Possibilities of artificial intelligence in assessing the impact of potentially dangerous texts in modern news discourse: problem of statement

Irina Karabulatova[1]*

[1]Peoples' Friendship University of Russia (RUDN University), 115093 Podolskoe Schosse 8/5, Moscow, Russian Federation

Abstract. The relevance of the stated problem reflects the study of the "friend-foe" dichotomy, which is clearly represented in the modern news discourse, since it reflects the most significant problems for society: migration, the COVID-19 pandemic, crime, various confrontations, problems of socially vulnerable citizens, etc. The subject of the research is to Refine the parameters for evaluating potentially dangerous texts for the subsequent creation of a library of software modules for theming and classifying news messages, including using AI technologies. Hypothesis: the proposed parameters of the system of interpretation of potentially dangerous text increase the chances of determining the prognostic level of the degree of propensity to illegal actions, so the creation of a digital library will help to quickly analyze the levels of potential dangers for the recipient. The use of digital technologies for psycholinguistic assessment of potentially dangerous texts optimizes the search and tracking of such texts, contributing to the development of measures to ensure the safety of the human psyche in conditions of massive impact on the recipient in order to change his personal attitudes. The author raises the problem of creating a single digital platform for evaluating such texts, noting the need for linguistic priority when creating semantic markup, which will allow us to qualitatively rank potentially dangerous texts. Such work requires the application of interdisciplinary efforts of specialists in the fields of linguistics, psychology, mythology, history, sociology, political science, cultural studies, mathematics, computer science and Digital Humanities. The practical value is unquestionable, since psycholinguodiagnostics of a person does not correlate with the potential danger of texts produced by such a person in society.

1 Introduction

As a result of epidemiological and geopolitical transformations, new challenges arise for the entire human civilization, as a result, being in the mode of forced self-isolation, emergency situations and quarantine measures, an increasing number of people experience maximum stress on the individual psyche. In this regard, the issues of preservation of the human psyche are becoming more relevant than ever.

The use of expressive means of influence by the majority of mass media aimed at attracting the attention of recipients breaks down the traditional protective psychoemotional

---

* Corresponding author: Karabulatova-is@rudn.ru

mechanisms of the individual, resulting in an increase in deviations and disorders [1, 2, 3, 4] due to the so-called social schizophrenia [5].

Modern news discourse actualizes archetypal fears and stereotypes that have developed in society [6, 7]. Such publications are suggested to be considered as potentially dangerous, since they destroy the objective perception of reality.

A potentially dangerous text is a cross-section of ethnolinguistic, socio-political, psycholinguistic and linguoculturological problems of a particular locus with an explicit or hidden "Overton window".

The problem of "Overton Windows" [8] is particularly acute in modern news discourse, being an important means of influencing recipients in the context of information warfare [9, 10, 11]. A well-established thesis States that a socially dangerous person is an extremist person, so his public statements and published texts can give a fairly complete description of the purpose and intentions of the addressee of speech [12, 13].

The current level of development of human civilization actualizes the achievements of psycholinguistics, focused on the study of the personality realized in the text, and pragmalinguistics, which analyzes the parameters of the impact on the recipient without taking into account the emotional and psychological characteristics of the individual.

Today we see an urgent need for scientific papers that would reveal the specifics of understanding and impact of explicit and/or potentially dangerous texts on the psyche of recipients.

2 Materials and methods

This work expands the empirical base of discursive linguistics, LSP theory and practice, motivology, emotive linguistics, conflictology, political science, media linguistics, journalism, and psychology, whose interests include consideration of the problem of the influence of emotions on language.

The purpose and hypotheses of the study led to the formulation and solution of the following problems.

1) analyze ideas about the ontological essence and psycholinguistic mechanisms of the phenomenon of potential danger in the text;

2) develop a concept of the phenomenon of potentially dangerous text parameters;

3) carry out predictive monitoring of modern publications in order to predict the possibility/ impossibility of a potential danger to humans and society.

4) development and testing of technologies for evaluating and improving the effectiveness of textual examinations, which are of fundamental importance for further high-tech developments in the field of motivology and psycholinguistics.

Such large-scale tasks predetermined the choice of main approaches to project development within the framework of news discourse analysis, namely:

1) psychological, historical and cultural studies;

2) proper linguistic (grammatical, textual, stylistic, etc.);

3) semiotic (actually semantic, syntactic, pragmatic);

4) philosophical (structuralist, post-structuralist, deconstructivist);

5) logical (argumentative, analytical);

6) informational (communicative and rhetorical).

To create semantic markup for AI training [14, 15], it is necessary to conduct preliminary methods such as interpretive, content analysis, structural and semantic analysis, implicative scaling, evaluation, and motivational-target analysis. These methods are associated with the methodology of social communication: cohort group surveys, testing, and in-depth interviewing [16]. The methodology of social communication provides feedback that helps determine how strong the effect of a particular dangerous text was [17].

Preliminary screening of open sources on the Internet (it was performed manually) showed that over the past five years there have been more than 2000 socio-political conflicts [18]. In the first half of 2020 alone, 94 social and labor conflicts were recorded [19], with over a thousand ethnically motivated conflicts of varying intensity. [20].

The research material is formed in the course of a continuous sampling of the development of politicization of news discourse, revealing potential and/or obvious dangers to society, describing deep internal social changes. These transformations are usually hidden from the surface view by various ideological attitudes and stereotypes.

The empirical base of the methodology is based on a secondary analysis of Russian and foreign research on this issue.

3 Revue and discussion

Identification and measurement of potentially dangerous types of news discourse on the Internet on topics and risks is carried out using interdisciplinary tools (psycholinguistics, computer science, psychology, cognitive science, political science, sociology, cultural studies, journalism) [21, 22, 23].

The problem of attribution of potentially dangerous text is due to the fact that in order to determine the existence of an offense, it is necessary to clearly define the target orientation of the analyzed text.

Based on this result, the motivational basis for the formation and promotion of potentially dangerous discourses in the media space is revealed [24, 25].

Psycholinguistic expert assessment of a potentially dangerous text is based on identifying the nature of the impact on the addressee of speech, namely: transformation of the value picture of the world, prompting to any actions and/or actions in the virtual world or in reality.

The mere fact of publication and/or disclosure of information is not a justification for the statement about the danger of the text.

Potentially dangerous text can be either spoken or written. In addition, it can be either a single utterance or a set of utterances and/or texts [26].

Establishing criminal intent is not our task, since it is a field of legal competence. However, due to the fact that the so-called LegalTech is currently actively developing as a digital assistant to lawyers, the combination of two such auxiliary systems could act as a regulator, helping to avoid unnecessary legal liability. However, there is still a risk of excessive confidence in the ability of AI to make any court decisions and restrictive measures, which can also become a potential danger to society.

To identify potentially dangerous text, we identify the author's purpose of a text message. This goal is always expressed directly or indirectly. It is identified by analyzing the language tools used in the message, taking into account the type of information and the specifics of the communicative situation. From the point of view of linguistics, semantic analysis is of leading importance, where the analysis of the meaning of what is said plays an important role in determining the dangers of the text [27]. Researchers make a fair distinction: meaning, content, and actual meaning [28], but, as a rule, these distinctions are not taken into account when evaluating potentially dangerous texts.

The key is to understand what are expressed and unexpressed language means, and what are the characteristics? In addition, potentially dangerous text reflects the addressee's subjective attitudes, which are aimed at changing the attitudes of the recipient of the message. Therefore, disputes may arise about the level of compliance of interpretations and the analyzed text [29], so the task is to assess the degree of reliability of the interpretation of one or another party [30].

A special place is occupied by the analysis of works of art in the context of potential danger. in this case, the use of literary tools is required.

Semantic markup for creating an algorithm for evaluating potentially dangerous text should establish seven main types of publication values, whose characteristics need to be established, namely: 1) a call to a particular action (about which the question is asked); 2) propaganda of specific views and ideologies (for example: a negative attitude towards someone/ something; proclamation of racial/ethnic/social/ gender or other superiority; indication of inferiority of a person on any basis); 3) arousal of negative feelings and emotions (discord, hostility and hatred) on the part of recipients to any person/ group of persons/ region/ country; 4) humiliation of human dignity on the grounds that need to be answered; 5) accusing the person indicated in the question of the actions that are asked in the question; 6) justification of actions and views indicated in the question; 7) threat of violence against the person/ group of persons indicated in the question. As a rule, these questions are answered manually by experts, but since the space of questions is strictly limited, they can be displayed in the plane of digital analysis. At the same time, the digital algorithm faces the same question as a regular expert: to identify the meaning and signs of its implementation in a specific text. Therefore, questions are formulated with an emphasis on the presence/ absence of specific signs of meaning expression.

It should be noted that the so-called "extremist" meanings are controversial among researchers [31], since the very concept of "extremism" belongs to the legal field, going beyond the competence of a psycholinguist.

4 Results

When we consider the issues that face digital expert assessment, first of all, we are faced with the problem of creating a methodology for creating marked-up samples of texts that would combine psychological, linguistic, historical and cultural, socio-political, and legal aspects.

Clustering potentially dangerous discourses in the news discourse on topics and risks allows you to create semantic markup for the work of experts. The linguistic assessment loop itself is designed to form marked-up samples that are used to train artificial intelligence methods in the signal system and semantic information search (in the monitoring loop). therefore, we propose to create an information technology for automatically identifying potentially dangerous types of information and news discourse based on understanding the psycholinguistic and socio-cultural mechanisms of their impact. We can divide the main actors or forces that drive the entire society towards a particular type of public behavior into four main groups. We present these groups as a table.

Table.1. The main actors of the potential dangers of contemporary news discourse

Societies	Publicities	Experts	AI	The main actors of potential dangers in modern news discourse
non-profit associations, social groups, national and cultural autonomies, individual state structures and departments, government and self-government bodies, mass media, etc.	specific individuals, passionaries, statesmen, public figures, artists, bloggers, journalists, politicians, and so on.	Professional analysts, accredited experts, researchers-scientists, field officials, recognized specialists, professionally oriented journals, programs, websites, etc.	Targeting, Parsing, digital funnels, CRM systems, etc.	

At the same time, we must understand that all these actors are implemented in texts that have an impact on the entire society. We consider the texts themselves, rather than the authors ' personalities, which allows us to unify approaches to automatic identification of potential dangers in the news discourse.

Many of the analysis steps that are obvious to an expert are not those for AI, so we need to rank each step fractionally so that the digital algorithm for evaluating potentially dangerous text works as closely as possible to the analysis performed by a human. Based on this, we identify the following six functional steps in solving expert tasks:

1) setting the specific value of the unit used in the analyzed text segment (type: Who does the author refer to by the word "Russian"/ "Ukrainian"/ "Caucasian"/ "disabled" (etc.) in this text? How does the author define the word "protest"/ "boycott"? What actions does the author of the statement "the First option is to boycott everyone" call for? etc.);

2) attribution of the form of the method of expressing the author's opinion from the position of its admissibility in the public sphere (type: What style of speech does the statement belong to? Is this expression acceptable within the literary style?);

3) establishment of lexemes with unambiguous negative meaning in the text (reference to murders, suicides, protests, violence, pressure, bullying, gaslighting, ostracism, public affront, etc.);

4) identification of the communicative purpose of the author's statement (for example: is the analyzed statement of the author a call or not? If "Yes", what are the signs that this is confirmed in the analyzed text? Is the offer being analyzed a threat? If "Yes", then on what grounds? – etc.);

5) establishing in the text the motivation for certain actions on the part of recipients (namely: whether the analyzed text contains calls for riots, murder, violence, and other antisocial forms of behavior (specify which ones); whether the text contains statements that justify illegal and/or antisocial / deviant forms of social behavior);

6) identifying the focus of a key statement, the entire text or a collection of texts on the impact on the recipient/ group of recipients, etc. (examples of clarifying questions: What feelings does this text evoke in the reader/ readers? What causes the emotional tension of the analyzed text? To which group of people is this text addressed and on what grounds? Is

this text, in the submitted materials, open and understandable to the "average" person, or is its content encoded and directed to a specific person or group of people? What is the General meaning of this text? What is the key fragment for determining the semantic direction of the entire text?).

The first three tasks are purely linguistic. When solving subsequent problems, the methods of cognitive science, psychology, and pragmatics are used, but in any case, the results of linguistic analysis are used as evidence.

However, each step contains intermediate small steps. For example, in order to determine the emotional tone of a text, it is necessary to distribute the analyzed tokens depending on the reflection of emotional States. When we consider potential dangers, we appeal to negative emotions, although the scale of tonality contains all the shades from minus to plus. In this connection, E. A. Nushikyan ranked 16 classes of emotional States, and within each of them the researcher placed emotives according to the degree of increase in the manifestation of emotional saturation [32]. Po, E.A. Nushikyan, the scale of manifestation of emotional saturation looks like the following emotional chains before affectation, both in their extreme expressions of positive and negative:

1) contentment → pleasure → joy → delight → exultation → ecstasy → happiness;
2) affectionate → sympathetic → caring → affectionate;
3) approval→ admiration → encouragement → praise;
4) bewilderment → surprise → amazement;
5) sadness → sorrow → yearning → the mountain → despair;
6) discontent → annoyance → resentment → anger → rage;
7) worry → concern → anxiety → confusion → fear → fright → horror;
8) uncertainty → doubt → hesitation → indecision → confusion → despair;
9) regret → chagrin → bitterness → resentment;
10) embarrassment → apology → excuse → awkwardness → remorse → shame;
11) dislike → antipathy → disgust → contempt;
12) friendly banter → joke → mockery → irony → humiliation → sarcasm;
13) chide → reproach → reprimand → accusation;
14) disapproval → objection → protest → outrage;
15) distrust → wariness → suspicion;
16) Reminder → warning → caution → threat [32].

Forensic pathopsychologists rely on the classification of V. I. Galunov [33], who distinguishes the following gradation of the emotional state: joy → sadness → excitement → depression → rage → fear → apathy → norm. The mobility of language semantics, which creates prerequisites for the trend of language variability and manifests itself in speech in updated semantic modifications, in the language system – in the asymmetry of signs, causes the appeal to more complex than the word units that determine the semantics of the language in action. Therefore, we see in the text an increase in negative ratings, and as a result, negative emotions in the recipients of the text, leading in General to the phenomenon of social schizophrenia, i.e. schizophrenic-like reactions in normal people under the influence of intense external factors. By E. A. Nushikyan [32], the most frequent emotions have a fairly high level of their identification (95-98%), so the negative emotions that are embedded in the analyzed document are clearly recognized by recipients.

A potentially dangerous text activates the activation of the provocation mechanism in the recipient's consciousness, which includes the passage of three stages: 1) subconscious attitudes that are not recognized by the addressee; 2) the activity stage with subsequent awakening of awareness; 3) the involution of awareness and acceptance of the status of automation of actions [34]. Therefore, potentially dangerous text is characterized by encoding types: soft, steep, or hard. Thus, a high degree of manipulation of the speech

recipient's consciousness is achieved, when the author of a potentially dangerous text provokes the transformation of the text recipient's communicative behavior in order to actualize psychosocial States that are not typical of the provoked recipient in order to achieve the goal of managing both the individual and society as a whole.

5 Conclusion

Potentially dangerous texts are United by a common theme of impact on recipients, which is due to the acute social orientation of publications in the news discourse. The situation of digitalization and the coronavirus pandemic actualizes public attention to texts distributed on the Internet both on official media platforms and in social media (social networks, blogs, streaming channels, vblogs, etc.). This variety of information is usually contradictory, where there is not only a standard text structure represented as a linear sequence of letters, but also a hypertext structure with the necessary hyperlinks. All this makes it necessary to analyze a large array of documents, which it would be advisable to analyze using AI methods [35, 36]. At the same time, the creation of a program to ensure the safety of the human psyche in the face of information pressure will serve as the basis for ensuring the health of the country and human civilization as a whole. The absence of analogues in analytical digital examinations of potentially dangerous texts and the creation of such a platform in Russian will be the first step for the subsequent creation of a multilingual platform "Pro et contra".

Acknowledgments

Reference

1. F. S. Safuanov, T. N. Sekerazh. Activities aimed at encouraging children to engage in suicidal behavior: opportunities for forensic psychological examination. Psychology and law, 2, 33-45 (2017)
2. A.V.Zeynalbdyeva. The Diversity of manifestations of protest vote. Scientific sheets of BelSU. Series: Philosophy. Sociology. Right. 10 (231), 104-108 (2016)
3. I. S. Karabulatova. The Manipulation of the Public consciousness through Russian Press Coverage of Events around the Paris Peace Conference, 1918-1920. The Paris Peace Conference (1918-1920) and Its Aftermath: Settlements, Problems, Perceptions. Edited by Sorin Arhire and Tudor Roșu. London: Cambridge Scholars Publishing, 209-227 (2020)
4. 4. A. A. Shteba. Constitutive signs of provocative communicative behavior. Language and law: actual problems of interaction. actual problems of interaction: materials of the III-th International scientific and practical conference. Ed. V. Yu. Melikyan. Rostov n/D: Donizdat, 3. 43-51 (2013)
5. Th. Charernboon, J. Patumanond. Social cognition in schizophrenia. Mental Illness 2017; volume 9:7054, 16-19.

6. E. Ermakova, M. Jilkisheva, G. Fayzullina, I. Karabulatova, Kh. Shagbanova. The media end fiction: postmodernist discourse of contemporary terrorism in the context of apocalyptic rhetoric. Central Asia and the Caucasus, 17, 2, 61-69 (2016)

7. V. Shaklein, I. Mitrofanova, S. Deryabina, S. Mikova, I. Karabulatova. The impact of ethnopsycholinguistic portrait of Serbian and Russian youth on the formation of Russian-Serbian relations. Advances in Social Science, Education and Humanities Research, International Conference on Man-Power-Law-Governance: Interdisciplinary Approaches (MPLG-IA 2019), 374. Atlantic Press, https://www.atlantis-press.com/proceedings/mplg-ia-19/125925326 (2019)

8. I. S. Karabulatova, I. P. Savchuk. Information and marketing war in modern advertising discourse: transformation of gender stereotypes. Scientific review. Series 2: Humanities, , 1-2, 51-65 (2019)

9. V. V. Barabash, A. S. Bobryshova, O. I. Lepilkina, I. S. Karabulatova. The specific of the interpretation of the tragedy over Sinay (October 31, 2015) as a focus of information attention of the "VKONTAKTE" social network. Astra Salvensis. 6. 1, 289-310 (2018)

10. M. Ebrahimi, A. Hossein Yazdavar, and A. Sheth. Challenges of Sentiment Analysis for Dynamic Events. Affective computing and sentiment analysis. Editor: E. Cambria, 70-75, (2017)

11. V. V. Barabash, E. A. Kotelenets, I. S. Karabulatova, M. Y. Lavrentyeva, Yu. S. Mitina. The confrontation between the Eastern and Western worldviews in the conceptual space of the information war against Russia: the genesis and evolution of the terminological apparatus. Amazonia investiga, 8 (19), 246 – 254 (2019)

12. L. M. Balabanova. Forensic pathopsychology (issues of determining the norm and deviations). Donetsk: Stalker, 432 (1998)

13. I. Karabulatova. The specifics of individual behavior under the influence of the work of a recruiter of extremist orientation in an electronic digital society. Advances in Social Science, Education and Humanities Research, 374. MPLG-IA 2019, Atlantic Press, https://www.atlantis-press.com/proceedings/mplg-ia-19/125925295 (2019)

14. A.V. Sukhareva, K. V. Vorontsov. "Building a complete set of topics for probabilistic thematic models", Intelligent systems. Theory and applications, 23:4 7-23, (2019).

15. Y. Liu, H. Yu and Q. Yang, Introduction to the Special Issue on Federated Machine Learning. IEEE Intelligent Systems, 35, 04, 3-4 (2020)

16. F. Lambotte, E. Wathelet. Le cycle texte-conversation en tant que dispositif accompagnant le changement de "catalyseur" des tensions dans la restructuration de l'organisation sociale. Communication and management. Researches – Pratiques, 107-124 (2016)

17. E. S. Palekha, E. A. Sadykova. Comments from users of social networks: once again to the question of aggressive text. Philology and culture. 2, 98-101 (2018)

18. I. S. Karabulatova, K. V. Vorontsov. Digital linguistical migrationology: the possibilities of artificial intelligence in the study of migration processes. II World Congress in real and virtual mode "West-East: crossing cultures", 2-6 October 2019, Japan, Kyoto Sange University, Scientific and practical materials of the 2nd world Congress in Japan 2019, Kyoto: Kyoto Sange, 2, 760-766 (2019)

19. E. Malysheva. The coronavirus provoked an increase in the number of labor conflicts in Russia. Almost four million workers were laid off during the pandemic. 7x7, 05.08.2020 (2020). https://7x7-journal.ru/articles/2020/08/05/koronavirus-sprovociroval-rost-chisla-trudovyh-konfliktov-v-rossii-za-vremya-pandemii-uvoleno-pochti-4-mln-rabotnikov

20. I.S. Karabulatova, K. V. Vorontsov Digital linguistic migratology and monitoring of potentially dangerous texts about migration. Migration bridges in Eurasia: new approaches to the formation of migration policy in the interests of sustainable

development. Materials of the XI International scientific and practical forum. Edited by S. V. Ryazantsev, M. N. Khramova, 189-204 (2020)

21. I. Karabulatova, Kh. Vildanov, A. Zinchenko, E. Vasilishina, A. Vassilenko. Problems of transformation matrices modern multicultural identity of the person in the variability of the discourse of identity Electronic Information Society. Pertanika, 25(S). Jul., 1-16.(2017)

22. S. V. Ionova. Aspects of research of a written text as an object of linguistic expertise. Bulletin of the Volgograd state University. Series 2, Linguistics, 16 (2), 28-38 (2017)

23. C.Vásquez, D. Schoeneborn, & V.Sergi. Summoning the spirits: Organizational textsand the (dis)ordering properties of communication. Human Relations, 0018726715589422.doi :10.1177/0018726715589422 (2015)

24. E. Riloff et al., Sarcasm as Contrast Between a Positive Sentiment and Negative Situation. Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), 704–714 (2013)

25. I. S. Karabulatova, P. V. Barsukov, I. V. Akhmetov, O. V. Mamatelashvili, F. F. Khizbullin "Network Wars" as a New Type of Devitiaon Processes in the Modern Electronic and Information Society in the Context of Social and Economic Security. MJSS, 6, 6, S.3, 150-159 (2015)

26. V. B. Chelpanov. Psychological and linguistic expertise of the text, Moscow: international Academy of lie research, 420 (2018)

27. S. M. Zenkevich. Author's designations of prosodic characteristics of emotional speech in the novel "Wives and daughters"by E. Gaskell. Professionally-oriented language teaching: reality and prospects: materials of the scientific practice conferences. Saint Petersburg: Saint Petersburg state University of Economics publishing house, 248-253 (2020)

28. M. M. Polekhina, V. A. Limonzeva, I. S. Karabulatova, M. S. Vyhrystyuk. The Evolution of the Concept of "Terror" / "Terrorism" in Modern Scientific Knowledge as a Factor in Ensuring the Security of Modern Society. In the: Astra Salvensis, 6 (12), 695-704 (2018)

29. I.A. Novikova, A.L. Novikov, M.A. Rybakov. Psychological and linguistic features of the Russian language acquisition by international students. RUDN Bulletin. Series: Psychology and Pedagogy,1, 61-65 (2015)

30. N. N. Panchenko. Reliability as a communicative category: autoref. Diss.... Ph. D. – Volgograd, (2010)

31. I. Mkrtumova, A. Dosanova, I. Karabulatova, V. Nifontov. The use of communication technologies of oppose political-religious terrorism as an ethnosocial deviation in the contemporary information-digital society // Central Asia and the Caucasus.17, 2, 54-61 (2016)

32. E. A. Nushikyan. Prosodic organization of emotional speech. Abstract. Diss. Doctor of Philological Sciences. Leningrad, 32 (1987)

33. V. I. Galunov. Speech, emotions, personality: problems and prospects. Speech, emotions, personality, 3-13 (1978)

34. Yu. B. Gippenreiter Introduction to General psychology, Moscow: AST, 420 (2009)

35. S. Choukah, P. Theophanidis. Emergence and ontogenetics: Towards a. communication without agent. Social Science Information, 55(3), 286-299 (2016)

36. A.Esuli, A. Moreo and F. Sebastiani, Cross-Lingual Sentiment Quantification. IEEE Intelligent Systems, 35, 03, 106-114 (2020)