

РЕДАКТОРСКАЯ СТАТЬЯ  
EDITORIAL

UDC [81'23+81'322]:004

DOI: 10.18413/2313-8912-2023-9-1-0-1

Valery D. Solovyev<sup>1</sup> 

Mihai Dascalu<sup>2</sup> 

Marina I. Solnyshkina<sup>3</sup> 

Discourse complexity: driving forces of the new paradigm

<sup>1</sup> Kazan (Volga region) Federal University  
18 Kremlevskaya St., Kazan, 420008, Russia  
E-mail: [maki.solovyev@mail.ru](mailto:maki.solovyev@mail.ru)

<sup>2</sup> Polytechnic University of Bucharest  
313 Splaiul Independentei, Sector 6, Bucharest, 060042, Romania  
E-mail: [mihai.dascalu@upb.ro](mailto:mihai.dascalu@upb.ro)


<sup>3</sup> Kazan (Volga region) Federal University  
18 Kremlevskaya St., Kazan, 420008, Russia  
E-mail: [mesoln@yandex.ru](mailto:mesoln@yandex.ru)


**How to cite:** Solovyev, V. D., Dascalu, M. and Solnyshkina, M. I. (2023). Discourse complexity: driving forces of the new paradigm, *Research Result. Theoretical and Applied Linguistics*, 9 (1), 4-10. DOI: 10.18413/2313-8912-2023-9-1-0-1

УДК [81'23+81'322]:004

DOI: 10.18413/2313-8912-2023-9-1-0-1

Соловьев В. Д.<sup>1</sup> 

Даскалу М.<sup>2</sup> 

Солнышкина М. И.<sup>3</sup> 

Движущие силы новой парадигмы дискурсивной  
комплексологии

<sup>1</sup> Казанский (Приволжский) федеральный университет  
ул. Кремлевская, 18, Казань, 420008, Россия  
E-mail: [maki.solovyev@mail.ru](mailto:maki.solovyev@mail.ru)

<sup>2</sup> Политехнический университет Бухареста  
313 Сплаиул Индепендетей, Сектор 6, Бухарест, 060042, Румыния  
E-mail: [mihai.dascalu@upb.ro](mailto:mihai.dascalu@upb.ro)

<sup>3</sup> Казанский (приволжский) федеральный университет  
ул. Кремлевская, 18, Казань, 420008, Россия  
E-mail: [mesoln@yandex.ru](mailto:mesoln@yandex.ru)

**Информация для цитирования:** Соловьев В. Д., Даскалу М., Солнышкина М. И. Движущие силы новой парадигмы дискурсивной комплексологии // Научный результат. Вопросы теоретической и прикладной лингвистики. 2023. Т. 9. № 1. С. 4-10. DOI: 10.18413/2313-8912-2023-9-1-0-1

The modern area of research into text/discourse complexity (hereinafter DC)

has made significant advances with the advent of Natural Language Processing techniques.

The latest achievements are numerous, and we suggest classifying them into three groups: (1) re-defining the notion of ‘complexity’ and differentiating it from related concepts, such as subjective-assessed difficulty or comprehensibility (Dascalu et al., 2018; Solnyshkina and Kisel'nikov, 2015; Botarleanu et al., 2022); (2) types of complexities identified and described: lexical and syntactic, ‘absolute’ and ‘relative’ (McNamara, 2011; Solnyshkina et al., 2022); (3) expanding research data from ‘linear text only’ to ‘non-linear texts’ (Wenger and Payne, 1996). All the above are outcomes of numerous intensive research aimed at quantitative documentation of numerous patterns of text types and distributions of text features (Biber et al., 2021; Corlatescu, Ruseti and Dascalu, 2022; Gatiyatullina et al., 2020) on the one hand, and readers’ abilities on the other (McNamara, Levinstein and Boonthum, 2004).

DC emerged and was developed to respond to social demands targeting the improvement of the population’s reading literacy. Since the first published research of Sherman (1893), research in the field has always pursued a pragmatic approach encapsulated in its main question – “what makes a text difficult/non-readable/incomprehensible?” In the early 2000s, after seminal works of psycholinguists (Kintsch, 1998; Wolfe et al., 1998), the question was specified to “what makes a text difficult for a certain category of readers” (Crossley, Greenfield and McNamara, 2008), thus widening the object of studies from ‘a text’ to ‘a text and a reader,’ or more specifically ‘text – reader alignment’ (McNamara, Levinstein and Boonthum, 2004).

Professional jargon traditionally distinguishes text features/parameters and readers’ characteristics: while texts are explored for ‘complexity predictors’ (i.e., text features impacting its comprehension), readers are examined for their ‘criteria’ (i.e., abilities to comprehend a certain category of texts). These abilities are usually defined as

cognitive and behavioral patterns, including motivation, working memory, anxiety, possible speech impairment, general and specific knowledge, and language proficiency (Dascalu, McNamara, Crossley, Trausan-Matu, 2016).

The pragmatic dimension of DC resulted in its broad inter-disciplinary focus (Solnyshkina, Kharkova and Kazachkova, 2020) and employment of neurological (Martínez-Santiago et al., 2023), cognitive (Putra, Lukmana, 2017; Lyashevskaya, Pyzhak and Vinogradova, 2022; Laposhina, Lebedeva and Berlin Khenis 2022), and Artificial Intelligence methods (Ivanov, 2022; Sharoff, 2022).

The prospects of modern research in DC lie in exploring mechanisms of text complexity adjustment (i.e., simplification) and identifying text features and interdependent clusters of text features (Shardlow, 2014).

The current issue is composed of three sections:

**SECTION I: Text complexity predictors: Methods and approaches for assessment,**

**SECTION II: Cognitive mechanisms of text comprehension and**

**SECTION III: Neural networks for Natural Language Processing.**

This division into sections is designed to make the presented information manageable and easier to discuss. Each section contains articles on one of the most important constituents of text comprehension analysis: the object (i.e., a text), the subject (i.e., either a reader or a listener), and the employed methods.

In **SECTION I: Text Complexity predictors: Methods and approaches for assessment**, we collected the research focused mainly on quantifying features predictive of text comprehension.

In “Classification of Russian Textbooks by Grade Level and Topic using Reader-Bench” by A. Paraschiv, M. Dascalu, and M. Solnyshkina, the reader finds analyses and the implementation of automated classifica-

tion methods applied to a dataset of 154 Russian textbooks. The authors' focus is on predicting the topic and text complexity. The authors measure text indices with the help of ReaderBench, a multilingual open-source platform, and then use them in conjunction with BERT-based models. The results indicate that text complexity indices complement the contextualized embeddings while improving the classification performance of BERT-based models.

The article "Terminology use in school textbooks: A corpus analysis" by S. I. Monakhov, V. V. Turchanenko, and D. N. Cherdakov presents an in-depth study of Russian school textbooks' terminological system. The research develops a method of terminology retrieval and contributes to compiling a database of Russian school terms assigned to a specific discipline and school level. The authors develop and apply an original approach using vector semantics based on the distributive hypothesis. They also consider (dis)similarities of terminology in school textbooks, science, popular literature, and vernacular. The researchers argue that the number and diversity of terms in a text are predictors of its lexical complexity. The authors conclude that the nature of interdependence between text complexity and principles of its didactic effectiveness is contradictory.

In "Lexical density as a complexity predictor: The case of Science and Social Studies textbooks", G. M. Gatiyatullina, M. I. Solnyshkina, R. V. Kupriyanov, and C. R. Ziganshina explore the ratio of different parts of speech and their effect on readability in American textbooks across grades (7-12) and disciplines. The analysis confirmed the trend of the strong positive growth of nouns and adjectives and the decrease in lexical verbs from grades 7 to 11. The study reveals minor, though statistically significant, differences between social studies and natural science textbooks which could be used in automatic text profiling. The authors conclude that multidirectional dynamics of verbal and nominal elements across grades result in the gen-

eral nominalization of both discourses with lower readability values in natural science textbooks.

In SECTION II: **Cognitive mechanisms of text comprehension**, we present studies exploring the subject of text comprehension, either a reader or a listener. The section opens with the article "Silent, but salient: Gestures in simultaneous interpreting" by O. K. Iriskhanova, A. J. Cienki, M. V. Tomskeya, and A. I. Nikolayeva, which explores salience in gestures of simultaneous interpreters. It is a landmark study of a specific communicative situation left beyond the research paradigm before. The authors conduct a rigorous empirical study of gestures in simultaneous translation and suggest classifying them into salient and non-salient types. The study advocates that the 2nd type of gesture is performed about twice as often as the salient gestures. Researchers also offer a detailed description of elementary discursive units, most often accompanied by salient gestures. The obtained results are also consistent with the earlier research that gestures are "windows into an individual's thoughts" and lead to a more robust interpretation of the multimodal nature of meaning in the communication of simultaneous interpreters.

The study presented by M. I. Kiose, A. I. Izmalkova, A. A. Rzhesheshevskaya, and S. D. Makeev in "Text and metatext event in the gaze behavior of impulsive and reflective readers" is focused on oculomotor behavior of readers. The authors use standard research tools and explore two questions. First, the authors investigate the effect of the structure of events in the text (play) on oculomotor behavior using an original corpus of MultiCORText. For this purpose, researchers annotated MultiCORText, developed in the framework of the current analysis, to enable marking specifics of constructing events. The study revealed differences in constructing events in the author's and characters' utterances. The second research question concerns the interdependence of oculomotor behavior and the cognitive styles of readers. To achieve it, the authors compared the

behavior of impulsive and reflective readers to find out statistically different peculiarities of constructing events.

The article “Numbers in simultaneous interpreting: a multimodal analysis” by A. Cienki, A. V. Leonteva, O. V. Agafonova, and A. A. Petrov defines and describes cognitive strategies of simultaneous interpreters while dealing with numbers in the source text. The study considers a multimodal analysis of numbers in simultaneous interpreting focused on the generated texts and accompanying gestures. The research shows that interpreters tend to skip numbers in the target language and that gestures function as adapters assisting interpreters in coping with the extra cognitive load imposed by numbers.

V. Solovyev, Yu. Vol'skaya, R. Akhtiamov, in their article “Range of associations to Russian abstract and concrete nouns,” focus on associations that native Russian speakers develop while acquiring abstract and concrete nouns. The dataset with 100 words having the highest degree of concreteness/abstractness was retrieved from the “Russian Associative Dictionary” by Yu. Karaulov. The research findings indicate that all abstract nouns develop a wider range of associations, while concrete nouns evolve much stronger associations, thus confirming their consistency with the Context Accessibility Theory (CAT). The authors also propose a classification of associations based on the type of interdependence of stimulus words and associations. The study also argues for a striking consistency between the results of lingo-statistical and neuro-physiological analyses of abstract/concrete words.

The article “Specifics of Text Derivatives Propositions in Ontogeny” by A. A. Petrova, I. V. Privalova, M. B. Kazachkova, and K. U. Yessenova explores the nature of text recalls generated by Russian 5th-graders and offers a classification based on the number of reproduced propositions. The study is focused on the concept of deep semantic roles and their transformations in recalls. The latter

reflects changes in exponential and contentive parts of the signs. The authors argue that the collected data demonstrate specifics of the cognitive growth of different groups of teenagers and are consistent with the principle of ‘generating virtual dialogue partners’. The corpus of recalls used in their study was provided by the “Text Analytics” Lab, the right-holder of the Corpus of Sounding Speech compiled at Kazan Federal University.

I. V. Blinnikova, M. D. Rabeson, G. B. Blinnikov, and A. I. Izmalkova present their research “Complexity of visual semantic search in the first and second languages: eye-movement analysis,” in which they compare the oculomotor activity of native and non-native speakers when performing a semantic search. They use a popular intellectual conundrum in which subjects search for words in squares with randomly arranged letters. The squares, sized 15x15, contain letters of 10 words lined up vertically and horizontally. As expected, the word search in the native language proves to be more effective, and its strategies in the native and foreign languages differ dramatically. Native speakers’ strategy consists of longer fixations and shorter saccades. Non-native speakers’ behavior is more chaotic, with longer saccades and shorter fixations. The findings also support the effectiveness and interdependence of the employed strategies on the one hand, and word frequency, letter overlap, and emotiveness on the other.

**SECTION III: Neural Networks in Natural Language Processing** focuses on the vanguard methods of language analysis (i.e., neural networks) and opens with the article “A deep neural method based on language models for processing natural language Russian commands in human-robot interactions” by A. G. Sboev, A. V. Gryaznov, R. B. Rybka, M. S. Skorokhodov, and I. A. Moloshnikov. The study is focused on the increasingly urgent problem of organizing effective human-robot speech interaction. The authors propose translating natural language commands into a format of formalized graphs adequate for subsequent processing. To fulfill

this and other complex problems, researchers successfully solve at least two challenging problems: identifying pronouns referents and reconstructing ellipsis. For these purposes, they apply language models based on the Transformer architecture. The algorithms were implemented and validated in a three-dimensional virtual model of a robotic device developed at the National Research Center “Kurchatov Institute”.

The article “Parametrizing number variation in Russian noun phrases with experimental studies and language modeling” by K. A. Studenikina explores the long-standing issue of the category of numbers in modifiers in Russian coordinative constructions and presents her view on the morphosyntactic factors impacting this choice. The author interviewed informants using Yandex.Toloka and trained a neural network to predict the modifiers' form. The findings imply that the neural network makes correct predictions in simple cases but does not cope well with ambiguous contexts.

### References

- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (2021). *Grammar of spoken and written English*, John Benjamins, Amsterdam, Netherlands.  
<https://doi.org/10.1075/z232> (In English)
- Botarleanu, R., Dascalu, M., Watanabe, M., Crossley, S. A. and McNamara, D. S. (2022). Age of Exposure 2.0: Estimating word complexity using Iterative models of word embeddings, *Behavior Research Methods*, 54, 3015-3042.  
<https://doi.org/10.3758/s13428-022-01797-5> (In English)
- Corlatescu, D., Ruseti, Ş, Dascalu, M. (2022). ReaderBench: Multilevel analysis of Russian text characteristics, *Russian Journal of Linguistics*, 26 (2), 342–370.  
<https://doi.org/10.22363/2687-0088-30145> (In English)
- Crossley, S. A., Greenfield, J. and McNamara, D. S. (2008). Assessing Text Readability Using Cognitively Based Indices. *TESOL Quarterly*, 42 (3), 475–493.  
<http://www.jstor.org/stable/40264479> (In English)
- Dascalu, M., Crossley, S. A., McNamara, D. S., Dessus, P. and Trausan-

Matu, S. (2018). Please Readerbench this text: A multi-dimensional textual complexity assessment framework, in Craig, S. (ed.), *Tutoring and Intelligent Tutoring Systems*, Nova Science Publishers, Hauppauge, NY, 251–271. (In English)

Dascalu, M., McNamara, D. S., Crossley, S. A. and Trausan-Matu, S. (2016). Age of exposure: A model of word learning, in Zilberstein, S., Schuurmans, D. and Wellman, M. (eds.), *Proceedings of the 30th Annual Meeting of the Association for the Advancement of Artificial Intelligence (AAAI'16)*, AAAI Press, Phoenix, AZ, 2928–2934. (In English)

Gatiyatullina, G., Solnyshkina, M., Solovyev, V., Danilov, A., Martynova, E. and Yarmakeev, I. (2020). Computing Russian Morphological distribution patterns using RusAC Online Server, *Proceedings of the 13th International Conference on Developments in eSystems Engineering (DeSE)*, Liverpool, United Kingdom, 393-398.  
<https://doi.org/10.1109/DeSE51703.2020.9450753> (In English)

Ivanov, V. V. (2022). Sentence-level complexity in Russian: An evaluation of BERT and graph neural networks, *Frontiers in Artificial Intelligence*, 5.  
<https://doi.org/10.3389/frai.2022.1008411> (In English)

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*, Cambridge University Press, Cambridge, MA. (In English)

Laposhina, A. N., Lebedeva, M. Yu. and Berlin Khenis, A. A. (2022). Word frequency and text complexity: an eye-tracking study of young Russian readers, *Russian Journal of Linguistics*, 26 (2), 493–514. <https://doi.org/10.22363/2687-0088-30084> (In Russian)

Lyashevskaya, O. I., Pyzhak, J. V., Vinogradova, O. N. (2022). Word-formation complexity: a learner corpus-based study, *Russian Journal of Linguistics*, 26 (2), 471–492.  
<https://doi.org/10.22363/2687-0088-31187> (In English)

Martínez-Santiago, F., Torres-García, A. A., Montejo-Ráez, A. et al. (2023). The impact of reading fluency level on interactive information retrieval, *Universal Access in the Information Society*, 22, 51–67.

<https://doi.org/10.1007/s10209-021-00826-y> (In English)

McNamara, D. S. (2011). *Coh-Metrix: Its role in readability and the case for cohesion*, Panel presentation for Exploring the Common Core standards' approach to text complexity at 57th Annual Convention of the International Reading Association, Orlando, FL. (In English).

McNamara, D. S., Levinstein, I. B. and Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking, *Behavior Research Methods, Instruments, & Computers*, 36 (2), 222-233. <https://doi.org/10.3758/bf03195567> (In English)

Putra, D. A. and Lukmana, I. (2017). Text complexity in senior high school English textbooks: A systemic functional perspective, *Indonesian Journal of Applied Linguistics*, 7 (2), 436-444. <https://doi.org/10.17509/ijal.v7i2.8352> (In English)

Shardlow, M. (2014). A Survey of Automated Text Simplification, *International Journal of Advanced Computer Science and Applications*, 4. <http://dx.doi.org/10.14569/SpecialIssue.2014.040109> (In English)

Sharoff, S. A. (2022). What neural networks know about linguistic complexity, *Russian Journal of Linguistics*, 26 (2), 371–390. <https://doi.org/10.22363/2687-0088-30178> (In English)

Sherman, L. A. (1893). *Analytics of Literature, a manual for the objective study of English prose and poetry*, Ginn & Company, Boston, MA. (In English)

Solnyshkina, M. I. and Kisel'nikov, A. S. (2015). Slozhnost' teksta: Ehtapy izucheniya v otechestvennom prikladnom yazykoznanii [Text complexity: Stages of study in domestic applied linguistics], *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya*, 6, 86–99. (In Russian)

Solnyshkina, M. I., Solovyev, V. D., Gafiyatova, E. V. and Martynova, E. V. (2022). Text complexity as interdisciplinary problem, *Voprosy Kognitivnoy Lingvistiki*, 1, 18-39. (In Russian)

Solnyshkina, M. I., Harkova, E. V. and Kazachkova, M. B. (2020). The structure of cross-linguistic differences: Meaning and context of 'readability' and its Russian equivalent 'chitabelnost', *Journal of Language and*

*Education*, 6 (1), 103-119. <http://doi.org/10.17323/jle.2020.7176> (In English)

Wenger, M. J. and Payne, D. G. (1996). Comprehension and Retention of Nonlinear Text: Considerations of Working Memory and Material-Appropriate Processing, *The American Journal of Psychology*, 109 (1), 93–130. <https://doi.org/10.2307/1422929> (In English)

Wolfe, M. B. W., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W. and Landauer, T. K. (1998). Learning from text: Matching readers and texts by latent semantic analysis, *Discourse Processes*, 25, 309-336. (In English)

**Все авторы прочитали и одобрили окончательный вариант рукописи.**

**All authors have read and approved the final manuscript.**

**Конфликты интересов: у авторов нет конфликтов интересов для декларации.**

**Conflicts of interests: the authors have no conflicts of interest to declare.**

**Valery D. Solovyev**, Doc. Sci. (Physics and Mathematics), Professor, Chief Researcher, Text Analytics Research Laboratory, Institute of Philology and Intercultural Communication, Kazan (Volga Region) Federal University, Kazan, Russia.

**Валерий Дмитриевич Соловьев**, доктор физико-математических наук, профессор, главный научный сотрудник НИЛ «Текстовая аналитика», Институт филологии и межкультурной коммуникации, Казанский (Приволжский) федеральный университет, г. Казань, Россия.

**Mihai Dascalu**, Ph.D. (CS), Ph.D. (Edu), Professor, Dr., Department of Computers, Polytechnic University of Bucharest, Bucharest, Romania.

**Михай Даскалу**, доктор наук (Информационные технологии, Образование), профессор, профессор кафедры вычислительной техники, Бухарестский политехнический университет, Бухарест, Румыния.

**Marina I. Solnyshkina**, Doctor of Philology, Head and Chief Researcher, Text Analytics Research Laboratory, Professor of the Department of Theory and Practice of Teaching Foreign Languages, Institute of Philology and Intercultural Communication, Kazan Federal University, Kazan, Russia.

**Марина Ивановна Солнышкина**, доктор

филологических наук, профессор, профессор кафедры теории и практики преподавания иностранных языков, руководитель и главный научный сотрудник, НИЛ «Текстовая аналитика», Институт филологии и межкультурной коммуникации, Казанский (Приволжский) федеральный университет, Казань, Россия.