**DOI: https://doi.org/10.31392/NPU-nc.series9.2024.27.02**
**UDC: 811.161'32]:004.032.26:004.055:004.423.46**

Oleksii V. Dovhan
PhD in Philology,
Doctoral Student at the Department of Slavic, Romance,
and Oriental Languages,
Faculty of Foreign Philology,
Mykhailo Dragomanov State University of Ukraine,
Kyiv, Ukraine
https://orcid.org/0000-0002-6728-818X,
e-mail: oleksiivdovhan@gmail.com

# SENSE DIFFERENTIATION OF TEXTS AS A COMPONENT OF NEURAL NETWORK MODELLING

Bibliographic Description:

Dovhan, O. (2024). Sense Differentiation of Texts as a Component of Neural Network Modelling. *Scientific Journal of Mykhailo Dragomanov State University of Ukraine. Series 9. Current Trends in Language Development*, 27, 15–27. https://doi.org/10.31392/NPU-nc.series9.2024.27.02

***Abstract***

*The article argues that the most productive for linguistic research at the present stage is the use of Artificial Neural Networks (ANNs) due to their productivity, representativeness, etc. It is emphasized that the basis for such use should be sense differentiation, thanks to which linguists can optimize the search, analysis, etc. of data for their research. In particular, taking into account semantic, morphological, syntactic, etc. features will allow the production of more reliable, fundamental results in various tasks of Natural Language Processing (NLP). The author emphasizes that the above will result in a qualitative leap in the scientific research of Ukrainian linguists, the possibility of presenting their results to world science, and further fruitful cooperation with foreign colleagues within the framework of grant programs.*

*Thus, the semantic differentiation of texts is an integral part of the actualization of Artificial Neural Networks (ANNs) (in particular, Bidirectional Long Short-Term Memory Network (BiLSTM), Convolutional Neural Networks (CNNs), Deep Learning Networks (DLNs), Deep Neural Networks (DNNs), Graph Neural Networks (GNNs), Recurrent Neural Networks (RNNs), etc.) in modern linguistic research within the digital humanities. In the author's opinion, the latter is produced by the focus on practical results, localization of implementation (in particular, Natural Language Processing (NLP), sentiment analysis, etc. Therefore, further study, improvement, and optimization of the existing innovative tools (in particular, neural network modelling of linguistic units) include work on more effective methods of working with context (through a combination of different types of Artificial Neural Networks (ANNs) with different layers, which is presented in the work of foreign colleagues), localization of language styles (essential in the process of fact-checking initiatives – as a milestone for validating text data), etc. without manual intervention in the above.*

***Keywords:*** *neural network modelling, applied linguistics, computational linguistics, artificial neural networks, neural network models, sense, sense differentiation.*

## *1. Introduction.*

Naturally, the genesis of scientific research correlates with socio-cultural transformational changes in society, which, in turn, produce a certain cyclicality and specificity of the latter's dynamics. Thus, digitalization forms a special culture of

consumption, worldview, and so on, affecting, in fact, the understanding of the essence of scientific research (in particular, humanitarian research). The originality of the above approach is in the positioning of any research as an integrated construct that functions in correlation with others that are close in topic, objects, methodology, hypothesis, etc., but are in a different subject field.

In the context of this approach, an innovative research toolkit that is inherent in understanding text (words, phrases, sentences, etc.), primarily as a data source, is indicative. Here it is worth mentioning Natural Language Processing (NLP) and Artificial Neural Networks (ANNs) as its tools (in particular, neural network modelling of linguistic units). The use of the above-mentioned tools creates several problems in conducting linguistic research, the core of which is the issue of effective processing of textual data, which is characterized by levelness, complexity, ambiguity, etc. We see the solution to this issue in the concept of sense differentiation of texts as an important component of neural network modelling of linguistic units in the context of Natural Language Processing (NLP).

***2. Literature Review.***

The foregoing demonstrates the originality of the problem of sense differentiation of texts as a component of neural network modelling: it refers to the algorithm of this process, the sources and layers updated in it, etc., as well as the ranking of the data used, the architectural features of neural network models, etc. In turn, this breadth of the problem field produces the integration of works related to the discursive nature of neural network modelling in the context of building modern humanitarian (in particular, linguistic) research. We also emphasize the relevance of such components of neural network modelling to linguistic analysis as taking into account the specifics of the machine and deep learning (number, types, succession of updated layers, etc.), data science tools (extraction, processing (analysis, processing, representation, etc.) and data updating, etc.), etc.

The study of H. Wang (Wang, 2024 URL: https://goo.su/seplAN) is devoted to the analysis of Natural Language Processing (NLP) in the context of solving several practical problems of modern linguistics and developing certain solutions. In the analysed study, the author highlights the specifics of Natural Language Processing (NLP), noting that such work usually consists of solving *the core task* (in fact, it is a problem solved with the help of various areas of Natural Language Processing (NLP), it includes language models, morphology, grammatical, semantic, etc. analyses, etc.) and *the application* (focused on certain specific tasks performed to solve the task: we are talking about such processes as machine translation, information retrieval, question-answering systems, dialogue systems, etc.).

The scientist notes that Natural Language Processing (NLP) is a productive environment for modern linguistic research, in which opinion mining, or sentiment analysis, is commonly distinguished. The researcher emphasizes that sentiment analysis is productive for processing textual data (in particular, texts of Internet discourse, including political discourse). The author sees the representativeness of working with the latter in their nature: it is unstructured data that can be organized for use in linguistic and pragmatic, psycholinguistic, etc. studies using Convolutional Neural Network (CNN). The scientist believes that convolutional operations, which are productive for processing text data and, as a result, for sentiment analysis, are a tool for the above.

A further study of the latter in the context of Deep Neural Networks (DNNs) as a determinant of the aforementioned Natural Language Processing (NLP) is presented in D. Tsirmpas et al. (Tsirmpas, Gkionis, Papadopoulos, & Mademlis, 2024 URL: https://goo.su/sKUQLc1). In the analysed study, the authors highlight the specifics of

analysing long and short texts in terms of processing requirements for this task. The scientists note that the exponential growth of data in Internet discourse makes its automated understanding extremely relevant. First of all, the researchers are referring to automated web search (we can partially see its application in the process of working with Google), analysis of legal, medical, financial, etc. data, news aggregation, etc.

The authors provide an introductory overview of the neural network architectures required to address the above issues and briefly characterize the core tasks of long text analysis (classification, summarization, etc.). The authors pay special attention to sentiment analysis in long texts, which is due to the common practice of considering the latter in the context of a special case of document classification. This study also provides a thorough taxonomy of common deep neural methods for analysing long documents and lists publicly available annotated datasets that are productive for linguistic research in this area.

The features of sentiment analysis and information communication, the specifics of their classification and parameterization of expression, etc. in English-language text data are highlighted in J. Hao, Y. Ding (Hao, & Ding, 2023 URL: https://goo.su/yA1jz). In the analysed study, the above is based on the correlation between emotions and human values, which allowed the authors to build a neural network model Bidirectional Long Short-Term Memory Network (BiLSTM) that studies sentiment analysis and information communication. The scientists claim that to optimize the latter's work with English-language text data, they encoded the emotional information of such text. This, in turn, made it possible to further extract emotional features from such data and solve the problem of losing the latter using the loss function.

The specifics of the existing methods of image-text matching (in particular, in the context of the relevance of image-text pairs with the fixation and aggregation of the affinity between words and images) are presented in X. Qin, L. Li, G. Pang (Qin, Li, & Pang, 2024 URL: https://goo.su/tidM). In the analysed study, the authors present the developed Multi-Scale Motivated Neural Network (MSMNN) designed to match images and text. According to the scientists, the peculiarity of this artificial neural network is that it extracts visual and textual features from three scales: local features, global features, and salient features. Researchers believe that this approach produces an actualization of multi-scale complementarity to reduce the error of single-scale matching, which, in turn, allows the use of a cross-modal interaction module. The latter combines the results of the above-mentioned process at three such levels at once and can be productive if applied to other initial results within the scope of the problem studied by the authors.

Another aspect of text data research in the context of innovative tools is presented in the study by X. Ren (Ren, 2024 URL: https://goo.su/OFRTj0), which highlights the specifics of automated distinction between handwritten and printed text. The latter is a problem in the context of automated processing since handwritten text is quite variable. The author sees the solution to this problem in the use of the Deep Spatio-Temporal Residual Convolutional Neural Network (DS-TRCNN) on the Maximum Qualitative Analysis 6 (MAXQDA6) dataset, encoded using the deductive method. The scientist notes that the Spatial-Temporal Filter (STF) he uses is productive for detecting data that changes over time and is tied to a specific location. The researcher also emphasizes that he has updated the Kernel Principal Component Analysis (KPCA), which helped to separate attributes from segmented data. The author notes that the proposed method is productive for automated storytelling, support for creative writing, and investigation of literary genres, which is important for modern linguistic research.

Text search as a milestone in the data selection process and its features are highlighted in the study by H. Tu (Tu, 2024 URL: https://goo.su/YrVu5), where the author

proposes to use the Convolution Neural Network (CNN) developed by him for this purpose. The scientist notes that the aforementioned neural network model is productive in the process of extracting the original features of the text and reducing its dimensionality. According to the researcher, the latter allows for real-time search by calculating the semantic similarity of text data. The author emphasizes that the developed Convolution Neural Network (CNN) is characterized by high accuracy, complexity of text data analysis, and low time consumption.

A logical continuation of the analysis of textual data is the study of the specifics of the communication actualized in them (sarcasm), conducted by C. Thaokar et al. (Thaokar, Rout, Rout, & Ray, 2024 URL: https://goo.su/SdFuvC), which investigates the use of words with the opposite meaning to the direct one for absurdization, comedy, etc. According to the authors, this communicative practice is a challenging problem for Natural Language Processing (NLP) (in particular, sentiment analysis and information communication), as it can actualize positive words to represent negativity. Scientists believe that this makes it difficult to detect sarcasm, as it changes the polarity of the sentence and the difference between the actualized words and the way they are used. The researchers note that they studied various word-level features to localize sarcasm in three benchmark datasets (the latter included the creation of an N-gram probability dictionary, negation words, and PoS tags). In addition, the authors reviewed and compared different machine learning and hybrid deep learning models with handcrafted features and word embedding features, which allows us to speak about the representativeness of the study.

The specifics of detecting false data in Internet discourse (in particular, social media) are discussed in R. Chelehchaleh et al. (Chelehchaleh, Salehi, Farahbakhsh, & Crespi, 2024 URL: https://goo.su/kb5Li). In the analysed study, the authors argue that previous approaches to detecting false data (misinformation, disinformation, propaganda, and, in particular, fake news) focused on the text, but, in their opinion, a comprehensive approach to solving this problem is needed. Thus, the scientists see a way out in the use of various machine learning techniques, in particular, the actualization of the Hybrid and Multi-Feature Framework (BRaG). The latter takes into account the content (e.g., text) and context (e.g., domain, style, sentiment analysis, etc.) of the analysed data and is based on a combination of Bidirectional Encoder Representations from Transformers (BERT) pre-trained language model, Recurrent Neural Network (RNN) and Graph Neural Network (GNN), which allows it to analyse the text, the sequence of users involved and the dynamics of the distribution of such data (respectively, the final data representation vector). The researchers emphasize that the proposed approach also analyses text emojis (in particular, it takes into account the contextual information they highlight), which makes the results more valid.

The aforementioned problem is still being studied by A. Shalini, S. Saxena, B. Kumar (Shalini, Saxena, & Kumar, 2024 URL: https://goo.su/A1j0dAE), who propose a system based on the deep learning model of Recurrent Neural Network-Long Short-Term Memory (RNN-LSTM). The authors note that the proposed approach is based on the conventional feature extraction process: Lemmas, Bi-Gram, Tri-Gram, N-gram, Term Frequency Inverse Document Frequency (TF-IDF), part-of-speech, and dependency-based natural language processing (NLP) features.

Instead, N. Ye et al. (Ye, Yu, Ma, Zhou, & Yan, 2024 URL: https://goo.su/EvuS) argue in their study that existing methods for detecting false information lose some of the data to varying degrees, and the authors see the solution to this problem in their proposed content-based recognition method with low computational costs. Thus, the researchers present a lightweight system for localizing English-language false data (news) in Internet discourse, which includes a new method of extracting textual features (specifically mapping English text and symbols to 0-255 using American Standard Code for Information

Interchange (ASCII) codes). According to the researchers, the essence of the latter is that the completed sequence of numbers is interpreted as the value of picture pixel points, and a computer vision model is used to detect them. The above allowed the authors to position their framework similar to Word2Vec, Glove, Bidirectional Encoder Representations from Transformers (BERT), etc.

A logical continuation of the analysis of false data in Internet discourse is the study by P. Krishnamoorthy, M. Sathiyanarayanan, H. Proença (Krishnamoorthy, Sathiyanarayanan, & Proença, 2024 URL: https://goo.su/AO1N), in which the authors explore the possibilities of categorizing spam and ham email documents. The researchers used different approaches to categorizing this type of data, based on data mining. The scientists tested the method on the publicly available Enron dataset: the latter was pre-processed to optimize the result and speed up its pre-processing (vectorization, clustering, etc.). The method proposed by the authors has been updated in several categories, including Logistic Regression (LR), Convolutional Neural Networks (CNNs), Random Forests (RF), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and the proposed Deep Neural Networks (DNNs).

The analysis of textual data from Internet discourse in the context of the Recommender System (RS) is presented in R. Patel, P. Thakkar, V. Ukani (Patel, Thakkar, & Ukani, 2024 URL: https://goo.su/TUJHg4), where the latter is positioned as an innovative tool for decision-making. The authors note that the three common forms of such systems that actualize explicit/implicit feedback to build their recommendations are: collaborative, content-based, and hybrid filtering. According to the researchers, ratings are the most productive form of such communication, but product descriptions, reviews, images, audio, and video materials are also important, as they play a corrective role, in improving the results of the above systems and their performance. The scientists emphasize that traditional Recommender Systems (RS) are based on the nearest neighbour method or other machine and deep learning models, but today, there are already systems based on the aforementioned Convolutional Neural Network (CNN).

The problem of textual data processing (in particular, the actualization of abbreviations in machine and deep learning and the success of a particular neural network model) in the context of working with senses is covered in D. Chopard, P. Corcoran, I. Spasić (Chopard, Corcoran, & Spasić, 2024 URL: https://goo.su/PGhi). In the analysed study, the authors note that the use of abbreviations without their clear decoding is common, as well as the existence of several industry-specific complications (patient confidentiality, manual commenting, etc.) in medical practice, which makes it difficult for machine and deep learning. The researchers see a way out of the above situation in the actualization of scientific abstracting as a means of overcoming these problems by forming a large automatically annotated dataset. Noteworthy that the authors propose to provide the latter with artificially modelled global abbreviations: this will allow integration of the neural network model with the recognition of multi-word terms. According to scientists, this approach optimizes the process of text data processing, significantly speeding up the extraction of senses of certain abbreviations from the data corpus for analysis.

Another aspect of the problem is studied by X. Huang, Y. Huang, C. Mercado (Huang, Huang, & Mercado, 2024 URL: https://goo.su/17w1b), who analyse in their study the methodology of automatic generation of professional text (in particular, neural network algorithms). The authors note the complexity, duration, limitations, etc. of the above approach (updating a large amount of annotated data, complex statistical algorithms, etc.), emphasizing the lack of convergence and local vulnerability. The scientists claim that they have developed (based on optical detection and an advanced neural network algorithm) the

most effective method of automatic text generation and data processing for a specific topic (in this case, recruiting texts).

The study of a human-like computational model for concept learning based on Spiking Neural Networks (SNNs) is presented in Y. Wang, Y. Zeng (Wang, & Zeng, 2024 URL: https://goo.su/seplAN), in which the authors analyse the specifics of concept acquisition in individuals. The authors identify the mechanisms actualized in human concept learning and present the findings from computational neuroscience and cognitive psychology. The scientists emphasize that the representation of concepts in the brain is based on two main milestones: *multisensory* and *text-derived representations*, which, in turn, are coordinated by a semantic control system, which leads to their (concepts') assimilation. The researchers note that the actualization of human mechanisms for working with concepts in the neural network model they created produced results similar to human cognition.

Thus, the analysis of the historiography of this issue has shown the originality of sense differentiation of texts as a component of neural network modelling (in particular, of linguistic units). In addition, the above analysis has revealed the lacunarity of this issue: for example, several contemporary studies that examine various aspects of this issue lack a comprehensive analysis of the sense differentiation of texts (in particular, Internet discourse). There are also many unresolved issues related to Natural Language Processing (NLP), sentiment analysis, information communication, etc., which naturally creates uncertainty in methodology, tools, and other modern linguistic research. The above naturally produces the relevance of this article, which aims to fill the existing gaps by highlighting the specifics of sense differentiation of texts as a component of neural network modelling.

### *3. Aim and Objectives.*

*The article aims* to consider the originality of sense differentiation of texts in the process of modern scientific linguistic research. *The subject* is the specificity of the above in the context of neural network modelling as an innovative tool of linguistic science.

Achievement of the above aim and subject involves the realization of the next *objectives:*

1. Systematize theoretical achievements in the semantic differentiation of texts in the context of neural network modelling of linguistic units.
2. Analyse the existing algorithms and methods of computing, computational, applied, etc. linguistics, machine and deep learning, data science, etc. of the above process.
3. Present neural network modelling of linguistic units as an innovative tool of modern linguistic science.

### *4. Methodology.*

A language poly system is a complex object of study not only for modern innovative linguistic research but also for conventional, classical research. First of all, this is due to the ambiguity of its existence, as well as the dynamic correlation with transformational changes produced by digitalization processes. In particular, Artificial Neural Networks (ANNs), which are becoming increasingly popular, are a productive innovative tool for modern humanities (including linguistic) research. The representative use of the latter is due to the possibility of modelling complex language structures, understanding the context, etc., which generally optimizes such research by speeding up the processing of textual data and the like. We consider it expedient to focus on the key concepts, methods, and examples of neural network modelling in such studies in the context of sense differentiation of texts.

First of all, it is advisable to focus on the key issue of this article – sense differentiation of texts, which is the process of highlighting, localizing, etc. the above-

mentioned aspects of linguistic units in textual data. Naturally, this process requires several discursive, background knowledge: contextualization, semantics, syntax, morphology, etc. (Huang, Huang, & Mercado, 2024 URL: https://goo.su/17w1b). The aforementioned milestones will allow Artificial Neural Networks (ANNs) to understand the originality of a particular meaning, usage features, and combinability of linguistic units. Mostly neural network methods are designed to automate the process of linguistic analysis, and the basis for their work is the above-mentioned Artificial Neural Networks (ANNs) (in particular, Bidirectional Long Short-Term Memory Network (BiLSTM), Convolutional Neural Networks (CNNs), Deep Learning Networks (DLNs), Deep Neural Networks (DNNs), Graph Neural Networks (GNNs), Recurrent Neural Networks (RNNs), etc.), which are similar in structure to the neural systems of the human brain.

To perform neural network modelling (of words, phrases, sentences, text), it is necessary to process text data in a certain way. We are talking about their pre-processing: *vectorization* (improves the distinction between word meanings in the context of the entire text data set), *clustering* (structuring, grouping similar texts into certain categories or clusters), and *the use of contextual models*, such as Bidirectional Encoder Representations from Transformers (BERT) (capable of localizing shades of sense, indirect meanings of words, etc.), *improving the quality of updated data* (we are talking about augmentation, data cleaning, etc., as well as the use of ready-made or manually collected by a linguist datasets: thus, a significant improvement in the "input" will lead to an exponential increase in the relevance of the "output" data), etc.

Representative examples of the methodology for updating the neural network modelling of words for sense differentiation in the context of Natural Language Processing (NLP) of texts are: a) *distinguishing the senses of words in context:* automated distinction of the meaning of words with its consideration (e.g., machine translation, etc.); b) *analysis of text semantics:* study of sentiment analysis, information communication, etc, studying the emotional colouring of the text as a fact-checking component, etc.); c) *text generation:* creation of text data taking into account the style, type, semantic and logical parameterization of the text, etc, annotated datasets as a component of neural network model training, etc.); d) *localization of synonymy*, *homonymy*, *etc.:* distinguishing complex linguistic phenomena (sarcasm, irony, etc. as elements of absurd sense) and distinguishing the meaning of linguistic units and the like (e.g., generating a video or image at the user's request, etc.) and others (Tu, 2024 URL: https://goo.su/YrVu5).

Thus, the methodology of sense differentiation of texts using neural network modelling is an important area of modern linguistic research (applied, computational, etc. linguistics, Natural Language Processing (NLP), and others). This is due to the effectiveness of using the above methodology to study sense correlations in textual data. The prospect of actualizing this approach is the genesis of methods for neural network modelling of linguistic units to optimize the results of neural network models: increasing accuracy, speed, etc.

### *5. Results.*

The above-mentioned Artificial Neural Networks (ANNs) (in particular, Bidirectional Long Short-Term Memory Network (BiLSTM), Convolutional Neural Networks (CNNs), Deep Learning Networks (DLNs), Deep Neural Networks (DNNs), Graph Neural Networks (GNNs), Recurrent Neural Networks (RNNs), etc.) are powerful innovative tools for modern humanities (in particular, linguistic) research in terms of processing, analysing, representing, etc. textual data. The core difficulty of conducting such a study in the digital humanities is textual variability: the originality of stylistics, sentiment analysis, semantics, morphology, etc. architecture, etc. (Wang, & Zeng, 2024 URL: https://goo.su/seplAN).

The concept of sense differentiation is designed to solve this problem, where the researcher performs certain processing of the analysed data to optimize the result of the neural network model. Thus, the aforementioned concept plays an important role in neural network modelling, as it produces the actualization of semantic, syntactic, morphology, and other differences in textual data and their relational analysis. The latter includes the ability to process similar but not identical texts, which is productive when checking academic texts for plagiarism (e.g., Plag's artificial intelligence content check) or processing texts from different cultures or dialects.

Meanwhile, the above-mentioned allows us to speak about the originality of sense differentiation of texts in the process of modern scientific linguistic research and its specificity about neural network modelling as an innovative tool. Thus, the theoretical and practical achievements in this process, mentioned in the literature review of this article, testify to the special place of sense differentiation. The point is that it is a core component of several Natural Language Processing (NLP) tasks: in particular, processing, analysis, representation of textual data, their generation concerning computing, computer science, applied, etc. linguistics, machine and deep learning, data science, etc. (Chelehchaleh, Salehi, Farahbakhsh, & Crespi, 2024 URL: https://goo.su/kb5Li). For example, in the above-mentioned plagiarism check, it is necessary to determine the degree of similarity of a text to similar or identical ones, and it is also productive to study its belonging to a certain category, etc.

As noted in the extensive historiography of the problem, ensuring the effective operation of neural network models of a particular type with architectural features (number, order, etc. of layers in them) is possible by actualizing various aspects of sense differentiation. The latter is done by optimizing the architecture of the neural network model, qualitative data pre-processing, and its advanced training on high-quality datasets. This approach ensures quick adaptation of the used Artificial Neural Networks (ANNs) and their updated layers to work with text data. For example, in the case of text generation (e.g., ChatGPT 3.5 and ChatGPT 4, etc.), the correct distinction between style, topic, and other textual data allows you to create more authentic and relevant text. As for machine translation (e.g., DeepL), the identification of semantically similar but distinct phrases can improve the quality of work, avoid misunderstanding of the text, and, as a result, significantly improve "the output" (Patel, Thakkar, & Ukani, 2024 URL: https://goo.su/TUJHg4).

Note that digitalization trends are affecting our lives right now (when searching Google, YouTube, etc., the recommendations of these platforms are based on the work of Artificial Neural Networks (ANNs)), which is why their use in scientific activities is a requirement of the times. At this stage, we consider the use of the aforementioned Artificial Neural Networks (ANNs) to be the most productive for linguistic research because of their performance, representativeness, etc. At the same time, it is advisable to make sense of semantic differentiation as the basis for this use, thanks to which linguists can optimize the search, analysis, etc. of data for their research. In particular, taking into account semantic, morphological, syntactic, etc. features will allow producing more reliable, fundamental results in various Natural Language Processing (NLP) tasks. The above will result in a qualitative leap in the research of Ukrainian linguists, the possibility of presenting their results to world science, and further fruitful cooperation with foreign colleagues within the framework of grant programs.

Thus, the sense differentiation of texts is an integral part of the actualization of Artificial Neural Networks (ANNs) (in particular, Bidirectional Long Short-Term Memory Network (BiLSTM), Convolutional Neural Networks (CNNs), Deep Learning Networks (DLNs), Deep Neural Networks (DNNs), Graph Neural Networks (GNNs),

Recurrent Neural Networks (RNNs), etc.) in modern linguistic research in the digital humanities. Focus on practical results, localization of implementation (in particular, Natural Language Processing (NLP), opinion mining, or sentiment analysis, etc.) (Shalini, Saxena, & Kumar, 2024 URL: https://goo.su/A1j0dAE). Further study, improvement, and optimization of the available innovative tools (in particular, neural network modelling of linguistic units) include work on more effective methods of working with context (through a combination of different types of Artificial Neural Networks (ANNs) with different layers, which is presented in the work of foreign colleagues), localization of language styles (essential in the process of fact-checking initiatives as a milestone for validating textual data), etc. without manual intervention in the above. In addition, it would be productive for Ukrainian linguistics to improve existing algorithms that would allow artificial neural network models to take into account individual differences in texts. We are talking about updating the gradients of sense weight for different elements of such texts and using contextual approaches to text data processing.

***6. Discussion.***

Natural Language Processing (NLP) is an integrated (of interest to linguists and other scientists, including those working with data) promising area of research. The latter's relevance is due to the exponential growth of textual (and other) data, which primarily functions in Internet discourse. Notably, the very concept of sense differentiation is an indication of the specifics of differences between textual data, and in the context of neural network modelling of linguistic units, it is the core concept. This, in turn, is related to the need to achieve greater accuracy, efficiency, speed, etc. of existing neural network models in the process of performing certain tasks of the above-mentioned Natural Language Processing (NLP) (Tsirmpas, Gkionis, Papadopoulos, & Mademlis, 2024 URL: https://goo.su/sKUQLc1; Wang, 2024 URL: https://goo.su/seplAN, etc.): machine and deep learning (Chopard, Corcoran, & Spasić, 2024 URL: https://goo.su/PGhi; Huang, Huang, & Mercado, 2024 URL: https://goo.su/17w1b, etc.), machine translation, sentiment analysis and information communication (Chelehchaleh, Salehi, Farahbakhsh, & Crespi, 2024 URL: https://goo.su/kb5Li; Hao, & Ding, 2023 URL: https://goo.su/yA1jz, etc.) and others.

Thus, the spectrum of possible actualizations of Artificial Neural Networks (ANNs) in modern linguistic research is quite representative in the literature review of our article. It should be noted that the very concept of spectrum is productive in this regard since the specifics of modern linguistic research conducted within the digital humanities are fundamentally open. This refers to the poly interpretability of linguists' understanding of modern innovative tools: e.g., the use of Artificial Neural Networks (ANNs) of any type (in particular, Bidirectional Long Short-Term Memory Networks (BiLSTM), Convolutional Neural Networks (CNNs), Deep Learning Networks (DLNs), Deep Neural Networks (DNNs), Graph Neural Networks (GNNs), Recurrent Neural Networks (RNNs), etc.) will naturally have several features.

Therefore, it is up to the linguist-researcher to decide which tool, in what order, and in what combination he or she can use, because a completely non-standard approach may turn out to be productive since algorithms are developing every day and tomorrow what is commonly used (e.g., Convolutional Neural Network (CNN), Graph Neural Network (GNN), Recurrent Neural Network (RNN), etc.) may be outdated and not relevant. What is important here is the plasticity, adaptability, openness, etc. of the methodology used for such research and the readiness to revise its core principles. The above gives rise to many controversial issues that cannot be resolved by the works mentioned in the literature review and this article, as they are too large-scale in nature, in particular:

1. *The issue of formalizing the concept of sense differentiation of texts concerning neural network modelling*. Notably modern humanitarian (in particular, linguistic) research is characterized by the poly interpretability of methods, approaches, and the variety of tools used. Thus, even if Artificial Neural Network (ANN) becomes the key tool, the problem of its type, architecture, originality of the algorithm, etc. arises. Thus, we are talking about an equation with a large number of possible variables, each of which significantly affects the dynamics, performance, etc. of the result. In this context, it should be borne in mind that the choice of all of the above may not necessarily be a good one, but even a negative, incomplete, etc. result of using an innovative linguistic tool is indicative and produces its further genesis.

2. *The issue of representativeness of certain approaches, tools, instruments, etc.* Directly related to the aforementioned variability, since today a linguist-researcher can refuse to use the traditional "gentleman's set" for neural network modelling of linguistic units (e.g., Convolutional Neural Network (CNN), Graph Neural Network (GNN), Recurrent Neural Network (RNN), etc.) and choose, for example, the "Sora" algorithms of the Artificial Neural Network. At the same time, he will be able to justify his choice by the performance of the latter with text data (the developers, OpenAI, have such successful products as ChatGPT 3.5 and ChatGPT 4, etc.) This approach may result in something other than typical Natural Language Processing (NLP): machine and deep learning, machine translation, sentiment analysis, information communication, etc. Nevertheless, the results of such a study will be just as valuable, evidentiary, representative, and reproducible.

3. *The issue of relevant factors that correlate with the sense differentiation of texts and the possibility of their automation*. This is an ambiguous question that cannot have a single correct answer: even the literature review of this article has revealed a wide range of possible solutions. It should be noted that in the process of writing our doctoral dissertation, we studied a much larger number of works by scholars on the analysed topic. The result of this fundamental analysis was the understanding of the openness of this issue, since in this case the tools of modern linguistic research are related to digitalization changes. In particular, we are talking about the dependence of research in the digital humanities on digitalization trends, the dynamics of information technology development, etc., which, in turn, illustrate the potential uncertainty of this issue.

4. *The issue of updating sense differentiation data to improve the performance of Artificial Neural Networks (ANNs) in the process of performing the above-mentioned Natural Language Processing (NLP) tasks (in particular, machine and deep learning, machine translation, sentiment analysis, information communication, etc.)* It should be noted that in this article, we highlighted the correlation between pre-processing textual data for analysis and, in fact, several possible actions within the concept of sense differentiation of texts. For example, we noted that this concept is pivotal because it allows localizing many differences between textual data (e.g., two texts containing the same information, but presented distinctively in terms of style, punctuation, morphology, etc.) In turn, automation of such laborious research work will help optimize scientific work and the genesis of modern science in general. In addition, the above will lead to the further development of new innovative approaches to conventional linguistic concepts (analysis, processing, representation, etc.): the latter will be understood not only from the standpoint of linguistics but also in an integrated way – as data for processing, which will significantly change the methodology, purpose, results, etc. of such research.

5. *The issue of existing challenges in the development of neural network models that would consider the sense differentiation of texts, and the problem of overcoming such difficulties*. If we talk about the world practice, this issue (working with Artificial Neural

Networks (ANNs) and their certain algorithmization) correlates with the views, place of work, subject orientation of the researcher, which is still available, etc. Thus, such research in the humanities is mostly conducted within the framework of digital humanities or information technology (from programming to machine and deep learning to data science as an integrator of the entire possible spectrum of actualization).

Thus, the sense differentiation of texts is a core element of the integration of the mathematical paradigm into humanities research. The development of the latter, in turn, correlates with the hardware, software, etc. of scientific, educational, and other institutions (Ukraine) and the dispersion of approaches to innovative linguistic tools (foreign countries). That is why it is important today to aggregate approaches to neural network modelling within the digital humanities, which will allow the development of more efficient, accurate, etc. neural network models capable of localizing the originality of sense differentiation of textual data. The discussion around the above and several related issues will produce further development of Natural Language Processing (NLP) methods (in particular, machine and deep learning, machine translation, sentiment analysis, information communication, etc.) and exponential growth of the use of Artificial Neural Networks (ANNs) in modern linguistic research (including in Ukraine).

### *7. Conclusions.*

Thus, the concept of sense differentiation of texts is the cornerstone of neural network modelling in the context of the aforementioned Natural Language Processing (NLP). To a greater extent, the relevance of such differentiation correlates with the originality of its implementation in the work of Artificial Neural Networks (ANNs). The latter, in turn, will produce an increase in the accuracy, quality, etc. of their work, which will result in the efficiency of processing (analysis, generation, etc.) of text data in several Natural Language Processing (NLP) tasks. The breadth of the spectrum of sense differentiation application correlates with the level, variability, and other specific tasks, the specifics of the actualized data, etc. analysed by the artificial neural network model. Naturally, the type, algorithm, specifics of training, the order and types of layers updated in its structure, etc. will affect the output. Socio-cultural determinants are relevant in such studies: for example, the meaning of the same phrase varies depending on the linguistic, national, and other worldviews of its speaker. The latter is a prospect for future work on neural network modelling in modern linguistic research in the context of the digital humanities.

### ***R e f e r e n c e s***

Chelehchaleh, R., Salehi, M., Farahbakhsh, R., & Crespi, N. (2024). BRaG: a hybrid multi-feature framework for fake news detection on social media. *Social Network Analysis and Mining*, 14 (1), 35. https://doi.org/10.1007/s13278-023-01185-7 *Springer Link*: website. URL: https://goo.su/kb5Li (date of application: 01.04.24).

Chopard, D., Corcoran, P., & Spasić, I. (2024). Word sense disambiguation of acronyms in clinical narratives. *Healthcare Text Analytics: Unlocking the Evidence from Free Text*, Volume IV. https://doi.org/10.3389/fdgth.2024.1282043 *Frontiers*: website. URL: https://goo.su/PGhi (date of application: 01.04.24).

Hao, J., & Ding, Y. (2023). Emotional Expression and Information Communication in English Texts Based on Artificial Intelligence Technology. *Applied Mathematics and Nonlinear Sciences*. https://doi.org/10.2478/amns.2023.2.01047 *Sciendo:* website. URL: https://goo.su/yA1jz (date of application: 01.04.24).

Huang, X., Huang, Y., & Mercado, C. (2024). Artificial intelligence recruitment text automatic generation based on light detection and improved neural network algorithm. *Optical and Quantum Electronics*, 56 (2), 162. https://doi.org/10.1007/s11082-023-05770-0 *Springer Link:* website. URL: https://goo.su/17w1b (date of application: 01.04.24).

Krishnamoorthy, P., Sathiyanarayanan, M., & Proença, H. P. (2024). A novel and secured email classification and emotion detection using hybrid deep neural network. *International Journal of Cognitive Computing in Engineering*, 5, 44–57. https://doi.org/10.1016/j.ijcce.2024.01.002 *ScienceDirect:* website. URL: https://goo.su/AO1N (date of application: 01.04.24).

Patel, R., Thakkar, P., & Ukani, V. (2024). CNNRec: Convolutional Neural Network based recommender systems-A survey. *Engineering Applications of Artificial Intelligence*, 133, Part A, 108062. https://doi.org/10.1016/j.engappai.2024.108062 *ScienceDirect*: website. URL: https://goo.su/TUJHg4 (date of application: 01.04.24).

Qin, X., Li, L., & Pang, G. (2024). Multi-scale motivated neural network for image-text matching. *Multimedia Tools and Applications*, 83 (2), 4383–4407. https://doi.org/10.1007/s11042-023-15321-0 *Springer Link*: website. URL: https://goo.su/tidM (date of application: 01.04.24).

Ren, X. (2024). Construction of a thinking model for Literary Writing based on Deep Spatio-Temporal Residual Convolutional Neural Networks. *Multimedia Tools and Applications*, 1–22. https://doi.org/10.1007/s11042-023-18016-8 *SpingerLink*: website. URL: https://goo.su/OFRTj0 (date of application: 01.04.24).

Shalini, A. K., Saxena, S., & Kumar, B. S. (2024). Original Research Article Automatic detection of fake news using recurrent neural network – Long short-term memory. *Journal of Autonomous Intelligence*, 7 (3). https://doi.org/10.32629/jai.v7i3.798 *Frontier Scientific Publishing:* website. URL: https://goo.su /A1j0dAE (date of application: 01.04.24).

Thaokar, C., Rout, J. K., Rout, M., & Ray, N. K. (2024). N-Gram Based Sarcasm Detection for News and Social Media Text Using Hybrid Deep Learning Models. *SN Computer Science*, 5 (1), 163. *Springer Link*: website. URL: https://goo.su/SdFuvC (date of application: 01.04.24).

Tsirmpas, D., Gkionis, I., Papadopoulos, G. T., & Mademlis, I. (2024). Neural natural language processing for long texts: A survey on classification and summarization. *Engineering Applications of Artificial Intelligence*, 133, Part C, 108231. https://doi.org/10.1016/j.engappai.2024.108231 *ScienceDirect*: website. URL: https://goo.su/sKUQLc1 (date of application: 01.04.24).

Tu, H. (2024). Online Text Retrieval Method Based on Convolution Neural Network. *Journal of Multiple-Valued Logic & Soft Computing*, 42 (1–3), 159–177. *EBSCO:* website. URL: https://goo.su/YrVu5 (date of application: 01.04.24).

Wang, H. (2024). An Overview of the Application of Convolutional Neural Networks in Sentiment Analysis. *Science and Technology of Engineering, Chemistry and Environmental Protection*, 1 (5). https://doi.org/10.61173/t4sg2v25 *Dean & Francis Academic Publishing:* website. URL: https://goo.su /zSpKdIe (date of application: 01.04.24).

Wang, Y., & Zeng, Y. (2024). A brain-inspired computational model for human-like concept learning. *arXiv preprint arXiv:2401.06471*. *arXiv:* website. URL: https://goo.su/seplAN (date of application: 01.04.24).

Ye, N., Yu, D., Ma, X., Zhou, Y., & Yan, Y. (2024). TAI: a lightweight network for content-based fake news detection. *Online Information Review*. https://doi.org/10.1108/OIR-11-2022-0629 *Emerald Insight*: website. URL: https://goo.su/EvuS (date of application: 01.04.24).

**Бібліографічний опис**:

Довгань, О. В. (2024). Смислова диференціація текстів як складник нейромережевого моделювання. *Науковий часопис Українського державного університету імені Михайла Драгоманова. Серія 9. Сучасні тенденції розвитку мов*, 27, 15–27. https://doi.org/10.31392/NPU-nc.series9.2024.27.02.

***Анотація***

*У статті стверджується, що найпродуктивнішим для лінгвістичних досліджень на сучасному етапі є використання Artificial Neural Networks (ANNs) через їх продуктивність і репрезентативність. Підкреслено, що підґрунтям такого використання доцільно зробити смислову диференціацію, завдяки якій лінгвісти можуть оптимізувати пошук, аналіз даних для своїх досліджень. Зокрема, врахування семантичних, морфологічних, синтаксичних особливостей дозволить продукувати достовірніші, фундаментальніші результати у різних задачах Natural Language Processing (NLP). Автор акцентує увагу на тому, що наслідком цього стане якісний стрибок наукових досліджень українських мовознавців, можливість представлення їх результатів світовій науці й подальше плідне співробітництво з зарубіжними колегами у межах грантових програм.*

*Смислова диференціація текстів є невід'ємним складником актуалізації Artificial Neural Networks (ANNs) (зокрема, Bidirectional Long Short-Term Memory Network (BiLSTM), Convolutional*

*Neural Networks (CNNs), Deep Learning Networks (DLNs), Deep Neural Networks (DNNs), Graph Neural Networks (GNNs), Recurrent Neural Networks (RNNs) тощо) у сучасних лінгвістичних дослідженнях у межах цифрових гуманітарних наук. На думку автора, останнє продиктоване орієнтацією на практичний результат, локалізацією впровадження (зокрема, Natural Language Processing (NLP), sentiment analysis тощо). Відтак, подальше вивчення, вдосконалення та оптимізація наявного інноваційного інструментарію (зокрема нейромережевого моделювання мовних одиниць) – це діяльність щодо методів роботи з контекстом (через поєднання різних типів Artificial Neural Networks (ANNs) з різними шарами, що наочно представлене у працях зарубіжних колег), локалізації мовних стилів (особливо важливо в процесі фактчекінгових ініціатив – як віха до валідації текстових даних) без ручного втручання у ці процеси.*

***Ключові слова:*** *нейромережеве моделювання, прикладна лінгвістика, комп'ютерна лінгвістика, штучні нейронні мережі, нейромережеві моделі, смисл, смислова диференціація.*