020. Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, 4(5):460–471, May.

Victoria Bobicev, Marina Sokolova, Yasser Jafer, and David Schramm. 2012. Learning sentiments from tweets with personal health information. In Leila Kosseim and Diana Inkpen, editors, *Advances in Artificial Intelligence - 25th Canadian Conference on Artificial Intelligence, Canadian AI 2012, Toronto, ON, Canada, May 28-30, 2012. Proceedings*, volume 7310 of *Lecture Notes in Computer Science*, pages 37–48. Springer.

Jorge Carrillo-de Albornoz, Javier Rodríguez Vidal, and Laura Plaza. 2018. Feature engineering for sentiment analysis in e-health forums. *PLOS ONE*, 13(11):1–25, 11.

Kerstin Denecke and Yihan Deng. 2015. Sentiment analysis in medical settings: New opportunities and challenges. *Artificial Intelligence in Medicine*, 64(1):17 – 27.

Warren J Ferguson and Lucy M Candib. 2002. Culture, language, and the doctor-patient relationship. *Family medicine*, 24:353–361.

Berkeley Franz and John W. Murphy. 2018. Reconsidering the role of language in medicine. *Philosophy, Ethics, and Humanities in Medicine*, 13(1), June.

Lorraine Goeuriot, Jin-Cheon Na, Wai Yan Min Kyaing, Christopher Khoo, Yun-Ke Chang, Yin-Leng Theng, and Jung-Jae Kim. 2012. Sentiment lexicons for health-related opinion mining. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, IHI '12, page 219–226, New York, NY, USA. Association for Computing Machinery.

J. Habermas. 1970. Technology and science as 'ideology.'. In J. Habermas, editor, *Toward a Rational Society*, pages 79 – 126. Boston: Beacon.

Munir Khan. 2019. The perspective of medical communication on the biomedical model of practice and patient centeredness: A review of the language of medical case presentation genre. *International Journal of Linguistics, Literature, and Translation*, 2:71–80.

N.V. Kharitonova, S.V. Baryshnikova, and O.V. Monastyretskaya. 2019. Peculiarity of lexical component of professionally oriented discourse in the internet (based on medical forum). *Philology. Theory and Practice*, 12:331–334.

Joel C. Kuipers. 1989. "medical discourse" in anthropological context: Views of language and power. *Medical Anthropology Quarterly*, 3(2):99–123.

Ernestina Menasalvas, María del Pilar Salas-Zárate, José Medina-Moreira, Katty Lagos-Ortiz, Harry Luna-Aveiga, Miguel Ángel Rodríguez-García, and Rafael Valencia-García. 2017. Sentiment analysis on tweets about diabetes: An aspect-level approach. *Computational and Mathematical Methods in Medicine*, 2017.

Benjamin Nye, Junyi Jessy Li, Roma Patel, Yinfei Yang, Iain Marshall, Ani Nenkova, and Byron Wallace. 2018. A corpus with multi-level annotations of patients, interventions and outcomes to support language processing for medical literature. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 197–207, Melbourne, Australia, July. Association for Computational Linguistics.

Adil Rajput. 2020. Chapter 3 - natural language processing, sentiment analysis, and clinical analytics. pages 79 – 97.

Ludmila Sergeevna Shuravina. 2013. Medical discourse as a type of institutional discourse. *Journal of Health and Social Behavior*, 328(37):65–67.

K.V. Staiano. 2016. *Interpreting Signs of Illness: A Case Study in Medical Semiotics*. Approaches to Semiotics [AS]. De Gruyter.

Howard Waitzkin. 1989. A critical theory of medical discourse: Ideology, social control, and the processing of social context in medical encounters. *Journal of Health and Social Behavior*, 30(2):220–239.

James M. Wilce. 2009. Medical discourse. *Annual Review of Anthropology*, 38(1):199–215.

Jun Xu, Yaoyun Zhang, Yonghui Wu, Jingqi Wang, Xiao Dong, and Hua Xu. 2015. Citation sentiment analysis in clinical trial papers. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2015:1334—1341.

Shweta Yadav, Asif Ekbal, Sriparna Saha, and Pushpak Bhattacharyya. 2018. Medical sentiment analysis using social media: Towards building a patient assisted system. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May. European Language Resources Association (ELRA).

Anastazia Zunic, Padraig Corcoran, and Irena Spasic. 2020. Sentiment analysis in health and well-being: Systematic review. *JMIR Med Inform*, 8(1):e16023, Jan.