Dovhan O. V.,
PhD in Philology,
Doctoral student at the Department of Slavic, Romance and Oriental Languages,
Drahomanov Ukrainian State University

POLITICAL INTERNET DISCOURSE: FEATURES OF TEXT ANALYSIS IN THE CONTEXT OF MACHINE LEARNING

Summary. The article analyzes the concept of discourse in the context of linguistic science and presents the structure and peculiarities of its interpretation, as well as possible misinterpretations that affect its functionality. In addition, the study highlights the identity of political activity (in particular, political (and Internet) discourse) in the context of machine learning tasks and prospects for possible study. In particular, the author focuses on the ontological and axiological ranking of politics as a phenomenon of human existence and highlights the features of political linguistics with its inherent anthropocentrism, expansionism, and functionality. The article also presents the peculiarities of analyzing the texts of political online discourse in the context of machine learning tasks: in particular, the specifics of representation of certain political events, parameterization of emotional support, etc. are highlighted.

Thus, the study of the peculiarities of analyzing political online discourse in the context of machine learning is relevant in the current international situation: in particular, the study of the originality of the process of analyzing the texts of the latter in the context of the specifics of the artificial neural network. In addition, such a study is of particular importance due to the hybrid nature of the russian-Ukrainian confrontation, which is characterized by several fake, distorted, actually false, and other data actualized in the media space.

The author emphasizes the inseparability of analyzing the texts of the above-mentioned discourse and combating the use of such false data (disinformation, misinformation, propaganda) constructed by the russian federation to destabilize the situation in Ukraine. The article proposes a solution to this issue through the development of machine learning algorithms capable of adapting to the fluid political online discourse and fully analyzing communication in it in all its forms and forms. Thus, the study of texts of political online discourse in the context of machine learning is directly related to the possibility of creating artificial neural network models capable of in-depth analysis.

Key words: discourse, political online discourse, analysis of online discourse, text analysis, machine learning, artificial neural networks.

A general statement of the problem and its connection with important scientific or practical tasks. Discourse (from the late Latin *discursus* – reasoning, proof) is a multivalent term that denotes speech in the general sense, as well as its use together with the system of concepts that it is conditioned by [1]. That is why the Cambridge Dictionary defines the above concept as "the use of language for communication in oral or written form" [2], while the Explanatory Dictionary of the Ukrainian Language

positions it as "a coherent text, a superior unity" [3]. Therefore, within the framework of Wittgenstein's "language game" [4, p. 5], discourse can be positioned as a complex object, the nature of which is in the mutual projection of linguistic and behavioral rules.

At the same time, linguistic research is characterized by an original understanding of discourse: it is usually correlated with such concepts as speech, utterance, text, and the activity mediated by it (monologic/dialogic), represented by a communicative act. This produces an interpretation of the concept of discourse in the context of several phenomena with an empirical nature of existence and a simultaneous distortion of the understanding of its essence in terms of localized or fragmented representation of the latter.

Thus, the concept of discourse is positioned in linguistic science as a way of communication and data exchange mediated by textual structures. In turn, this marks its understanding as a socio-cultural product that is representative of the specifics of speech activity in a particular context. That is why, in our opinion, it is productive to interpret discourse as a text in the process of becoming before the interpreter's mental view in all the differentiated nature of this concept (we are talking about the complex structure of the latter: it consists of sentences or their fragments) [5]. In this sense, it is advisable to understand Internet discourse *as all forms of communication on the Internet*, represented by blogs, forums, e-mail, social networks, various messengers, etc.

Frequently, the content of the above discourse is concentrated around a kind of "cornerstone" – *the core concept* (or *topical discourse*), and the logical content of individual sentences (*discursive components*) is called propositions. The role of the latter is to build connections that ensure the functioning of logical relations within and the stability of their existence for the sake of the unity of all elements: *disjunction* – "or", *conjunction* – "and", *implication* – "if-then", etc. The mechanism of realization of the aforementioned discourse is as follows: the interpreter (a native speaker of the language poly system and, accordingly, of the discourse) accumulates the generalized sense of elementary sentences while updating new data segments. The latter becomes a kind of "announcement" of the next information (the sentence that will be interpreted shortly) within the actualized information (the result of the previous, current, or intermediate interpretation).

In the context of the above, political Internet discourse is noteworthy as a special type of discourse, namely, the originality of the connection between politics and language, which is the communication basis of any activity. This means, first of all, that no political regime can exist without communication, i.e., the language polysystem in the political function. In this sense,

politics is considered a certain communicative modus of human activity, in which the language poly system plays a functional role (relatively speaking, a toolkit for political action) because politicians actualize the language poly system for informing (propaganda: appeals, instructions, persuasion, etc.; lawmaking: creation, amendment, etc. of regulatory documents, etc.).

At the same time, the originality of political activity about other types lies in its predominantly discursive nature: in turn, it is associated with the linguistic essence of political actions. We are talking about the growing influence of communicative and semantic components on the existence of political processes, which, in turn, produces an increasing role of semiotic parameterization. Thus, the latter has become the basis for the representation of political activity as exclusively linguistic, and for modern political science, it has given grounds for considering language not as a representative of such activity, but as a component of the political field. The latter has naturally led to the possibility of a reverse perspective, in which politics is seen as a specific form of speech activity, realized by institutionalized speech acts of adapting ontological reality to words.

The above is connected with the ontological and axiological ranking of politics as a phenomenon of human existence: it is not without reason that the disclosure of its essence is represented in several special (sociological, philosophical, political science) dictionaries. Modern political linguistics is characterized by the following features: a) *anthropocentrism* (linguistic personality becomes a point of reference when studying phenomena); b) *expansionism* (the linguistic field of research includes related issues, expanding it); c) *functionality* (language is studied in the process of everyday life, i.e. in action); d) *explanatory* (the desire to describe linguistic phenomena along with explaining the essence of what is happening) [6], etc.

We are talking about the aforementioned performativity (the speech essence of political actions) and autoreferentiality (self-presentation, presentation of data about oneself) of institutional facts. The latter constructs a social ontology, causing the text to shape it under the guise of representing reality. This is in line with the approach to the language poly system formed in the twentieth century within linguistics and philosophy, in which language was considered not only as a tool for describing ontological reality but also as a mechanism and form of its construction. The latest russian disinformation and manipulative campaigns unfolding in the temporarily occupied territories are illustrative in this regard: In particular, a) *monopolization of the religious space* (delivery of propaganda literature to churches in Zaporizhzhia region); b) *inspection of our fellow citizens' phones* (checking Telegram channels subscribed by compatriots in Donetsk region) c) *the formation of pseudo-volunteer movements* (enrollment of Ukrainian children in the "Eaglets of russia" – a children's division of the kremlin-funded organization in the Kherson region); d) the creation of a network of free "district newspapers" (which are looking for local collaborators in Luhansk region to fill the position of chief editors) [7], etc.

Thus, the linguosemantic characteristics of the language poly system in the political function can be supplemented by linguocommunicative ones if we add another element to R. Jakobson's scheme – the political one. The latter can be positioned as an inverse magic function, within which an absent or inanimate "third party" will turn not only into the addressee but also the addressee of the message. At the same time, in the process of political communication, not only the addressee and the addressee are institutionalized but also the communicative process itself, which acquires the features of a semiotic counterpart of itself. The choice of the latter as the basis for the magic function is justified by similar processes of deployment. Similarly, by default, we expect that a speech act will cause a change in ontological reality, reconstructing it by the sense it contains. Therefore, its (the act's) participants are endowed with the appropriate power, the source of which is not myths, constructs, etc., but the social stratification of society.

Thus, the process of communication can be positioned as one of the pivotal ones for human ontology, since it is characterized by a special type of causal relationship with the language polysystem. We are talking about the actualization of several functions of the linguistic phenomenon: communicative, magical, identification, epistemological, thought-creating, and others. That is why, in the context of the above, political Internet discourse is the most representative of the way a society lives, and the specifics of the functioning of narratives in it, which is related to the universal nature of political communication. We are talking about the actualization of relevant spheres of human ontology within it, worldview role in terms of interests and target beliefs, etc. [8].

The main sources of studying political language, given the number of correlations between the latter and all forms of communication on the Internet (including the press, radio, television, etc.) and directly institutional (including leaflets, parliamentary debates, speeches at rallies, political party documents, etc.) discourse, are political Internet discourse in all its diversity of forms and types. The latter is a syncretic construct that arose as a result of the merger of two discourses – political discourse and Internet discourse itself, which led to the originality of its nature.

Analysis of recent research and publications on this topic, highlighting previously unresolved parts of the general problem to which this article is devoted. The problem of the peculiarities of analyzing political online discourse in the context of machine learning has a multilayered nature, which, first of all, produces its actualization in several interdisciplinary and integrated works. Thus, O. Bolkariova [8] argues that the analysis of political discourse (including political Internet discourse) is a new way of studying politics, which, in turn, creates the need for new tools (methods, criteria, models of text, and context analysis). We agree with the researcher: as stated above, machine learning and artificial neural networks as a means of solving the problem of new tools for studying political discourse (including political online discourse) are the solutions to the problem of new tools for studying political discourse.

The analysis of the linguistic and pragmatic features of pleonastic units formed based on the principle of redundancy in political speech is devoted to the work of B. Zayniddinovna [9]. The scientist highlights the specifics of this speech, studying in detail the originality of pleonastic combinations actualized within its framework. The researcher emphasizes that, despite the apparent redundancy and semantic repetition, pleonastic combinations of political speech have a specific methodological task from a pragmatic linguistic point of view.

The study of the possibilities of using artificial intelligence in the context of the development of Internet discourse is contained in A. Robertson, M. Maccarone [10], where the authors update

the narrative analysis used for the main trends in the use of such intelligence. The researchers argue that narratives are critical because they resist the impulse of rapid and, in some cases, unlimited technological progress, and also offset apocalyptic narratives about artificial intelligence familiar from mass culture. The study contributes to the understanding of the socio-technical perceptions of a category of actors that have been insufficiently developed to date. In contrast, the work of G. Longo [5] is aimed at studying the correlations between Internet discourse, family, and related concepts, which allows us to explore how rethinking the Internet as a social institution produces a deeper understanding of the everyday life of this phenomenon.

The fundamental role of political Internet discourse in the context of social networks for all spheres of modern life is highlighted in the work of A. Chaves-Montero [11]. This study focuses on the analysis of tweets related to social services published on Twitter (currently – X) during the election campaigns in Spain in 2015-2019. In their work, the researchers used the *Gephi* tool to observe how these messages are spread on the aforementioned network. The results show that users use Twitter for informal communication during the election period, focusing on messages, condemnation, and positive evaluation to increase their visibility and influence.

Instead, a fundamental analysis of the socio-cultural features of comments under political news that turn into a public forum in the context of online discussions is presented in C. Küchler et al. [12]. The study of the specifics of the political reaction to the prospect and reality of refugees entering their local communities on the example of comments in political online discourse using a structural and thematic model is presented in the work of C. Kelling, B. Monroe [13]. The results of the linguistic and stylistic analysis of German-language comments in the context of political Internet discourse are presented in D. Kuchina [14], in which the author analyzed and identified the distinctive features inherent in the above genre.

The work of G. Ramya [15] continues the study of commentary texts based on the study of online political and property data for their further moderation. The author highlights the features of training a neural network model using such algorithms as LSTM and BERT (Bidirectional Encoder Representations from Transformers), as well as GPT2 (Generative Pre-Trained Transformer 2), which helps in text generation by increasing the size of the data set for training in various classification models. The above produces the ability to train such a model in such a way that it can give accurate results during text classification, as well as in the process of identifying false/true comments contained in the data source.

The study of the role of the perception of threats from other cultures caused by terrorist attacks and criminal events in public discourse and support for right-wing radical parties is presented in the work of F. Giavazzi et al. [16] The authors propose to divide Twitter users (currently – X) into electoral districts in Germany and use machine learning to calculate the degree of textual similarity between the tweets they produce and the tweets of the accounts of the main German parties. The study of Internet discourse continues with the work of J. Govers et al. [17], in which the authors state that the influence of social networks is directly related to the problem of protecting the security of Internet users and freedom of expression on the Internet. The researchers propose and highlight a unique approach to the contextual analysis of hate speech, which integrates all other studies in this area.

The analysis of Internet discourse texts in the context of hate speech is studied in the work of M. Fazil et al. [18] in which the authors emphasize that social networks are a means of facilitating the process of real-time communication for users, while simultaneously producing problems of hate speech and false data. The researchers present an automatic method for detecting hateful messages, namely a convolutional BiLSTM neural network for classifying the latter. This neural network model updates existing methods for representing words in a multi-channel environment with multiple filters with unequal kernel sizes to localize semantic relationships in different windows.

The study by K. Bagchi, E. Banker, I. Ogunleye [19] is devoted to the representation of the work with the problem of false data mentioned in the previous work, the authors prove the falsehood of the assumption that there is no content moderation by Internet platforms. Thus, scientists have identified three current trends in the dissemination of such content: medical misinformation, hate speech, and child exploitation. In each of these areas, they highlight recent measures taken by policymakers to regulate how online platforms moderate these undesirable forms of content. They then detail actions that platforms are already taking in these areas to combat the challenges posed by online discourse.

The problem of hate speech in political Internet discourse continues to be studied by V. Wanniarachchi et al. [20], who propose a methodological framework that uses a new mixed-methods approach to identify hate speech patterns from large text arrays collected from social networks. The authors actualize computer-mediated quantitative methods, including natural language processing (NLP) methods: sentiment analysis, emotion analysis, and topic modeling, as well as qualitative analysis of online discourse.

According to P. Kar and S. Debbarma [21], the main problem of text analysis for hate speech detection using machine learning is the reduction of the problem to a binary classification that does not take into account either the thematic focus or the target nature. Thus, since there is no unified annotated dataset and no scientific study that could give an idea of the relationship between the features of such a language, the available methods usually study only one or two of its features at a time. In addition, these methods are not effective for multilingual environments where most conversations are mixed, so scientists have proposed using a hybrid diagonal recurrent neural network (RNN) with diagonal shunting (FE-DGRNN) to detect hate speech and analyze sentiment in multilingual mixed-code texts. The experience of creating and training a transformational neural network model for analyzing emotions and sentiments in Hungarian political texts is presented in I. Üveges, O. Ring [22], in which the training data were manually annotated texts of parliamentary speeches from 2014 to 2018, rich in various emotions.

The work of Y. Lupu et al. [6] classified the types of hate speech and hostility on mainstream and marginal online resources: the authors conducted a supervised machine learning analysis of 7 types of hate speech in online discourse on 6 interconnected online platforms. The researchers found that offline events, such as protests and elections, are often followed by an increase in the number of types of hate speech in online discourse that seem to have little to do with the main event. The urgency of localizing hate speech in political online discourse is evidenced by the study by F. Jafri et al. [23], in which the authors present a new dataset called IEHate, which contains 11,457 manually annotated Hindi

tweets related to the Indian Assembly election campaign from November 1, 2021, to March 9, 2022. The above dataset, according to the authors, is a valuable resource for researchers and practitioners working to develop and evaluate methods for detecting hate speech in languages with limited resources.

One of the important components of machine learning (at the beginning) and natural language processing by an artificial neural network (later) with preliminary linguistic analysis is the aforementioned sentiment analysis, the features of which are presented in V. Derbentsev et al. [24]. The authors argue for the relevance of the latter in the context of working with Internet discourse, presenting a comparison of deep learning models of deep artificial neural networks (convolutional (CNN), with long short-term memory (LSTM) layers and bidirectional LSTM with CNN layers (BiLSTM-CNN)) within their study. E. del Valle, L. de la Fuente [25] study political debates in online discourse (in particular, social media), in which the authors consider the above phenomenon in the context of its derivatives: first of all, hate speech. The scholars note that such derivative problems are of pivotal importance for social development due to the degree of their influence on public opinion and, accordingly, the communication strategies of political parties, state institutions, media corporations, and lobbyists.

The problem of an automated methodology for analyzing political online discourse is also addressed in the work of E. Fournier-Tombs, G. di Marzo Serugendo [26], in which the authors used a quantitative approach based on machine learning to measure the quality of discourse in political discussions on the Internet. The researchers used the DelibAnalysis framework, which made their research methodology accessible, representative, and reproducible. This system updated a simplified version of the DQI to train a classifier, which was then used by the researchers to predict the discourse quality of any uncoded comment in a particular political discussion on the Internet.

The work of P. Törnberg [27] presents an innovative toolkit, namely, large language models (LLM), which the author positions as a universal method of text analysis in the social sciences. The latter, according to the scientist, is due to the simplicity, speed, accessibility, etc. of their actualization and wide possibilities for text analysis tasks: from annotation to classification of sentiment analysis and critical analysis of online discourse. The researcher in his paper discusses in detail the stages of the above analysis (software installation, API setup, data upload, development of analysis prompts, text analysis, and verification of results). The results and the course of the study are illustrated by the author's accomplishment of the difficult task of identifying populism in political texts.

Therefore, the study of the peculiarities of analyzing political online discourse in the context of machine learning is relevant in the current international situation: in particular, the study of the originality of the process of analyzing the latter's texts in the context of the specifics of the artificial neural network. In addition, such a study is of particular importance due to the hybrid nature of the Russian-Ukrainian confrontation, which is characterized by the existence of several fake, distorted, actually false, and other data actualized in the media space.

In turn, the existence of the above data (misinformation, propaganda, etc.) in political online discourse is productive due to the favorable nature of the latter. Thus, political Internet discourse is by its nature an environment in which various groups, individuals, and others express their position on certain events, and subjectively interpret them, thus influencing public opinion [18]. First of all, this raises the question of the authenticity (reliability, truthfulness, veracity, etc.) of the data that exists in this type of discourse, becoming the core problem of its functioning and research. In turn, the latter requires the development of innovative tools for processing its data to identify unreliable segments and protect against manipulative paradigms of the enemy, which constitutes the unresolved tasks of the problem we are studying and the relevance of studying text analysis in the context of machine learning.

Formation of the purpose of the article (statement of the task). *The purpose* of the article is to consider the peculiarities of analyzing the texts of political Internet discourse as an object of research. *The subject* is the specifics of the above phenomenon in the context of machine learning and the work of artificial neural networks as an innovative tool of linguistic science.

Presentation of the main research material with full justification of the scientific results obtained. The above-mentioned is the pivotal role of the language poly system in the existence of ontological reality, within which it is a mechanism and form of constructing the latter. Thus, the modes of actualization of the linguistic poly system corresponding to various social functions naturally produce the creation of new types of reality, or rather, ideas about the latter. At the same time, it is obvious that the above-mentioned *representations of reality* (or *ideas about reality*) are *axiologically determined*: first of all, it means that such a representation does not exist outside the language poly system, and cannot carry a socially significant sense under such conditions.

In this context, political Internet discourse can be positioned as a special type of discourse precisely because, in our opinion, it includes all the others. In turn, this is due to the specific functionality of such discourse: for example, participants (voters, candidates, politicians, etc.) within the communication process actualize the language poly system mainly for some kind of influence, often manipulative [16]. Thus, the specificity of this construct allows us to talk about the productivity of using machine learning as a powerful tool for its analysis. However, the success of updating the latter naturally correlates with the thoroughness and quality of the preliminary linguistic analysis of the text, which must be done manually to update its algorithm in machine learning, and finally in the work of an artificial neural network.

It is advisable to take into account the following components inherited from the political discourse itself: a) *subject matter* (representation of the life of a group or person); b) *original function* (the core function is considered to be informational, while among others we can mention realization-forming (formation of specific political realities), protest (expression of disagreement with the actions of the current government and calls for its replacement) and magic (formation of belief in the fundamental "otherness" of the next candidate, party, etc;) c) *emotional engagement and influence of texts* (we are talking about the impact on public consciousness and political processes: It is actualized through the mediating role of political Internet discourse as a specific environment for interaction between political actors and the target audience (voters, politicians, etc.).

At the same time, such an influence can be both constructive (promoting discussion and understanding of complex political issues) and destructive (increasing political polarization and a sense

of distrust in the political system as a whole through critical pathos, since one of the tasks of Internet texts is to assess the actions of the authorities, which produces the use of several linguistic clichés (clichés, terminology, etc.)). The above suggests that the analysis of political online discourse texts in the context of machine learning should be built on three possible levels:

1. *Stylistic*, which, along with the analyzed data (since the content aspect is pragmatic, it is a description of certain facts of ontological reality), also contains its axiological ranking. Thus, we are talking about assessment, modality, etc.: for example, reports on the course of the russian-Ukrainian war can be positioned as an attack by fascist extremists on the "friendly" russian military forces defending the representatives of the "russian world" and the national liberation struggle against the occupiers).

2. *Manipulative* (or *rhetorical*), in which a certain desire by the addressee's idea of it is imposed under the guise of information (the above-mentioned construction of ontological reality by the language poly system). In the context of this level, it is advisable to talk about the existence of both direct and indirect (disguised) false information: misinformation, disinformation, etc. For example, the diplomatic fake used in the video with the Commander-in-Chief of the Armed Forces of Ukraine Valeriy Zaluzhnyi illustrates the existence of such data quite vividly. In essence, they are something that has taken place in the process of description or representation, transformed in such a way as to relay Russian narratives. In this case, by narrative, we mean unsubstantiated stories that aim to reproduce a fundamentally different image of an event or an image of another event by the addressee [28].

The above, according to L. de Saussure, is a special type of actualization of the linguistic poly system, which can be determined by such grammatical characteristics as goal setting, intentions, etc. The latter, in turn, are linguistically formalized in the quantitative predominance of linguistic sense that serves for the correct realization of the addressee's goal [29, p. 119]. It is noteworthy that such levels (stylistic and manipulative) imply the existence of a certain reality autonomous from the linguistic poly system, while the role of the latter is reduced to a kind of formal design (seemingly objective or deliberately distorted), which is inherent in political Internet discourse.

3. *Semantic*, in which the linguistic poly system acts in a dual role: as a form (functional role) and a medium (actually, we are talking about discourse) of sense. The point is that the language poly system is a form of construction and interpretation of ontological reality and a special type of social (speech) behavior. Thus, at the semantic level, the connection between the signified and the signifier is lost: "reality" (the signified) turns out to be equivalent to "language" (the signifier). However, the semantics of a linguistic poly system is not a self-reflexive construct that represents and reproduces in various forms exclusively its constructions.

The above approach is a significant simplification of the semantics of political online discourse: the idea of a referential emptiness or the resulting unverifiability (or specific falsity) of this discourse arises only when we apply the same procedures of semantic evaluation as for an "ordinary" statement. Instead, the sense and significance (reference) of political online discourse requires a special approach, since it is a complex, socio-cultural, and multilevel phenomenon, the description and evaluation of which requires the updating of the tools of modal semantics

(semantics of possible worlds) and pragmatics (theory of speech acts and performatives).

In analyzing the texts of political online discourse, it is necessary to take into account that it is not as biased as the discourse of other media, which leads to a greater degree of democracy and representation of opinion. It is about the absence of monopoly over certain data, which, in turn, causes fluctuations (oscillating nature) of sense within it. Nevertheless, the addressee has the opportunity to express his or her opinion: the transition from one level of information to another is a rather lengthy process. Thus, in some cases, this process is potentially open, i.e., it does not imply completion, since the network is constantly updating data and adding new information to a particular resource.

Let us turn to the above-mentioned understanding of the concept of discourse in linguistic science, where its positioning as a *tertium comparisonis* (third party) prevails. Thus, we are talking about its interpretation by a certain area at the intersection of language and speech, which, in turn, produces the seeming status of a special linguistic object. In our opinion, the above opinion is wrong, because discourse is a certain perspective, an interpretation of the description of the linguistic polysystem, within which non-universal and contextually determined dependencies are localized. (By the way, the latter are extremely productive for the process of setting the weights of the layers of artificial neural networks of various types and kinds.)

Instead, the language poly system is an ordered set, not of the elements themselves, but of context-dependent nonobligatory models of their actualization (language pragmatics). Such an understanding is representative and productive in the context of the purpose of our study, as it avoids a formalized or strict (i.e., based on the above-mentioned bonded characteristics) definition of political online discourse. Instead, it allows us to actualize T. van Dijk's non-exhaustive definition, representing such discourse as a description of the specific properties of any actualization of the language poly system used in certain behavioral patterns (in our case, for political purposes).

The basis for analyzing the texts of political online discourse in the context of machine learning is the correlation and indivisibility of words and social behavior. Here, we are talking about explicating the genetic basis of political action in the most general sense (as an ontological movement extrapolated epistemologically and ranked axiologically), namely, myth, rite, and ritual as the core elements of the latter. At the same time, importantly, special rules of linguistic behavior are actualized at the semantic level, disguised as common (ostensibly informing facts), the purpose of which is manipulation.

Thus, the above cannot be positioned as a communicative act or a description of certain facts of ontological reality. This is because the purpose of such rules is not to increase or cumulate new and known data, but to achieve a certain interpretation of the existing information with the subsequent acquisition of political success. Nevertheless, the Internet provides ample opportunities for creating a virtual dialog, which results in a modification of the addressee factor: while in political discourse the addressee is usually active, while the addressee is passive, in political Internet discourse both participants in communication are active.

In turn, this means that the addressee has the opportunity to interact directly with the addressee: to express his or her attitude to a particular issue through a message (in messengers, an e-mail, etc.), which leads to the diffusion of genres, their constant

modification and updating. Thus, political online discourse is an important component of the modern information space, where political thoughts, ideas, reactions, etc., as well as communication as such, play a pivotal role. As for the latter, in this discourse, unlike a conventional reference statement, the criterion is not the truth/falsity of the information actualized within it, but its impact on the addressee and the degree of success ranked by him/her. Therefore, political statements, regardless of the variety of grammatical forms, should be linguistically positioned as imperatives (obligations). The specific feature of which is not correlated with ontological reality, but success, expediency, and efficiency, which is naturally associated with the breadth of influence on public opinion [11].

Therefore, political Internet discourse is an original construct that differs from political and Internet discourse itself, and has a number of specific features, including: *political orientation* (thematic peculiarity, within which the issues of politics, public administration, good governance, etc. are discussed, as well as relevant socio-political topics and ideas), *actualization of communicators* (a special kind of interaction, in which communicators (including ordinary citizens-addresses) get the opportunity to participate in political communication: express their opinions and proposals at debates, round tables, etc.), *specificity of linguistic means* (localized (in relation to political discourse) vocabulary, jargon, etc. inherent in political language), information saturation (a special kind of information content, which does not always consist of a particular message about the events of ontological reality, but rather a specific representation of the latter in the context of a certain impact on the addressee), *a variety of actualized forms and types* (within the above discourse, a number of ways of implementation are actualized: from text messages to video blogs (for example, "circles" in Telegram) or live broadcasts.

The above naturally suggests that the main aspects that ensure the effectiveness of linguistic analysis of political Internet discourse texts for machine learning purposes are:

1. *Textual analysis*, which identifies relevant concepts, localizes concepts, topics, moods, etc. In the context of machine learning, this becomes the basis for an artificial neural network to recognize distinctive patterns and non-obvious correlations in such texts. In particular, in the context of linguistic science, political online discourse is productive for research:

a) *lexicology* (study of the original vocabulary actualized within it allows one to understand the peculiarities of the existence of certain ideas and specific concepts in the context of pragmatics (message, speech, visualization, etc.));

b) *syntax* (analysis of the used syntactic constructions and structures to establish the way political messages are formed and arguments for certain ideas are presented);

c) *stylistics* (tracking the variety of styles and genres in political online discourse illustrates the full range of expressive means and communication tools);

d) *semantics* (the identification of semantic cells and structures of political online discourse produces an understanding of the peculiarities of perception of a particular message, as well as the dynamics of its axiological ranking and determination of its impact on citizens; it also refers to the differentiation of shades of sense (distinguishing semantic differences in words and expressions by an artificial neural network that help to comprehend the context, i.e. the ability to work with semantic complexity)), etc.

2. *Analysis of the level of emotionality* (or *sentiment analysis*), which is used to study possible transformations of emotions/sentiments present in the texts of political online discourse. In turn, this is the basis for determining the positivity/negativity/neutrality of a particular statement or text and is productive for localizing hate speech by producing an understanding of the emotional context of a particular communicative act/acts.

3. *The analysis of the correlations between textual and ontological reality* is directly related to aspects 1-2 since data on references (meanings and senses) and sentiments (emotionality) correlate with its truth. Thus, the aforementioned aspect is important for recognizing and localizing linguistic manipulations and structural, semantic, indirect, etc. distortions to construct a certain interpretation of the events of ontological reality.

4. *Cultural and sociological analysis* (or *discourse analysis*) involves considering the number of realities, gaps, etc. determined by linguistic and cultural transformations. It is important to adapt an artificial neural network to work with the aforementioned features of political online discourse texts written by different addressees (representatives of social groups, communities, etc.). In addition, this aspect is representative of the socio-cultural context within which these components become culturally conditioned (productive in terms of taking the latter into account for the correct interpretation of the intentions of political statements/texts).

5. *Retrospective analysis* involves updating and grading references to the historical context: direct/indirect, implicit/explicit, etc. It is relevant for the artificial neural network to highlight the influence and rank its degree about the existence of certain trends in contemporary political (and Internet) discourse, studying their "novelty." We are talking about distinguishing between new and long-standing discussions, new discussions in themselves, new data in old discussions, and so on.

6. As a separate aspect, we consider it appropriate to highlight *speech analytics*, which partially actualizes and overlaps with the above, but the originality of its functioning requires its separate consideration. Thus, within the framework of the above-mentioned analytics, the peculiarities of interaction between political actors (addressees) and information consumers (addressees), as well as the language strategies used in this process, are studied [14]:

a) *argumentation*, which involves highlighting the arguments and logical strategies actualized in the process of political communication (actually, convincing the addressees-citizens in a certain interpretation of the event) (productive means of implementation are logical inferences and partial substitution of the interpretation of the facts presented, etc., emotionality and citing statistics with inaccurate indication of their sources);

b) *frames*, which study the specifics of the correlation between certain words and expressions of texts in the context of their general perception, in particular, the manipulative practices of establishing a specific linguistic frame to control the dynamics of discussion of a particular topic and, ultimately, to achieve a certain understanding of events;

c) *rhetoric* is the analysis of the frequency of use of certain rhetorical figures (metaphors, antitheses, etc.) to enhance the effect of the message and its memorability;

d) *several linguistic techniques/structures* involve distinguishing the structure of texts and their style: the use of active/passive constructions, the difference between written/oral messages, and the actualization of different language levels;

g) *communication strategies* that highlight the specifics of the algorithm of political actors' response to criticism and attacks on social media and other online platforms: retreat, ignoring, accusations, etc.

Thus, the above shows the integration of political Internet discourse as a phenomenon of the modern information space: it is about the originality of this type of discourse, which is characterized by democracy, lack of monopoly, etc. First of all, the above-mentioned aspects demonstrate the importance of linguistic studies of political (and Internet) discourse for understanding the specifics of its functioning as a reference environment, as well as the gradation of its influence on the public consciousness. Thus, from the linguistic point of view, this type of discourse is an important object of study, since its study reveals the complexity of the implementation of speech strategies, the influence of the language poly system on political activity and vice versa, as well as its constructive nature about ontological reality.

Conclusions from the study and prospects for further research in this area. Thus, analyzing the texts of political online discourse in the context of machine learning requires a comprehensive, fundamental, and integrated study of the available source base not only of linguistics but also of several related sciences (data science, machine learning, statistics, etc.). First of all, such an approach to solving the above-mentioned task will help to actualize several aspects of linguistic expression, including the socio-cultural context, etc.

At the same time, the above analysis is pivotal and representative of the existence of such discourse in the context of machine learning. In turn, this is due to the possibility of tracking the originality of the representation of the linguistic expression of certain political ideas, parameterization of emotional support, etc., which will significantly improve the efficiency and accuracy of such training. In addition, it will facilitate the emergence of a new generation of tools – artificial neural networks that will be able to analyze political Internet discourse in general and the parameterization of its texts (in particular, contributing to the understanding of the system of correlations, peculiarities of life, genesis, etc. of political processes).

Naturally, the above-mentioned analysis of political Internet discourse is a complex but significant task: first of all, it is about combating Russian disinformation, misinformation, etc. The development of algorithms capable of adapting to a fluid political environment and fully analyzing communication in it in all its forms and types can be a solution to this relevant task. Thus, the development of the latter will lead to a qualitative and effective analysis of political Internet discourse and will contribute to the development and improvement of the quality of democratic dialogue in the digital space of modern Ukraine and the world.

In addition, the analysis of political online discourse is representative of the dominant speech and communication strategies of political actors, the peculiarities of influencing political decisions and shaping public consciousness. At the same time, the aforementioned type of discourse is a dynamic phenomenon that is constantly evolving with the use of new technologies (in particular, artificial neural networks), which is why our future research will focus on the impact of artificial intelligence, as well as the development and improvement of methodologies for detecting false data in political online discourse.

Bibliography:

1. Rajandran K., Lee C. Politics in Malaysia : A discourse perspective. *Discursive approaches to politics in Malaysia : Legitimising governance.* Singapore : Springer, 2023. P. 1–15. *Open Acess* : website. URL: https://goo.su/dxqjyV (date of application: 20.11.23).
2. Discourse. *Cambridge Dictionary* : website. URL: https://goo.su/RGvpm (date of application: 20.11.23).
3. Дискурс. *Горох* : онлайн-бібліотека. URL: https://goo.su/OxlSp2 (дата звернення: 20.11.23).
4. Wittgenstein L. Philosophical investigations. London : Basil Blackwell, 1958. 250 p. *Squarespace* : website. URL: https://goo.su/9swv8R (date of application: 20.11.23).
5. Longo G. M. The Internet as a social institution : Rethinking concepts for family scholarship. *Family Relations.* 2023. Volume 72, Issue 2. P. 621–636. *Wiley* : online library. URL: https://goo.su/5G7WP (date of application: 20.11.23).
6. Offline events and online hate / Y. Lupu et al. *PLoS one.* 2023. Volume 18, Issue 1. https://doi.org/10.1371/journal.pone.0278511 *PLoS one* : website. URL: https://goo.su/hhnR5Y (date of application: 20.11.23).
7. Хроніки окупованих територій : станом на 14.11.2023. *Центр протидії дезінформації РНБО* : Телеграм-канал. URL: https://goo.su/o1hov (дата звернення: 20.11.23).
8. Болкарьова О. В. Аналіз політичного дискурсу в сучасному українському та англомовному просторі мас-медіа. *Вчені Записки ТНУ імені В. І. Вернадського. Серія: Філологія. Соціальні комунікації.* 2020. Т. 31 (70). № 1, Ч. 2. С. 20–23. *Вчені Записки Таврійського національного університету імені В. І. Вернадського. Серія: Філологія. Журналістика* : вебсайт. URL: https://goo.su/HUwc (дата звернення: 20.11.23).
9. Zayniddinovna B. G. Distinctive features of pleonasms in political speech. *Intellectual Education Technological Solutions And Innovative Digital Tools.* 2023. Volume 2, Issue 14. P. 168–174. *International Scientific Online Conferences* : website. URL: https://goo.su/ecbjD (date of application: 20.11.23).
10. Robertson A., Maccarone M. AI narratives and unequal conditions. Analyzing the discourse of liminal expert voices in discursive communicative spaces. *Telecommunications Policy.* 2023. Volume 47, Issue 5. https://doi.org/10.1016/j.telpol.2022.102462 *ScienceDirect* : website. URL: https://goo.su/MGjNIs (date of application: 20.11.23).
11. Chaves-Montero A. Social Services and Twitter : Analysis of Socio-Political Discourse in Spain from 2015 to 2019. *Sustainability.* 2023. Volume 15, Issue 4. https://doi.org/10.3390/su15043177 *MDPI* : website. URL: https://goo.su/neFjrnT (date of application: 20.11.23).
12. Gender-related differences in online comment sections: findings from a large-scale content analysis of commenting behavior / C. Küchler et al. *Social Science Computer Review.* 2023. Volume 41, Issue 3. P. 728–747. https://doi.org/10.1177/0894439321105204 *Sage Journals* : website. URL: https://goo.su/LR1gdPk (date of application: 20.11.23).
13. Kelling C., Monroe B. L. Analysing community reaction to refugees through text analysis of social media data. *Journal of Ethnic and Migration Studies.* 2023. Volume 49, Issue 2. P. 492–534. https://doi.org/10.1080/1369183X.2022.2100551 *Taylor & Francis Online* : website. URL: https://goo.su/jxhpP (date of application: 20.11.23).
14. Kuchina D. D. Metaphor as Invective in the Genre of Internet Commentary : A Focus on German Political Discourse. *Professional Discourse & Communication.* 2023. Volume 5, Issue 3. P. 61–75. *Research Gate* : website. URL: https://goo.su/6B2YX (date of application: 20.11.23).
15. Ramya G. R. Identifying fake reviews in relation with property and political data using deep learning. *Procedia Computer Science.* 2023. Volume 218. P. 1742–1751. https://doi.org/10.1016/j.procs.2023.01.152 *ScienceDirect* : website. URL: https://goo.su/ipOM (date of application: 20.11.23).

16. Terrorist Attacks, Cultural Incidents, and the Vote for Radical Parties : Analyzing Text from Twitter / F. Giavazzi et al. *American Journal of Political Science*. 2023. https://doi.org/10.1111/ajps.12764 *American Journal of Political Science* : website. URL: https://goo.su/4rUUY (date of application: 20.11.23).

17. Down the Rabbit Hole : Detecting Online Extremism, Radicalisation, and Politicised Hate Speech / J. Govers et al. *ACM Computing Surveys*. 2023. Volume 55, Issue 14. P. 1–35. https://doi.org/10.1145/3583067 *ACM DL* : website. URL: https://goo.su/zEjm1 (date of application: 20.11.23).

18. Attentional multi-channel convolution with bidirectional LSTM cell toward hate speech prediction / M. Fazil et al. *IEEE Access*. 2023. № 11. P. 16801–16811. https://doi.org/10.1109/ACCESS.2023.3246388. *IEEE xplore* : website. URL: https://goo.su/hON8 (date of application: 20.11.23).

19. Bagchi K., Banker E., Ogunleye I. Disrupting the Narrative : Diving Deeper into Section 230 Political Discourse. *The Center for Growth and Opportunity*. 2023. *The Center for Growth and Opportunity* : website. URL: https://goo.su/IrNLpN (date of application: 20.11.23).

20. Hate Speech Patterns in Social Media : A Methodological Framework and Fat Stigma Investigation Incorporating Sentiment Analysis, Topic Modelling and Discourse Analysis / V. U. Wanniarachchi et al. *Australasian Journal of Information Systems*. 2023. Volume 27. https://doi.org/10.3127/ajis.v27i0.3929 *Australasian Journal of Information Systems* : website. URL: https://goo.su/N4oT93s (date of application: 20.11.23).

21. Kar P., Debbarma S. Multilingual hate speech detection sentimental analysis on social media platforms using optimal feature extraction and hybrid diagonal gated recurrent neural network. *The Journal of Supercomputing*. 2023. P. 1–32. https://doi.org/10.1007/s11227-023-05361-6 *Springer Link* : website. URL: https://goo.su/jzrOzT (date of application: 20.11.23).

22. Üveges I., Ring O. HunEmBERT : a fine-tuned BERT-model for classifying sentiment and emotion in political communication. *IEEE Access*. 2023. Volume 11. P. 60267-60278. https://doi.org/10.1109/ACCESS.2023.3285536. *IEEE Xplore* : website. URL: https://goo.su/hqRIVC (date of application: 20.11.23).

23. Uncovering Political Hate Speech During Indian Election Campaign : A New Low-Resource Dataset and Baselines / F. A. Jafri et al. https://doi.org/10.48550/arXiv.2306.14764 *Cornell University* : website. 2023. URL: https://goo.su/iUSUV (date of application: 20.11.23).

24. A comparative study of deep learning models for sentiment analysis of social media texts / V. D. Derbentsev et al. *CEUR Workshop Proceedings*. 2023. P. 168–188. *CEUR Workshop Proceedings* : website. URL: https://goo.su/3ZNtYtu (date of application: 20.11.23).

25. del Valle E., de la Fuente L. Sentiment analysis methods for politics and hate speech contents in Spanish language : a systematic review. *IEEE Latin America Transactions*. 2023. Volume 21, Issue 3. P. 408–418. https://doi.org/10.1109/TLA.2023.10068844 *IEEE Xplore* : website. URL: https://goo.su/SN3d (date of application: 20.11.23).

26. Fournier-Tombs E., Di Marzo Serugendo G. DelibAnalysis : Understanding the quality of online political discourse with machine learning. *Journal of Information Science*. 2020. Volume 46, Issue 6. P. 810–822. https://doi.org/10.1177/0165551519871828 *Sage Journals* : website. URL: https://goo.su/FC6qzg (date of application: 20.11.23).

27. Törnberg P. How to use LLMs for Text Analysis. https://doi.org/10.48550/arXiv.2307.13106 *Cornell University* : website. URL: https://goo.su/7zisF5k (date of application: 20.11.23).

28. Narrative. *Cambridge Dictionary* : website. URL: https://goo.su/9MzHm3 (date of application: 20.11.23).

29. Saussure L. de. Manipulation and cognitive pragmatics : Preliminary hypotheses. Manipulation and ideologies in the twentieth century : Discourse, language, mind / L. de Saussure, P. Schulz (eds.). Amsterdam ; Philadelphia : J. Benjamins Pub. Co., 2005. P. 113–146. *Research Gate* : website. URL: https://goo.su/39lfIW (date of application: 20.11.23).

Довгань О. Політичний інтернет-дискурс: особливості аналізу тексту в контексті машинного навчання

Анотація. Стаття містить аналіз поняття дискурсу у контексті лінгвістичної науки, репрезентує структуру та особливості його тлумачення, а також можливі різночитання, які впливають на функційність. Окрім того, у дослідженні висвітлено самобутність політичної діяльності (зокрема, політичного (та інтернет-) дискурсу в контексті завдань машинного навчання та перспектив можливого вивчення. Зокрема, зосереджено увагу на онтологічному й аксіологічному ранжуванні політики як феномена людського буття та виокремлено ознаки політичної лінгвістики з притаманними для неї антропоцентризмом, експансіонізмом, функційністю. Також у статті репрезентовано особливості аналізу текстів політичного інтернет-дискурсу у контексті завдань машинного навчання: зокрема, висвітлено специфіку репрезентації тих чи тих політичних подій, параметризацію емоційного супроводу тощо.

Таким чином, дослідження особливостей аналізу політичного інтернет-дискурсу в контексті машинного навчання є актуальним в умовах сучасної міжнародної ситуації: зокрема, вивчення самобутності процесу аналізу текстів останнього у розрізі специфіки роботи штучної нейронної мережі. Окрім того, таке дослідження набуває особливої значущості ще й через гібридний характер російсько-українського протистояння, для якого притаманне побутування низки фейкових, викривлених, власне неправдивих та інших даних, актуалізованих у медіапросторі.

Автором підкреслено нерозривність аналізу текстів вищезазначеного дискурсу й боротьби з побутуванням таких неправдивих даних (дезінформації, мізінформації, пропаганди), сконструйованих російською федерацією з метою дестабілізації ситуації в Україні. У статті запропоновано вирішення означеного питання через розвиток алгоритмів машинного навчання, здатних до адаптації до плинного політичного інтернет-дискурсу, повноцінного аналізу комунікації у ньому у всіх видах та формах. Таким чином, вивчення текстів політичного інтернет-дискурсу в контексті машинного навчання безпосередньо пов'язано зі змогою створення штучних нейромережевих моделей, здатних до його ґрунтовного аналізу.

Ключові слова: дискурс, політичний інтернет-дискурс, аналіз інтернет-дискурсу, текстовий аналіз, машинне навчання, штучні нейронні мережі.