Discourseology of Linguistic Consciousness: Neural Network Modeling of Some Structural and Semantic Relationships

Дискурсологія мовної свідомості: нейромережеве моделювання деяких структурних і семантичних взаємозв'зків

Vitalii Shymko
Dr. Sc. in Psychology,
Professor

Віталій Шимко
доктор психологічних наук,
професор

E-mail: shymko@outlook.com
https://orcid.org/0000-0003-4937-6976

*Pereiaslav-Khmelnytskyi
Hrihorii Skovoroda
State Pedagogical University*
✉ 30, Sukhomlynskyi Str.,
Pereiaslav, Kyiv Reg.,
Ukraine, 08401

*ДВНЗ «Переяслав-
Хмельницький державний
педагогічний університет
імені Григорія Сковороди»*
✉ вул. Сухомлинського, 30,
м. Переяслав, Київська обл.,
Україна, 08401

ABSTRACT
Objective. *Study of the validity and reliability of the discourse approach for the psycholinguistic understanding of the nature, structure, and features of the linguistic consciousness functioning.*
Materials & Methods. *This paper analyzes artificial neural network models built on the corpus of texts, which were obtained in the process of experimental research of the coronavirus quarantine concept as a new category of linguistic consciousness. The methodology of feedforward artificial neural networks (multilayer perceptron)*

was used in order to assess the possibility of predicting the leading texts semantics based on the discourses ranks and their place in the respective linear sequence. Same baseline parameters were used to predict respondents' self-assessments of changes in their psychological well-being and in daily life routine during the quarantine, as well as to predict their preferences of the quarantine strategies. The study relied on basic ideas about discourse as a meaning constituted by the dispersion of other meanings (Foucault). The same dispersion mechanism realizes itself in interdiscourse interaction, forming a discursive formation at a higher level. The method of T-units (Hunt) was used to identify and count discourses in the texts. The ranking of discourses was provided based on the criterion of their semantic-syntactic autonomy.

Results. *The conducted neural network modeling revealed a high accuracy in predicting the work of the linguistic consciousness functions associated with retrospective self-assessment and anticipatory imagination of the respondents. Another result of this modeling is a partial confirmation of the assumption concerning existence a relationship between the structural parameters of the discursive field (the rank of the discourses and their place in the respective linear sequence) and the leading semantics of the text.*

Conclusions. *A discourse approach to the study of linguistic consciousness, understanding of its structure and functioning features seems to be reasonably appropriate. The implementation of the approach presupposes the need to form a base of linguistic corpora with the inclusion in each text markup of such parameters as: the presence of specific discourses, their ranks, positions in the linear sequence of discourses.*

Key words: *discourse, discursive field, discourseology, linguistic consciousness, semantics, neural network, neural network modeling, corpus.*

Introduction

Linguistic consciousness (hereinafter – LC) occupies an important place in the subject matter field of psycholinguistics and several other humanitarian disciplines (Abildinova, 2018; Benda, 1959; Chafe, 1974; Chomsky, 2006; Galperin, 1992; Leontev, 1969 et al.). Researchers pay special attention to discourse as a qualitative and quantitative unit of LC (Foucault, 1972; Grosz, Joshi & Weinstein, 1995; Hart, 2016; Hunt, 1977; Joshi, Prince & Walker, 1998; Lacan, 1971; Machin, 2013; Ponton & Larina, 2016; Taboada & Zabala, 2008 et al.). A separate applied demand for the study of LC discourseology is actualized by the computer sciences dealing with the natural language processing issue (Batista-Navarro et al., 2013; Crossley et al., 2014; Forsythand &

Martell, 2007; Grosz & Sidner, 1986; Péry-Woodley & Scott, 2006; McNamara et al., 2018 et al.).

Discursive conceptualization of LC is not fundamentally new; however, the corresponding empirical studies are still at the early days stage. Among other things, interdiscourse dynamics remains poorly studied, the features of the interconnections of these dynamics with thinking and other mental processes have not been comprehensively investigated. Problems of this kind involve the need to consider at least dozens of different parameters simultaneously. This significantly complicates the use of traditional mathematical and statistical tools in the sense of cumbersome interpretations and generalizations of the findings. An obvious solution to this problem is the use of neural network modeling methodology. A corresponding attempt has been implemented in our work aimed at understanding some of the structural and semantic parameters of interdiscourse interactions.

This article is devoted to the analysis of a series of artificial neural network models (hereinafter – models), built on the corpus of texts obtained in the course of our earlier research (Shymko & Babadzhanova, 2020), in turn directed to clarify the content and structure of the coronavirus quarantine concept as an emerging category of LC. In the current work, we considered forecasting of the leading semantics of respondents' quarantine definitions based on the rank of discourses and their place in the corresponding linear sequence. The same parameters were used to predict the respondents' assessments of how much their psychological well-being and daily life routine changed during the quarantine, as well as to predict the quarantine strategies preferred by them. Thus, we pursued the main goal of this article – to study the validity and reliability of the discourse approach to the phenomenology of LC, to the relevant extent that can be substantiated by the size of the empirical sample and its other qualitative characteristics.

Methods and Techniques of the Research

The primary dataset (Shymko & Babadzhanova, 2020) has been properly modified for the purposes of the current study, the statistical results of which are posted separately (Shymko, 2020b). Below we provide a description of those sampling parameters that have undergone

the necessary restructuring and/or which were not detailed in the primary dataset but are essential for this study. So, firstly, the input layer of neural networks in all cases considered here is represented by various parameters that characterize the discursive field, consisting of 9 discourses (Shymko, 2020b: [list of discourses.pdf]). That is, the prediction was carried out based on data of the discourses' presence/ absence, their rank (see below) and place in the corresponding linear sequence. Any others, incl. socio-demographic data were not used in the construction of models for reasons of maintaining the homogeneity of predictors. Secondly, for the identification and counting of discourses in the texts we used the method of 'minimum terminable unit' or T-unit (Hunt, 1977), taking into account the generalizations of the empirical experience of this approach application made by Taboada & Zabala (2008). Thirdly, the ranking of discourses was based on the criterion of their semantic and syntactic autonomy. Namely, the only discourse in the text or one that is commented, explained (causally or in a purpose-oriented way), developed (semantically and/or syntactically) by another discourse (discourses) and at the same time itself does not perform such functions – it was evaluated as *independent* one. All the other discourses were assessed as *auxiliary* ones.

Example 1. (hereinafter, in square brackets – texts in the original language). «*Quarantine is the restriction of people contacts, aimed at preventing the spread of the virus*» [Карантин – это ограничение контактов людей, направленное на предотвращение распространения вируса]. In this text, we have identified two discourses: CONTACT RESTRICTION (independent) and VIRUS DISSEMINATION (auxiliary).

Example 2. «*Isolation of sick people from healthy people, for which it is necessary to introduce new tough laws. A large complex of anti-epidemiological measures (masks, antiseptics, etc.)*» [Изоляция больных людей от здоровых, для чего необходимо вводить новые жесткие законы. Большой комплекс противоэпидемиологических мероприятий (маски, антисептики и т.д.)]. Here we have recorded three discourses (ISOLATION OF INFECTED – independent; BUREAUCRATIC RESPONSE – auxiliary; SANITATION AND HYGIENE – independent).

Example 3. «*Quarantine is a set of regulatory and restrictive measures that are aimed at preventing the mass spread of an infectious*

disease and providing for a special regime of economic and other activities, restricting the movement of the population, vehicles, cargo, goods and animals» [Карантин – комплекс регуляторно-ограничительных мер, которые направлены на недопущение массового распространения инфекционного заболевания и предусматривающие особый режим хозяйственной и иной деятельности, ограничение передвижения населения, транспортных средств, грузов, товаров и животных]. In this example, we found one independent discourse (BUREAUCRATIC RESPONSE) and three auxiliary ones (in the order of their appearance in the text – VIRUS DISSEMINATION, LIFESTYLE CHANGES and CONTACT RESTRICTION).

Taking into account the above, the scale of variables reflecting the presence/absence of discourses in the primary dataset was reformatted from nominal to ordinal one and the data was recoded according to the identified ranks (the absence of discourse in the text was denoted by «zero» rank). Also, additional variables were added to the dataset: (a) «leading discourse» (nominal scale) – to display an independent discourse (if there were two or more such discourses in the text, the first one was accounted); (b) «first discourse» (nominal scale); (c) «second discourse» (nominal scale); (d) third discourse (nominal scale); (e) the number of discourses (metric scale); (f) «sequence of discourses» (nominal scale).

The construction of feedforward artificial neural networks was implemented by using the software IBM SPSS Statistics V 26 (Multilayer Perceptron). Given the small number of texts in the corpus, each model went through ten machine learning cycles. At the same time, each iteration was preceded by a random split of the dataset for cross-validation into the training and testing parts of the sample (the preset approximate proportion was 70% to 30%, respectively). All models have one hidden layer and a given range of the number of neurons in it: from 1 to 50. This paper discusses the general characteristics of the obtained models and the corresponding predictors importance without delving into the layer's synaptic architecture features. The latter would require a different publication format. However, detailed statistical reports for each iteration are published in the Harvard Dataverse repository and are available for review (Shymko, 2020b).

Results

First of all, those models were generated and compared that predict the texts semantics (leading discourse) based on the ranks of discourses (ordinal scale) and simply by the fact of their presence/absence (nominal scale). As expected, models with an ordinal scale in the input layer of the neural network have demonstrated a significantly lower percentage of errors (Table 1) with a generalized predictive reliability level of 99.74% and 93.88% at the stage of machine learning and testing, respectively.

Table 1. Models summary of the leading discourse prediction (predictors – discourses in texts)

Stages of neural network modeling	Average % of incorrect forecasts by 10 iterations	
	Nominal scale models	Ordinal scale models
Machine learning	11.1	0.26
Testing	21.06	6.12

It is noteworthy that at the 8th iteration, the only model (Fig. 1) with an ordinal scale in this series was obtained, the predictive reliability of which was 100%, both in the process of machine learning and during testing. Importance indicators of the predictors for this model are shown in Table 2.

Table 2. Predictors importance (ordinal scale, 8th iteration)

| Discourse	Predictor's code	Importance coefficient	Normalized Importance, %
CONTACT RESTRICTION	(D 1)	0.149	100.0
RIGHTS AND FREEDOMS INFRINGEMENT	(D 8)	0.129	86.2
VIRUS DISSEMINATION	(D 6)	0.127	85.3
SANITATION AND HYGIENE	(D 2)	0.115	76.9
TOTAL ISOLATION	(D 4)	0.111	74.1
HEALTH CARE	(D 5)	0.103	69.0
ISOLATION OF INFECTED	(D 3)	0.099	66.6
LIFESTYLE CHANGES	(D 7)	0.095	63.6
BUREAUCRATIC RESPONSE	(D 9)	0.073	48.7
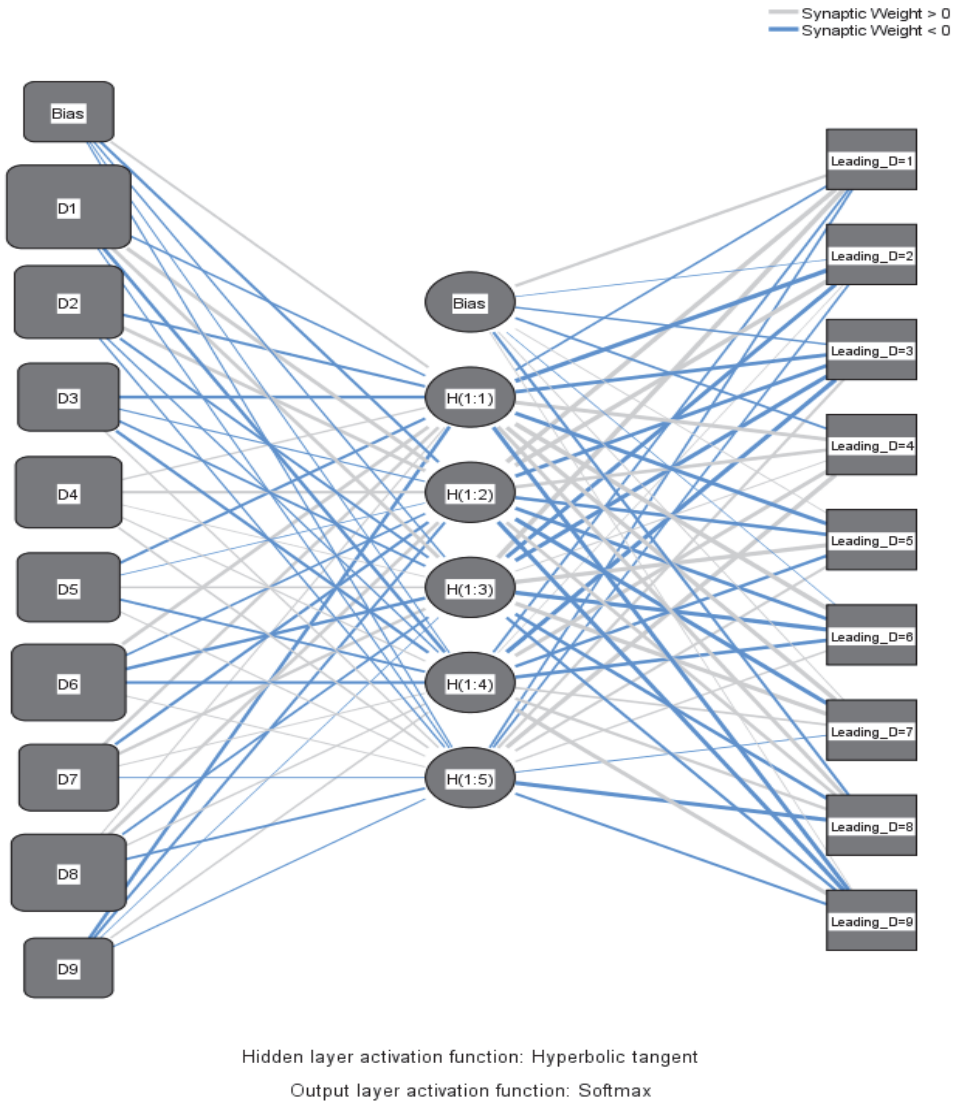
Figure 1. *The structure of the artificial neural network (ordinal scale, 8th iteration)*

Comparative analysis of the predictors importance in different iterations of this and other modeling series indicates the absence of a single (repeatable) hierarchy of variables. Considering the frequency parameters of the discourses' appearance in texts (Shymko & Babadzhanova, 2020: NKMRP4), this indirectly might point to the greater importance of interdiscourse interconnections in comparison with

the semantics of the discourses themselves. In other words, to predict the leading discourse in the text, the significance of the structural characteristics of the discursive field prevails over the semantic features of its units.

We find confirmation of the foregoing in the next two series of models, with the help of which we have determined the leading discourse on the basis of nominal distinction in a linear sequence of no more than the two first and no more than the three first discourses (Table 3). Let us recall that the corpus consists of texts with the number of discourses in each from 1 to 5.

Table 3. Models summary of the leading discourse prediction (predictors – few first discourses in texts)

Stages of neural network modeling	Average % of incorrect forecasts by 10 iterations	
	Models with the first two discourses	Models with the first three discourses
Machine learning	3.16	0.00
Testing	7.86	0.00

As one can see, subject to distinguishing the first three discourses, neural network modeling provides the most reliable prediction of the leading discourse. We additionally verified this conclusion by conducting another 10 iterations in two series of models with an input layer that simultaneously distinguishes the first three discourses and takes into account the presence/absence of discourses (one series – using the nominal scale, other one – using the ordinal). In all cases and at all stages level of 100% predictive accuracy were obtained (Shymko, 2020b).

Using a similar architecture of the input layers of neural networks, we continued to predict how much the psychological well-being and daily life routine of the respondents changed during the quarantine span, as well as what were their preferences for respective quarantine strategies. Note that in the latter case, we predicted the corresponding rank, set by the respondents to each of the four strategies they were proposed to evaluate. The results are summarized in Table 4.

In most cases, the quality of models with ordinal scales used for discourses is higher. Some exceptions to the effectiveness of machine learning (DAILY LIFE ROUTINE and SYSTEMIC STRATEGY) may

be appeared due to the complication of interdiscourse connections because of the discourses rank parameters consideration. That is, we assume the increase in the number of errors has been observed due to the processing of more complex structures. Nevertheless, statistics on incorrect predictions at the stage of models testing confirm the higher productivity of neural network architectures that distinguish not only interdiscourse interactions, but also the corresponding hierarchical characteristics.

Table 4. Models summary of the non-discourse parameters prediction

Predicted variables	Average % of incorrect forecasts by 10 iterations			
	Machine learning		Testing	
	Scale used for discourses		Scale used for discourses	
	Nominal	Ordinal	Nominal	Ordinal
Psychological well-being	28.44	8.30	0.00	0.00
Daily life routine	7.52	18.23	2.5	0.00
Quarantine strategies DEMOCRATIC.	14.33	11.79	3.33	0.00
SYSTEMIC	8.14	12.77	0.00	0.00
MILITARY LIKE	17.96	8.22	0.00	0.00
SOCIAL PROTECTION	29.48	16.15	5.00	0.00

Discussions

Discussion of the results described above requires consideration of the limitations that are objectively inherent to any research. Firstly, it is necessary to pay attention to the quantitative characteristics of the primary sample (Shymko & Babadzhanova, 2020: NKMRP4), which does not allow proper generalizing the conclusions. The same applies to the qualitative aspect of the research empirical base, namely, the presence in the corpus of only Russian-language texts and, as a consequence, the impossibility of making comparisons with texts in other languages of the Slavic group, as well as languages from other groups and branches.

Secondly, certain limitations are directly due to the design of the study itself. Thus, the authors of texts with quarantine definitions were asked to assess changes in their psychological well-being and everyday life routine during quarantine, i.e. the corresponding measurements were carried out within the framework of the phenomenologically associated

semantic space. It seems promising to expand such a space by making discourse measurements of LC activity products in other semantic spheres and in variety of others (non-laboratory) situations.

Thirdly, when interpreting the results of this study, it is important to consider the limitations caused by its theoretical (conceptual) focuses and methodological (procedural) aspects. So, we proceeded from the basic ideas about discourse, as of the meaning generated by the dissemination of other meanings (Foucault, 1972). The same mechanism realizes itself at a higher level (in interdiscourse interaction), forming a new structural-semantic «layer» – a discourse formation (Shymko, 2020a). We believe that such dynamics are involved in the organization and functioning of the entire LC. At the same time, we do not exclude alternative views both on the nature of discourses and on the technical methods of their localization and counting in texts. Cross-examination of texts from different theoretical and methodological positions and the subsequent comparative analysis contains a very fruitful perspective from a heuristic point of view.

Given the above constraints, what inferences and extrapolations are possible from the described series of neural network modeling? In our opinion, the useful potential of this study is primarily methodological. So, on the one hand, the effectiveness of the discourse approach to the study of LC phenomena has been empirically tested and confirmed. In particular, this manifested itself in the prediction results of respondents' self-assessments and preferences based on discourse characteristics of their texts. The very fact of the possibility of such a prediction determines the need to formulate several hypotheses regarding the place and role of discursive mechanisms in the work of a thinking as function of LC. We assume that it is advisable to consider thinking not only as operating with information that is organized discursively. Thinking, by itself, functions discursively, i.e. in a procedural sense, thinking is a discursive process. Among other things, such a view would well explain why interdiscourse relationships are more important for predicting the leading discourse than the semantics of the discourses themselves, as discussed above.

One way or another, we convinced that the texts discourseology can be effectively used for predictive assessment of the LC functions and its activity products, in particular: retrospective assessment (changes in psychological well-being and daily life routine) and anticipatory

imagination (quarantine strategies ranking). We also believe that the hypothesis about the relationship of the discursive field structural parameters (the discourses ranks and their place in the linier sequence) with the leading semantics of the text has been partially substantiated. However, here it is necessary to distinguish between the discursive field of a text and the discursive field of the corpus. We emphasize that the specified relationship is come out if the composition of both fields is known and accounted for. Thus, the question of the corpus approach to discourseology of LC is actualized. Full verification of this hypothesis presupposes overcoming the above limitations of our research.

On the other hand, this study can be useful in terms of some experience in applying the methodology of artificial neural networks to solve research problems in psycholinguistics on relatively small samples. The need to simultaneously consider a large number of inhomogeneous factors is a typical feature of interdisciplinary research. However, the use of neural networks is not only a matter of the convenience of a holistic view on a patchwork picture of variables. Multifactoriality in psycholinguistics is accompanied by additional complexity in the form of such a phenomenon as the entropy of language (Bentz et al., 2017). All this reduces the sensitivity of traditionally used statistical data processing tools to psycholinguistic patterns. In this respect, artificial neural networks are perhaps the best methodological alternative for now.

Conclusions

A discourse approach to the study of linguistic consciousness, understanding of its structure and functioning features seems to be reasonably appropriate. The implementation of this approach presupposes the need to form a base of linguistic corpora with the inclusion in the markup of each text of such parameters as – the presence of specific discourses, their ranks, positions in the corresponding linear sequence. The conducted neural network modeling on the data of the local corpus (texts of Russian-speaking Ukrainians containing the coronavirus quarantine definitions) reveals a high accuracy in predicting the activity results of some linguistic consciousness functions – retrospective self-assessments and products of the anticipatory imagination. Another result of this modeling is a partial confirmation of the assumption about

relationship between the structural parameters of the discursive field (the rank of the discourses and their place in the respective linear sequence) and the leading semantics of the text.

The most important, in our opinion, prospects for further research are associated with overcoming the limitations of current research. Namely, the quantitative and qualitative increase in the corpus base, as well as the diversification of text sources, which would provide the possibility of forming a cross-linguistic discourseology of linguistic consciousness.

References

Abildinova, Z.B. (2018). Yazykovoe soznanie kak psikholingvisticheskiy fenomen [Linguistic consciousness as a psycholinguistic phenomenon]. *Neofilologiya – Neophilology, 4*(14), 33–43. https://doi.org/10.20310/2587-6953-2018-4-14-33-43 [in Russian].

Batista-Navarro, R.T. et al. (2013). Facilitating the Analysis of Discourse Phenomena in an Interoperable NLP Platform. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing. CICLing' 2013*. Lecture Notes in Computer Science (Vol. 7816). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37247-6_45

Benda, C. (1959). The linguistic basis of consciousness. *ETC: A Review of General Semantics, 16*(3), 343–355. Retrieved January 20, 2021, from http://www.jstor.org/stable/24234383

Bentz, C., Alikaniotis, D., Cysouw, M., & Ferrer-i-Cancho, R. (2017). The Entropy of Words – Learnability and Expressivity across More than 1000 Languages. *Entropy*, *19*, 275. https://doi.org/10.3390/e19060275

Chafe, W. (1974). Language and Consciousness. *Language, 50*(1), 111–133. https://doi.org/10.2307/412014

Chomsky, N. (2006). *Language and Mind* (3rd ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511791222

Crossley, S., Allen, L., Kyle, K., & McNamara, D. (2014). Analyzing Discourse Processing Using a Simple Natural Language Processing Tool. *Discourse Processes, 51*(5–6), 511–534. https://doi.org/10.1080/0163853X.2014.910723

Forsythand, E.N., & Martell, C.H. (2007). Lexical and Discourse Analysis of Online Chat Dialog. *Proceedings of the International Conference on Semantic Computing (ICSC 2007)*, (Irvine, California, 17–19 September, 2007), (pp. 19–26). Irvine, CA. https://doi.org/10.1109/ICSC.2007.55

Foucault, M. (1972). *Archaeology of knowledge and the discourse on language*. New York: Pantheon.

Galperin, P.I. (1992). Linguistic Consciousness and Some Questions of the Relationship between Language and Thought. *Journal of Russian & East European Psychology, 30*(4), 81–92. https://doi.org/10.2753/RPO1061-0405300481

Grosz, B.J., & Sidner, C.L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics, 12*(3), 175–204.

Grosz, B.J., Joshi, A.K., & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics, 21*(2), 203–225. https://doi.org/10.21236/ADA324949

Hart, C. (2016). The visual basis of linguistic meaning and its implications for critical discourse studies: Integrating cognitive linguistic and multimodal methods. *Discourse & Society, 27*(3), 335–350. https://doi.org/10.1177/0957926516630896

Hunt, K.W. (1977). Early blooming and late blooming syntactic structures. In C.R. Cooper, & L. Odell (Eds.), *Evaluating Writing: Describing, Measuring and Judging* (pp. 91–104). Urbana, Illinois: NCTE.

Lacan, J. (1971). *The seminar of Jacques Lacan. Book XVIII: On a discourse that might not be a semblance 1971.* (Gallagher, C., Trans.). Retrieved from: https://hdl.handle.net/10788/167

Leontev, A.A. (1969). *Psiholingvisticheskie edinicy i porozhdenie rechevogo vyskazyvanija [Psycholinguistic units and the generation of speech utterance].* Moscow. Nauka [in Russian].

Machin, D. (2013). What is multimodal critical discourse studies? *Critical Discourse Studies, 10*(4), 347–355. https://doi.org/10.1080/17405904.2013.813770

McNamara, D., Allen, L.K., McCarthy, K.S., & Balyan, R. (2018). NLP: Getting computers to understand discourse. In *Deep Comprehension: Multi-Disciplinary Approaches to Understanding, Enhancing, and Measuring Comprehension* (pp. 224–236). Taylor and Francis. https://doi.org/10.4324/9781315109503

Péry-Woodley, M.-P., & Scott, D. (2006). Computational approaches to discourse and document processing. *TAL, 2*(47), 7–19.

Ponton, D.M., & Larina, T.V. (2016). Discourse Analysis in the 21st Century: Theory and Practice (I). *Russian Journal of Linguistics, 20*(4), 7–25. https://doi.org/10.22363/2312-9182-2016-20-4-7-25

Shymko, V. (2020a). Linguistic Semantics of the Covid-19 Quarantine Ideation. In L. Kalmykova, N. Kharchenko, & I. Mysan (Eds.), *Proceedings of the XV International and Practical Conference «Psycholinguistics in a Modern World – 2020» (Pereiaslav-Khmelnytskyi, 22–23 October, 2020),* (Vol. 15, pp. 297–302). https://doi.org/10.31470/2706-7904-2020-15-297-302

Shymko, V. (2020b). Neural network predictive modeling of Covid-19 related quarantine strategies preferences. *Harvard Dataverse.* https://doi.org/10.7910/DVN/OH8HQC

Shymko, V., & Babadzhanova, A. (2020). Study of the COVID-19 related quarantine notion as the language consciousness category on the sample of Russian speaking Ukrainians. *Harvard Dataverse.* https://doi.org/10.7910/DVN/Y2YGH7

Taboada, M., & Zabala, L.H. (2008). Deciding on units of analysis within Centering Theory. *Corpus Linguistics and Linguistic Theory*, *4*(1), 63–108. https://doi.org/10.1515/CLLT.2008.003

Walker, M., Joshi, A., & Prince, E. (Eds.). (1998). *Centering Theory in Discourse.* Clarendon Press.

АНОТАЦІЯ

Мета. *Вивчення валідності та надійності дискурсологічного підходу для психолінгвістичного розуміння природи, побудови та особливостей функціонування мовної свідомості.*

Методики дослідження. *У даній роботі проведено аналіз штучних нейромережевих моделей, збудованих на корпусі текстів, які були отримані в процесі експериментального дослідження концепту коронавірусного карантину, як нової категорії мовної свідомості. Використана методологія нейромереж прямого поширення (багатошаровий перцептрон) з метою оцінки можливостей прогнозування провідної семантики текстів на основі рангових характеристик дискурсів та їх місця у відповідній лінійної послідовності. Аналогічні вихідні параметри використані для предикції самооцінок респондентів щодо зміни їх психологічно самопочуття і повсякденного життя в період карантину, а також для передбачення їх переважань щодо доцільних карантинних стратегій. При цьому дослідження спиралося на базові уявлення про дискурс, як про значення, що породжується розсіюванням інших значень (Foucault). Цей же механізм розсіювання реалізує себе в міждискурсній взаємодії, формуючи на більш високому рівні дискурсивну формацію. Для ідентифікації і підрахунку дискурсів в текстах використовувався метод Т-юнітів (Hunt). Ранжування дискурсів реалізовано за критерієм їх семантико-синтаксичної автономії.*

Результати. *Проведене нейромережеве моделювання виявило високу точність предикції результатів роботи функцій мовної свідомості, пов'язаних з ретроспективним самооцінювання і антиципуючою уявою респондентів. Ще одним результатом зазначеного моделювання є часткове підтвердження припущення про наявність взаємозв'язку між структурними параметрами дискурсивного поля (ранг присутніх дискурсів та їх місце у відповідній лінійній послідовності) та провідною семантикою тексту.*

Висновки. *Дискурсологічний підхід до вивчення мовної свідомості, розуміння її побудови та особливостей функціонування – представляється обґрунтовано доцільним. Реалізація такого підходу передбачає необхідність формування бази лінгвістичних корпусів з включенням в розмітку кожного тексту таких параметрів, як-от: наявність конкретних дискурсів, їх ранги, позиції в лінійній послідовності дискурсів тексту.*

Ключові слова: *дискурс, дискурсивне поле, дискурсологія, мовна свідомість, семантика, нейронна мережа, нейромережеве моделювання, корпус.*

Шимко Виталий. Дискурсология языкового сознания: нейросетевое моделирование некоторых структурных и семантических взаимосвязей

АННОТАЦИЯ
Цель. *Изучение валидности и надежности дискурсологического подхода для психолингвистического понимания природы, устройства и особенностей функционирования языкового сознания.*

Методики исследования. *В данной работе проведен анализ искусственных нейросетевых моделей, построенных на корпусе текстов, которые получены в процессе экспериментального исследования концепта коронавирусного карантина, как новой категории языкового сознания. Использована методология*

© *Шимко Віталій*

нейросетей прямого распространения (многослойный перцептрон) с целью оценки возможностей прогнозирования ведущей семантики текстов на основе ранговых характеристик, содержащихся в них дискурсов и их места в соответствующей линейной последовательности. Аналогичные исходные параметры использованы для предикции самооценок респондентов об изменениях их психологическое самочувствие и повседневной жизни в период карантина, а также для предсказания предпочитаемых ими карантинных стратегий. При этом исследование опиралось на базовые представления о дискурсе, как о значении, порождаемом рассеиванием других значений (Foucault). Этот же механизм рассеивания реализует себя в междискурсном взаимодействии, формируя на более высоком уровне дискурсивную формацию. Для идентификации и подсчета дискурсов в текстах использовался метод Т-юнитов (Hunt). Ранжирование дискурсов реализовано по критерию их семантико-синтаксической автономии.

Результаты. Проведенное нейросетевое моделирование обнаруживает высокую точность предикции работы функций языкового сознания, связанных с ретроспективным самооцениванием и антиципирующим воображением респондентов. Еще одним результатом указанного моделирования является частичное подтверждение предположения о наличии взаимосвязи между структурными параметрами дискурсивного поля (ранг присутствующих дискурсов и их линейная последовательность) и ведущей семантикой текста.

Выводы. Дискурсологический подход к изучению языкового сознания, пониманию его устройства и особенностей функционирования – представляется обоснованно целесообразным. Реализация такого подхода предполагает необходимость формирования базы лингвистических корпусов с включением в разметку каждого текста таких параметров, как – наличие конкретных дискурсов, их ранги, позиции в линейной последовательности дискурсов текста.

Ключевые слова: дискурс, дискурсивное поле, дискурсология, языковое сознание, семантика, нейронная сеть, нейросетевое моедлирование, корпус.