

ISSN 2313-8912

# НАУЧНЫЙ РЕЗУЛЬТАТ

ВОПРОСЫ ТЕОРЕТИЧЕСКОЙ И ПРИКЛАДНОЙ ЛИНГВИСТИКИ

RESEARCH RESULT. THEORETICAL AND APPLIED LINGUISTICS

**12(2) 2026**

**16+**

Сайт журнала:  
*rilinguistics.ru*

сетевой научный рецензируемый журнал  
online scholarly peer-reviewed journal



# НАУЧНЫЙ РЕЗУЛЬТАТ. ВОПРОСЫ ТЕОРЕТИЧЕСКОЙ И ПРИКЛАДНОЙ ЛИНГВИСТИКИ

## RESEARCH RESULT. THEORETICAL AND APPLIED LINGUISTICS

Журнал зарегистрирован в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций  
(Роскомнадзор) Свидетельство о регистрации средства массовой информации Эл. № ФС77-69052 от 13 марта 2017 г.  
The journal has been registered at the Federal service for supervision of communications information technology and mass media (Roskomnadzor)  
Mass media registration certificate El. № FS 77-69052 of March 13, 2017



Том 12, № 2. 2026  
СЕТЕВОЙ НАУЧНО-ПРАКТИЧЕСКИЙ  
ЖУРНАЛ  
Издается с 2014 г.  
ISSN 2313-8912



Volume 12 (2). 2026  
ONLINE SCHOLARLY PEER-REVIEWED  
JOURNAL  
First published online: 2014  
ISSN 2313-8912

### РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

ГЛАВНЫЙ РЕДАКТОР: **Дехнич О. В.**, кандидат филологических наук, доцент, ФГАОУ ВО «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ»), г. Белгород, Россия  
ОТВЕТСТВЕННЫЙ СЕКРЕТАРЬ: **Ли К. К.**, кандидат филологических наук, ФГАОУ ВО «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ»), г. Белгород, Россия  
РЕДАКТОР АНГЛИЙСКИХ ТЕКСТОВ: **Ляшенко И. В.**, кандидат филологических наук, доцент, ФГАОУ ВО «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ»), г. Белгород, Россия

### ЧЛЕНЫ РЕДАКЦИОННОЙ КОЛЛЕГИИ:

**Алефиренко Н. Ф.**, доктор филологических наук, профессор, ФГАОУ ВО «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ»), г. Белгород, Россия

**Алпатов В. М.**, доктор филологических наук, академик РАН, Институт языкознания РАН (ИЯЗ РАН), Москва, Россия

**Аматов А. М.**, доктор филологических наук, профессор, ФГАОУ ВО «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ»), г. Белгород, Россия

**Багана Ж.**, доктор филологических наук, профессор, ФГАОУ ВО «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ»), г. Белгород, Россия

**Вальтер Г.**, доктор филологических наук, профессор, университет им. Эрнста Морица Арндта г. Грайфсвальда, Философский факультет, Институт славистики, Германия

**Вишнякова О. Д.**, доктор филологических наук, профессор, МГУ имени М.В. Ломоносова, г. Москва, Россия

**Воркачев С. Г.**, доктор филологических наук, профессор, ФГБОУ ВПО «Кубанский государственный технологический университет», г. Краснодар, Россия

**Златев Дж.**, доктор философии, профессор, Лундский университет, г. Лунд, Швеция

**Красик В. И.**, доктор филологических наук, профессор, ФГАОУ ВО «Волгоградский государственный университет», г. Волгоград, Россия

**Клепикова Т. А.**, доктор филологических наук, профессор, Санкт-Петербургский федеральный государственный бюджетный университет экономики и финансов, г. Санкт-Петербург, Россия

**Магировская О. В.**, доктор филологических наук, профессор, Институт филологии и языковой коммуникации Сибирского федерального университета, г. Красноярск, Россия

**Огнева Е. А.**, доктор филологических наук, профессор, ФГАОУ ВО «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ»), г. Белгород, Россия

**Панасенко Н. И.**, доктор филологических наук, профессор, Университет Св. Кирилла и Мефодия в Трнаве, г. Трнава, Словакия

**Риё Ж.**, доктор филологических наук, профессор, университет Ниццы-Сопии Антиполис, г. Ницца, Франция

**Соколова Я.**, кандидат филологических наук, профессор, Университет Константина Философа, г. Нитра, Словакия

**Хлебда В.**, доктор филологических наук, профессор, Опольский университет, г. Ополье, Польша

**Чекулай И. В.**, доктор филологических наук, профессор, ФГАОУ ВО «Белгородский государственный национальный исследовательский университет» (НИУ «БелГУ»), г. Белгород, Россия

### EDITORIAL TEAM:

EDITOR-IN-CHIEF: **Olga V. Dekhnich**, Ph.D. in Philology, Federal State Autonomous Educational Institution of Higher Education "Belgorod State National Research University", Associate Professor, Belgorod, Russia

EXECUTIVE SECRETARY: **Kira K. Lee**, Ph.D. in Philology, Federal State Autonomous Educational Institution of Higher Education "Belgorod State National Research University", Belgorod, Russia

ENGLISH TEXT EDITOR: **Igor V. Lyashenko**, Ph.D. in Philology, Federal State Autonomous Educational Institution of Higher Education "Belgorod State National Research University", Associate Professor, Belgorod, Russia

### EDITORIAL BOARD:

**Nikolai F. Alefirenko**, Doctor of Philology, Federal State Autonomous Educational Institution of Higher Education "Belgorod State National Research University", Professor, Belgorod, Russia

**Vladimir M. Alpatov**, Doctor of Philology, Full member of the Russian Academy of Sciences, Institute of Linguistics, Russian Academy of Sciences, Moscow, Russia

**Alexander M. Amатов**, Doctor of Philology, Federal State Autonomous Educational Institution of Higher Education "Belgorod State National Research University", Professor, Belgorod, Russia

**Jerome Baghana**, Doctor of Philology, Federal State Autonomous Educational Institution of Higher Education "Belgorod State National Research University", Professor, Belgorod, Russia

**Igor V. Chekulai**, Doctor of Philology, Professor, Federal State Autonomous Educational Institution of Higher Education "Belgorod State National Research University", Professor, Belgorod, Russia, Belgorod, Russia

**Wojciech Chlebda**, Doctor of Philology, Professor, University of Opole, Opole, Poland

**Vladimir I. Karasik**, Doctor of Philology, Professor, Federal State Autonomous Educational Institution of Higher Education "Volograd State National Research University", Volgograd, Russia

**Tatiana A. Klepikova**, Doctor of Philology, professor, Saint-Petersburg State University of Economics, Saint Petersburg, Russia

**Oksana V. Magirovskaya**, Doctor of Philology, Professor, School of Philology and Language communication of the Siberian Federal University, Krasnoyarsk, Russia

**Elena A. Oгнева**, Doctor of Philology, Federal State Autonomous Educational Institution of Higher Education "Belgorod State National Research University", Professor, Belgorod, Russia

**Nataliya I. Panasenko**, Doctor of Philology, Professor, University of St. Cyril and Methodius in Trnava, Trnava, Slovak Republic

**Josiane Rieu**, Doctor of Philology, Professor, Nice Sophia Antipolis University, Nice, France

**Jana Sokolova**, Ph.D. in Philology, Professor, Constantine the Philosopher University in Nitra, Nitra, Slovak Republic

**Olga D. Vishnyakova**, Doctor of Philology, Professor, Federal State Budget Educational Institution of Higher Education M.V. Lomonosov Moscow State University, Moscow, Russia

**Sergey G. Vorkachev**, Doctor of Philology, Professor, Kuban State Technological University, Krasnodar, Russia

**Harry Walter**, Doctor of Philology, Professor, University of Greifswald, Greifswald, Germany

**Jordan Zlatev**, Doctor of Philosophy, Professor of General Linguistics, Lund University, Lund, Sweden

Учредитель: Федеральное государственное автономное образовательное учреждение высшего образования «Белгородский государственный национальный исследовательский университет»

Издатель: НИУ «БелГУ». Адрес издателя: 308015 г. Белгород, ул. Победы, 85.  
Журнал выходит 4 раза в год

Founder: Federal state autonomous educational establishment of higher education «Belgorod State National Research University»

Publisher: Belgorod State National Research University  
Address of publisher: 85 Pobeda St., Belgorod, 308015, Russia  
Publication frequency: 4/year

**СОДЕРЖАНИЕ**




**CONTENTS**

<b>РАЗДЕЛ I. СОПОСТАВИТЕЛЬНОЕ ЯЗЫКОЗНАНИЕ</b>	<b>SECTION I. COMPARATIVE LINGUISTICS</b>
<b>Al-Harashsheh A., Al-Husban S., Almahasees Z.</b> Human, ChatGPT and DeepSeek Translation of Arabic Taboo Expressions in the Novel “Woman at Point Zero” into English <b>3</b>	<b>Al-Harashsheh A., Al-Husban S., Almahasees Z.</b> Human, ChatGPT and DeepSeek Translation of Arabic Taboo Expressions in the Novel “Woman at Point Zero” into English <b>3</b>
<b>Баркович А. А., Ван Ц.</b> Квалитативная и квантитативная специфика корпусной репрезентации частеречной категориальности русского и китайского языков <b>25</b>	<b>Barkovich Aliksandr A., Wang Q.</b> Qualitative and quantitative specificity of corpus representation of part-of-speech categoriality in Russian and Chinese <b>25</b>
<b>Ковальчук А. В.</b> Выявление смыслового ядра как метод преодоления текстойдности <b>61</b>	<b>Kovalchuk Aleksandr V.</b> Semantic core identification as a method to overcome textoidness <b>61</b>
<b>Моисеева И. Ю., Релишский А. И.</b> Потенциал и ограничения ChatGPT: анализ качества перевода и переводческих ошибок больших языковых моделей <b>84</b>	<b>Moiseyeva Irina Yu., Relishsky Aleksandr I.</b> Assessing promise and limitations of ChatGPT: analysis of translation quality and translation errors of large language models <b>84</b>
<b>РАЗДЕЛ II. ПРИКЛАДНАЯ ЛИНГВИСТИКА</b>	<b>SECTION II. APPLIED LINGUISTICS</b>
<b>Калинин О. И., Солопова О. А., Солопов А. Д.</b> Metaphor Analytics: нейросетевой подход к автоматизированному выявлению метафорического речевого воздействия <b>111</b>	<b>Kalinin Oleg I., Solopova Olga A., Solopov Alexander D.</b> Metaphor Analytics: A neural network approach to automated identification of metaphorical speech impact <b>111</b>
<b>Краснощекова С. В., Галкина Е. В., Ван И.</b> Распределение эвиденциальных значений в заданных нарративах на родном и неродном русском языке <b>139</b>	<b>Krasnoshchekova Sofia V., Galkina Elena V., Wang Yiling</b> Distribution of Evidential Meanings in Elicited Narratives in Native and Non-Native Russian <b>139</b>

РАЗДЕЛ I. СОПОСТАВИТЕЛЬНОЕ ЯЗЫКОЗНАНИЕ  
SECTION I. COMPARATIVE LINGUISTICS

UDC 81

DOI: 10.18413/2313-8912-2026-12-2-0-1

Ahmad Al-Harashseh<sup>1</sup>  
Sujoon Al-Husban<sup>2</sup>  
Zakaryia Almahasees<sup>3</sup>

Human, ChatGPT and DeepSeek Translation of Arabic Taboo Expressions in the Novel “Woman at Point Zero” into English

<sup>1</sup> Yarmouk University, Irbid, Jordan  
E-mail: [harashseh77@yu.edu.jo](mailto:harashseh77@yu.edu.jo)  
ORCID: 0000-0001-9381-641X

<sup>2</sup> Yarmouk University, Irbid, Jordan  
E-mail: [sujoonmoneed12@icloud.com](mailto:sujoonmoneed12@icloud.com)  
ORCID: 0009-0003-0731-1821

<sup>3</sup> Yarmouk University, Irbid, Jordan  
E-mail: [z.almahasees@yu.edu.jo](mailto:z.almahasees@yu.edu.jo)  
ORCID: 0000-0002-4035-7165

*Received 17 December 2026; accepted 15 June 2026; published 30 June 2026*

**Abstract:** The present study compares human translation of Arabic taboo expressions found in *Woman at Point Zero* with translations of the same expressions generated by ChatGPT and DeepSeek. The dataset included 20 taboo expressions that were manually identified in the text of the novel. First, a quantitative analysis was carried out to evaluate the quality of machine translations of the expressions in question according to the metrics of the Multidimensional Quality Metrics (MQM) framework. The analysis encompassed such dimensions as accuracy, fluency, terminology, and style. In addition, eight selected examples that represent the major categories of taboo language, as well as the most common difficulties in translation of the latter, were examined qualitatively. As a result, the present study provides one of the very few systematic assessments of the ability of generative AI to translate taboo expressions from Arabic literature. Human translation turns out to be superior to machine translation in terms of maintaining semantic accuracy, cultural significance, and pragmatics of taboo language. While ChatGPT and DeepSeek produce fluently written and terminologically accurate translations, they tend to neutralize, weaken, or misinterpret culturally encoded messages, thus reducing the stylistic and emotional effect of the source text. It appears that contemporary generative AI technologies are unable to cope with the translation of literary language that is culturally sensitive and requires interpretation in a particular context. One of the limitations of the study is its qualitative analysis based on a number of selected examples taken from one particular literary text. Therefore, the findings cannot be generalized to any other literary genre or language pair. However, the research yields original results for the rapidly growing field of AI-assisted literary translation, as it involves comparative MQM assessment of human and large language

model translations of Arabic taboo expressions and emphasizes the role of human post-editing in literary translation.

**Keywords:** Human Translation, ChatGPT, DeepSeek, MQM, Cultural Translation, Arabic-English Literary Translation, Taboo Translation, El-Saadawi

**How to cite:** Al-Harashseh, A., Al-Husban, S., Almahasees, Z. (2026). Human, ChatGPT and DeepSeek Translation of Arabic Taboo Expressions in the Novel “Woman at Point Zero” into English, *Research Result. Theoretical and Applied Linguistics*, 12 (2), 3–24. DOI: 10.18413/2313-8912- 2026-12-2-0-1

## 1. Introduction

Taboo expressions are among the most sensitive areas of language because they convey meanings beyond those associated with the words themselves. Taboo language includes terms or phrases that are perceived as offensive, rude, or socially inappropriate by members of a certain cultural group (Allan and Burridge, 2006; Jay, 1992). Cultural norms are not the same for all societies; thus, an expression that is taboo in one culture could be acceptable in another culture. This means that the translation of taboos goes beyond the simple task of finding equivalent terms.

In translating taboo expressions, the difficulty arises in the fact that the taboo expressions serve different functions in communication. Apart from conveying an expression of annoyance, a taboo expression could also be used to express anger, intimacy, humor, social identity, power struggle, or even opposition to prevailing ideologies (Jay, 2009). Taboo expressions in literature are normally used by writers with the intention of constructing their characters, bringing out social injustices, contesting against prevailing ideologies, or simply projecting their stylistic voice. The translator, thus, has to deal with the challenge of balancing between semantic accuracy and cultural appropriateness of the translated expression.

Previous studies have categorized the elements of taboo speech into various types such as profanity, obscenity, vulgarity, epithets, insults and slurs, scatology, and cursing (Jay, 1992). These categories do not denote separate groups of vocabulary but rather different levels of cultural and emotional sensitivity which call for diverse

translation strategies. Therefore, the translation of taboo speech is not an easy task that only requires linguistic skills, but intercultural sensitivity as well (Allan and Burridge, 2006).

Recent developments in AI technology have revolutionized translation by way of developing advanced language models like ChatGPT and DeepSeek. Both have displayed excellent results in translation tasks for various languages. However, literary translation poses one of the most difficult challenges to AI because literary content often relies heavily on implicit cultural connotations and emotions. This problem becomes even more pressing in the case of taboo expressions, which involve making ideological choices about how language is used within specific cultural contexts. While many scholars claim that recent developments in generative AI show promising results in ordinary translation tasks, there is still insufficient research on its translation of literary content, especially taboo expressions in Arabic literature.

Additionally, previous work has been mainly concerned with measuring the quality of translation or comparison between AI and human translators in terms of their performance in translating technical and non-literary texts. Relatively little research has been conducted regarding taboo translation in literature, and even less has used a systematic approach for assessing the quality of translation to compare human translators with AI. It is especially important to take into account the need for context-based decisions in literary translation, which brings up a question about how well the current

generation of AI is capable of maintaining the role of taboos in language.

In order to fill the gap mentioned above, the current research seeks to compare the English translation of Nawal El Saadawi's *Woman at Point Zero* by Sherif Hetata (1983) with the translations produced by the AI translators ChatGPT and DeepSeek. A total of 50 taboo terms have been collected manually from the Arabic novel for analysis. To assess the quality of the translations, the MQM criteria will be employed, which include four main dimensions – namely, accuracy, fluency, terminology, and style. Moreover, some chosen examples will be analyzed in depth to see how translation strategies used by the human translator and the AI models impact the transmission of cultural meaning and stylistic elements.

This paper adds to the emerging body of research on literary translation through AI in several key ways. It is one of the first studies that compares human and generative AI translations of taboo terms in Arabic based on the MQM system. The paper addresses the issue of the current big language models' handling of the conflict between adaptation and faithfulness in literary translation. Finally, it offers valuable information on both the potentials and constraints of AI-assisted literary translation.

Therefore, the present study attempts to answer the following research questions:

1. What is the nature of the differences between the human translator's translation, ChatGPT's translation, and DeepSeek's translation of the taboo terms used in *Woman at Point Zero* in terms of accuracy, fluency, terminology, and style using the MQM approach?

2. What are the translation techniques adopted by the human translator, ChatGPT, and DeepSeek for translating taboo terms from Arabic into English and what are their implications on meaning transfer?

## **2. Literature Review**

### **2.1 Theoretical Background**

Taboo expressions are one of the most culturally delicate aspects of translation since

they acquire their meaning not only based on their vocabulary but also on social conventions, religious beliefs, gender roles, ideologies, and contexts. Taboo expressions are words or phrases that are considered offensive, rude, or socially inappropriate in a particular society (Allan and Burrige, 2006; Jay, 2009). Yet, taboo language does not have an intrinsically offensive nature. Its offensiveness is dependent on who says it, whom it is directed at, the context in which it is used, and its pragmatic role. In the field of literature, taboo expressions may have several purposes; they may be used to describe a character, express rage, unveil violence, criticize social hypocrisy, or voice marginalized voices.

The translation of taboos in literary texts is especially difficult since they tend to have symbolic meaning as well. It is possible to use literal translation, euphemisms, omission, substitution, or even culturalization when translating such words, but all of these ways influence the way in which the target audience will perceive the text (Zagood et al., 2022). While literal translation may keep the shock effect from the source text, it will be unnatural and too shocking for the target audience. Euphemisms may make the translation less shocking, but they also make the emotion weaker, and the ideology of the text is lost. Moreover, omission may cause the loss of essential meaning related to characterization, realistic depiction, and social criticism (Lestari and Sutrisno, 2023).

The problem is even more acute in the case of feminist literary translations. Feminist translation theory understands translation as an ideological practice rather than a simple linguistic transfer (Simon, 1996; Castro and Ergun, 2017). In a feminist text, the use of taboo terms could serve as a means to unveil gender oppression, exploitation of the body, violence, and control by patriarchy. The novel by Nawal El-Saadawi, *Woman at Point Zero*, can be used as an illustration in this regard since the body, sexuality, violence, and women's resistance become key elements in a patriarchal society of Egypt. Taboo terms are

integral parts of the feminist discourse in the novel, and their softening or neutralization will result in the loss of its radicalism.

Pragmatic equivalence is the key to this debate. Baker (2018) suggests that when translating from one language into another, a translator has to take into consideration the role of context, implicature, and communicative purpose. Likewise, House (2015) believes that translation quality has to be judged based on how well the translated text accomplishes the same communicative task as the source text. As far as taboo words are concerned, pragmatic equivalence implies not just conveying the lexical meaning of words but also the desired effect that might be shock, humiliation, anger, resistance, or any other emotion.

Recently, machine translation and artificial intelligence developments have added yet another aspect to this debate. Machine learning models like ChatGPT and DeepSeek are capable of delivering fluent and consistent translations. However, literary translation is quite a complicated task since it relies on implied meanings, styles, cultural memories, and ideological positioning. Scientific evidence recently gathered concerning machine literary translation demonstrates that post-editing of creative translations must balance efficiency and style/creativity preservation. For instance, a recent 2024 paper written by Macken explores the potential of ChatGPT in post-editing literary machine translations, while Castaldo et al. (2025) analyze LLM-generated literary translations along with their post-editing by professional translators.

## 2.2 Empirical Studies

From recent research on artificial intelligence-assisted translations, it is evident that large language models have demonstrated high levels of effectiveness in general translation, although their efficiency depends on the text types, language pairs, domains, and evaluation measures used. According to Zhao et al. (2023), ChatGPT was assessed using human assessment of its translation performance in various domains and was

found to perform effectively in general translation, while it lacked domain-specific terminology and style in some cases. This applies to the current study due to the context-dependence of taboo terms in literary language.

Similarly, Hendy et al. (2023) analyzed the use of GPT in machine translation and found that the models can compete with neural machine translation. Nevertheless, there are studies comparing human translation, neural machine translation, and ChatGPT translation results, which show that while being good in general translations, LLMs do not have to be good at translating voice, ideology, or culturally meaningful nuances. This is important because the book under consideration implies the translation of taboo expressions used in the context of feminist theory, social criticism, and psychological trauma.

Research in machine translation of literature has paid attention to the limitations of machine translations. Guerberof-Arenas and Toral (2022) concluded that machine-translated literary texts require post-editing that might limit the creativity of the translators and affect the effectiveness of their translations negatively. The following works have kept addressing the role of AI for literary translators instead of taking their jobs. Macken et al. (2022) has discussed literary translation as a process consisting of machine translation, post-editing, and revision.

Recent research has been centered around LLMs and literary post-editing. Macken (2024) analyzed the possibility of automatic literary post-editing by ChatGPT and concluded that LLMs could enhance the qualities of machine translations, although professionals would be required to ensure literary translation. Additionally, Castaldo et al. (2025) revealed that the use of LLMs for creating literary translations could save the time spent on editing while preserving a certain degree of creativity; however, the research is another proof that the evaluation and intervention of humans are required when it comes to creative translation.

Previous research in the area of taboo translation tends to concentrate on audiovisual translation, such as film subtitling and dubbing, as well as Netflix series and TV drama translation. It should be noted that such research is important and informative since it helps understand how tabooed language tends to get mitigated, censored, and adapted for target audiences. However, the results of such research cannot be applied directly to literary translation since, in contrast to audiovisual translation, it is not limited by time, space, and multimodality but rather depends on narrative voice and character description. That is why the current research is going to concentrate on Arabic feminist literary discourse.

Scholars who have studied the writings of Nawal El-Saadawi have largely centered their discussions around feminism, patriarchy, sexual violence, and the resistance of women. Analysis of *Woman at Point Zero* usually concentrates on the opposition of Firdaus to oppression from the institutions, sexually, and socially (Amireh, 2000; Balaa, 2018). Yet there has been insufficient research on how the taboo expressions in the book are translated into English, and how AI machines translate them. This is because the taboos in El-Saadawi's writings are very related to the concepts of voice, body, oppression, and resistance.

Al-Yasin and Rabab'ah (2019) investigate how taboo language is rendered in American *Hip Hop* films that are subtitled into Arabic, with special interest in fan-produced subtitles. The study reviews 512 occurrences of English swearwords from three profanity-laden films, where official English subtitles are compared with those of Arabic fan-made ones. Based on Ljung's (2011) swearing typology, the researchers classify the English swearwords according to function and intensity, and examine the way in which these expressions are translated into Arabic. Four major strategies used in fan subtitles are determined by the study: Literal translation of taboo expressions, cultural substitution (i.e., taboo replaced by a

culturally acceptable equivalent), omission or deletion of taboo terms, and, lastly, euphemistic rendering (i.e., expressions minimizing the shock in the target language). The findings indicate that rather than using additions, the Arabic fan subtitlers preferred either substitution or omission as they were sensitive to the norms of the target audience. The study also highlights the emergence of tension between the need to retain the expressive strength of the source text and the need to be culturally conformist in the Arab context (Al-Yasin and Rabab'ah, 2019).

Abedaal and Al Sarhani (2021) investigate the translation of English swear words and taboo expressions into the Arabic subtitles. Using a mixed-method approach, they used a qualitative approach to identify translational strategies and quantitative methods to assess subtitle quality. The study comprises 40 cases of swear words and taboo expressions in the 2001 American crime drama "*Training Day*". The analysis exploited Pedersen's (2005, 2011) paradigms of a subtitling strategy type and a quality assessment model with the tweaks made by *Abdelaal* and Sarhani (2021). It was established that the most recurrent way used was the usage of omission and euphemistic translation, which was caused by cultural sensitivities as well as the conservatism of the Arab audiences. Other issues were: holding on to language, specifying, generalizing, and formalizing the informal language. Results of the quality assessment, which was designed taking into account the FAR model (Functional equivalence, Acceptability, Readability), showed that the majority of subtitles were at a good level of quality with minor incongruities that did not distort the impression about the viewer's comprehension. The research finds that although omission and euphemizing are effective devices and are culturally constrained, they are still necessary to strike a balance of sensitive subtitling and retention of the expressive content of the source material. The authors suggest reworking the existing models of the quality assessments to

incorporate the quantitative and qualitative criteria for a more integral analysis (Abdelaal and Sarhani, 2021).

Haider et al. (2023) examine how the text editors translate taboo expressions from the Jordanian Arabic Series *Jinn* to English. This series was selected because of its importance as Netflix's first Arabic-Jordanian production of explicit content and culturally sensitive language. The cultural taboo expressions were translated into English using four main strategies: Transforming a source culture taboo into a target culture taboo of the same or greater or smaller strength, omitting the source taboo expression, replacing it with a neutral or less offensive one, and using euphemism or circumlocution. The results demonstrate that the majority of taboo expressions were borrowed using substitution and euphemism to fit the target culture norms. However, in many instances, the original level of offensiveness was achieved or increased. Moreover, the study indicates that the most commonly spoken implied functions of taboo language were 'unfriendly suggestions' and 'noun supports'. It recommends future research on cross-cultural taboo-language transfer within subtitled and dubbed media (Haider et al., 2023).

Olimat et al. (2023) analyze the subtitling of sexually sensitive Arabic utterances for English-speaking viewers in two Netflix series: *Jinn* and *Al-Rawabi School for Girls*. The study examines 284 cases of sexually sensitive or culturally taboo language found in the original Arabic dialogue, in terms of how they are translated in the subtitles placed by Netflix for the English version. The researchers identify five major strategies utilized in translation process: Direct translation of the taboo lexeme from taboos to taboo lexeme direct standards, substitution of culture in this case, replacing the Arabic phrases that offend with the more contextually appropriate adventures of English, generalization that, a reduction in specifics and impact, omission which removes the sensitive term from the text and

euphemistic rendering to palliate that disruptiveness of the original text.

The outcomes demonstrate that even though direct translation and cultural substitution were the most widely employed and effective strategies, the excessive use of generalization and omission often resulted in diluted meanings and a diminished effectiveness on the target audience. Furthermore, the study covers the other data set from the American Sitcom *How I Met Your Mother*, and here the authors describe the subtitling for the comedic and the sexually suggestive dialogue. It is observed that the timeliness of conveying explanatory subtitles during short micro dialogic sections is dramatically compressed in its comedic feel and the delivery of comedy. In conclusion, the results indicate that translating culturally sensitive content, such as sexual or religious themes, is a complex task. The translator must accurately convey the intended meaning of the original text while simultaneously showing sensitivity and respect toward the target culture. Striking this balance presents a significant challenge, especially when there are notable cultural differences between the source and target audiences (Olimat et al., 2023).

Badmus et al. (2024) investigate taboo expressions in the Idoma language pragmatically, and they focus on the courtesy and indirect nature of such terms. The research comprises 215 examples of taboo expressions distributed with reference to their social, sexual, and cultural relevance. The survey views those expressions as words that are regularly used by adults to teach their young children the virtues of moral behavior, but are avoided in a public setting, for social reasons. The authors use a typology of taboo language, distinguishing between those used in private and those in the public domain, and dealing with the indirect or euphemistic ways in which they are mellowed. The same strategies that are used by the speakers are also categorized according to the study: use of taboo terms directly, omission of taboo terms, circumlocutions or euphemisms, and

culturally specific terms that describe community perception. The findings suggest that taboo language should remain for private conversation, and critical documentation of such expressions is very important for the preservation of sources of the Idoma language. The research also calls for greater analysis on the cultural and Ideological character of taboo language in the teaching of the youth on proper use and avoidance of such jargon (Badmus et al., 2024).

Shormani (2024) analyses the accuracy of translating Arabic oath phrases into English by ChatGPT and whether ChatGPT can capture swearing nuances or not. The research analyses 30 Arabic oaths drawn from literary sources, namely, it compares the transcripts of the AI and the human translations. The research states that there are really many weaknesses in how ChatGPT represents the subtlety of Arabic oaths, namely the inability to understand the religious figure specific to culture, the non-oath aspect of the Arabic language. Translation strategies assessed include direct translation of oaths, contextual alignment to fit in issues concerning the target language, removal of culture-specific elements or expressions, as well as literal translation of non-oaths' aspects. The finding points to the fact that the translations of the Arabic oaths made using ChatGPT never guarantee the desired cultural as well as emotional impact of the Arabic oaths, which requires further improvement of the AI-based translation systems. The study concludes with proposals towards obtaining more data for subsequent training on the Arabic oath expressions and their cultural contextualization to build a better-performing system (Shormani, 2024).

Brahimaj (2024) investigates the sociolinguistic nature of taboo language in Kosovar society. The research involves the usage of a combination of quantitative and qualitative methodologies (mixed-methods approach), meaning that the given study permits having an overall idea about the usage and understanding of expressions that can be attributed to the taboo class. Data collection

was made by a structured survey based on ten questions, in which patterns of usage and frequency of usage, social acceptability, and perceived offensiveness of taboo words were sought. The study was based on a sample consisting of 101 subjects of both sexes whose ages ranged from 17 to 60 years. Other than personal use of taboo language, the study acknowledged its relevance in context as well as the external influences that make an individual prone to uttering such language, like media exposure. Notwithstanding, no specific theoretical framework is referred to by name, yet the research is carried out in the context of the sociolinguistic approach to language use and identification of the way taboo language works on the informational level in social relations and how cultural and societal norms form a part of its use. It also tries its hand at language evolution in ways that implicitly concur with theories that deal with the dynamics and adaptiveness of language in society.

Al-Harashsheh and Shehab (2024) assess the ideological, cultural, and social implications related to the epithet parts of El-Saadawi's *Woman at Point Zero*'s English cover pages. The study employed Pierce's model of sign process. The study found that these epigraphic elements reflect El-Saadawi's literary style in writing and the deep social, ideological, and cultural implications that she struggles to convey about the Egyptian society's view toward women. In addition, it found that these epigraphs affect the readers' perception and interpretation of the novel's core message, theme, style, and its cultural context.

At broader conclusions, the study reports that the taboo language is common in the Kosovar society and is dynamically influenced by factors such as age in society, social environment, mood, media, among others. The use of taboo expressions is accompanied by anger or pain, and it depends on the context or the audience. The study is centered around the dynamism of language and the possibility of some change in the norms related to the changing opinions on the

taboo expressions. According to the results, it is necessary to continue considering the connection between the utilization of taboo language and the general cultural and social transformations (Brahimaj, 2024).

Fitriansyah (2023) investigates the taboo in the *Gran Torino Movie*, which looks at the translation of such words in the film. The study assessed the subtitling of 270 taboo words that were subtitled into Arabic. Experts were brought together in FGDs to determine the quality of translations. The research divides taboo words into ten groups, and among them are sexual acts, profanity, insults about race, and scatological references. They served four main functions: They involve the option of connecting, disconnecting, or distinguishing words or speakers; in addition, they were called auxiliary, abusive, expletive, or humorous by some. Most taboo words went through changes in meaning and role when moved from one language to another. More often than not, the approach taken for translation involved reduction, with established equivalent, exploitation, discursive creation, and literal translation being the next most popular. Still, reduction was the primary method used, but this lowered the quality of translations. From the assessment of the authors, Nababan et al. (2012), the scores were 1.78 for accuracy, 2.12 for acceptability, and 1.92 for the total quality, suggesting a poor overall result. It summarizes that avoiding taboo vocabulary is okay, but using substitutions is more helpful for preserving the meaning and intensity (Fitriansyah, 2023).

The reviewed literature, therefore, reveals three major gaps. First, recent studies on taboo translation remain heavily oriented toward audiovisual translation, with limited attention to literary texts. Second, research on El-Saadawi has emphasized feminist themes but has rarely examined taboo expressions from a translation-quality perspective. Third, although recent studies on ChatGPT, LLMs, machine translation, and post-editing are rapidly increasing, few have examined how generative AI translates culturally sensitive

taboo expressions in Arabic feminist literature. The present study addresses these gaps by comparing Sherif Hetata's human translation of *Woman at Point Zero* with translations generated by ChatGPT and DeepSeek, using the MQM framework to evaluate accuracy, fluency, terminology, and style, while also considering pragmatic equivalence and feminist literary context.

### 3. Methods and Procedures

#### 3.1. Data collection

The dataset for the current study was compiled out of 50 Arabic taboos manually collected from the original Arabic novel titled "*Woman at Point Zero*" by Nawal El Saadawi. These taboos were then paired with the official English translation of the book provided by Sherif Hetata (1983). The same taboos were translated via ChatGPT and DeepSeek with the same set of prompts used for all three translations in order to ensure consistent comparison. The 50 collected taboos belong to various categories, including sexual, gender, violent, body, religion-related, and other culturally sensitive taboos relevant to the feminist discourse of the novel.

All 20 expressions (See Appendix A) that were extracted for analysis were used for the quantitative study through the MQM approach. With this in-depth analysis, the researchers were able to analyze and compare the quality of the translation of the three translation results and find out the most prevalent patterns and errors in translation. The overall MQM scores from the whole corpus are shown in Figure 1.

For the qualitative analysis, eight examples were chosen through purposive sampling. This was done based on pre-defined criteria of methodology rather than convenience. First, the chosen examples are representative of the different categories of taboos, namely sexual, gendered, violent, and culturally-sensitive. Second, these examples show the various methods used by the human translator, ChatGPT, and DeepSeek to translate taboo expressions, which include, but not limited to, literal translation, euphemization, omission, cultural

substitution, and semantic change. Third, the examples were selected as they have interesting linguistic, pragmatic, and cultural features that would allow discussing the issues in the translation of taboo expressions. Lastly, the chosen examples are those that had the highest frequency rates for the MQM error types found in the qualitative analysis of the full dataset. Therefore, the qualitative analysis aimed to exemplify and explain the tendencies revealed throughout the analysis of all 50 expressions rather than to substitute the results of the quantitative analysis. The use of only eight examples, however, allowed conducting a deep enough analysis without exceeding the scope of a journal article.

The choice to use Nawal El Saadawi's *Woman at Point Zero* is due to the presence of many taboo terms that are intricately related to problems of gender, sexuality, violence, religion, and oppression in society. These aspects of the book make it suitable for analyzing the ways through which human and modern AI translators translate culturally-sensitive texts in literature.

### 3.2. Data Analysis

This research utilized a mixed-method analysis approach, which is a combination of qualitative and quantitative research designs. During the initial phase of the analysis, the translations done by Sherif Hetata, ChatGPT, and DeepSeek were assessed quantitatively based on the MQM empirical metrics framework. The MQM was chosen since it offers an international model for evaluating translation quality. MQM assesses translation quality in relation to accuracy, fluency, terminology, and style; hence, it is important in evaluating translations of culturally sensitive literature.

The translation errors were assigned based on their severity levels in terms of MQM. Errors of minor severity (weight = 1) are those errors that have no influence or very little influence on the overall understanding or reading of the translated text. Errors of major severity (weight = 5) have a considerable

effect on the usability or understandability of the translation, but the meaning is maintained nonetheless. The errors of critical severity (weight = 10) are those that change the meaning of the source text significantly, thus having the most significant impact on translation quality. All of the 50 items were assigned their score for MQM in a consistent manner.

After performing quantitative analysis, the study carried out a detailed qualitative analysis of the eight selected examples. The objective of this phase was to provide a rationale for the results of the quantitative analysis by examining how each translation dealt with cultural, pragmatic, and ideological aspects of the taboo expressions in question in terms of the literary context in which they occurred.

The identification and comparison of the translation strategies used by the human translator, ChatGPT, and DeepSeek in this research have been conducted using the classification developed by Khoshsaligheh and Ameri (2014). According to this classification, taboo expression translation is divided into four main strategies, namely: (1) taboo to taboo; (2) deletion; (3) taboo to non-taboo; and (4) euphemism. In order to conduct an in-depth comparison of translation quality and strategy usage, all eight selected examples have been analyzed on the basis of two classifications – the MQM one and the strategy taxonomy.

### Findings of the study

The analysis focused on how the taboo translations were subtitled and translated in terms of fluency, accuracy, and the ability to deal with the social and cultural context of these expressions. The MQM framework was used to accurately identify the types of issues and categorize errors, to assess the quality of the translation. There was a clear contrast between AI translation and human translation, highlighting the strengths and limitations of each type of translation.

**Table 1.** A taboo expression quoted from *Woman at Zero Point* (page 19)

ST	HT	Penalties	ChatGPT	Penalties	DeepSeek	Penalties
وأنت امرأة معها مطواه أو شفرة موسى، و قطعوا قطعة من اللحم بين فخذي. (ص:19)	Then she brought a woman who was carrying a small knife or maybe a razor blade. They cut a piece of flesh from between my thighs.	Total words in the text=26  P= $1*1+0*5+0*10/26=0.038$ Q= $100-(10*0.038)=99.62$	She brought a woman carrying a pocketknife or a razor blade, and they cut a piece of flesh from between my thighs.	Total words in the text=22  P= $0*1+0*5+0*10/22=0$ Q= $100-(10*0)=100$	She brought a woman carrying a dagger or razor blade, and they cut a piece of flesh from between my thighs.	Total words =21 P= $1*0+5*1+10*0/22=0.22$ Q= $100-(10*0.22)=97.8$
	Minor=1 Major=0 Critical =0	The use of “small knife” instead of “pocketknife”	Minor=0 Major=1 Critical=0		Minor=0 Major=1 Critical=0	*(use the word “dagger” instead of “small knife” or “pocketknife”

The above expression in Table 1 is considered a vulgarity since it describes female genital mutilation through the depiction of the injury to the body. In this case, the human translator succeeded in accurately translating the taboo and maintaining the feminist interpretation of El Saadawi's message about the violation of women's rights. There are no MQM errors in this translation apart from the inaccurate translation of the word مطواه into “small knife” rather than into its actual English translation. However, this translation still conveys the full meaning and emotional intensity of the message without any MQM errors. On the contrary, ChatGPT offers a

better translation of مطواه as “pocketknife.” Thus, ChatGPT manages to convey both semantic and pragmatic equivalence with no MQM errors. Nevertheless, DeepSeek keeps the taboo reference explicit; however, مطواه is translated as “dagger,” which constitutes a terminology error since the lexical item used does not correspond to the actual type of weapon used. As for the translation strategy, human translation and ChatGPT use taboo-for-taboo transfer; they use the domestication strategy and maintain the face-threatening acts of the expression in the target language. However, DeepSeek uses the combination of literal translation and lexical substitution.

**Table 2.** A taboo expression quoted from *Woman at Zero Point* (page 36)

ST	HT	Penalties	ChatGPT	Penalties	DeepSeek	Penalties
حينما يجف الثقب أتركه يقبلني، و أحس الورم الكبير فوق شفتي ووجهي مثل	When the hole was dried up, I let him kiss me. I could feel the	Total words =34 (penalties=issue minor * severity multiplier minor +	“When the sore dries up, I let it kiss me, and I feel the large swelling	Total words =33 (penalties=issue minor * severity multiplier minor + issues	When the hole dries, I let him kiss me, and I feel the large swelling over my lip	Total words =29 (penalties=issue minor * severity multiplier minor + issues

الكيس أو القربة المائلة بالماء الرائد أو الدهن) (ص 36)	swelling on my face and lips like a small purse, or a water skin, full of a stagnant greasy fluid.	issues major * severity multiplier major)/word count) P= $1*0+0*5+0*$ $10/34=0$ Q= 100- $(10*0)=$ 100%	over my lip and face like a sack or a skin filled with stagnant water or fat.”	major * severity multiplier major)/word count) P= $1*0+5*1+10*$ $0/33=0.30$ Q= 100- $(10*0.03)=$ 99.69%	and face like a waterskin filled with stagnant water or grease.	major * severity multiplier major)/word count) P= $1*0+0*5+0*10$ $/29=0$ Q= 100- $(10*0)=100%$
	Minor=0 Major=0 Critical=0		Minor=0 Major=1 Critical=0	(grammar mistake).	Minor=0 Major=0 Critical=0	

The expression in Table 2 falls into the category of vulgarity because it denotes an injury of a physical nature and the feeling of pain and suffering, which underlines the protagonist's trauma and the feminist representation of violence done to women in the novel. Human translation manages to keep all the information provided in the source text, including the sequence of actions, imagery, and simile, without any changes, making it a fully coherent text with no MQM errors. The translator employs the domestication strategy to transfer the taboo

meaning to the target readers. ChatGPT, despite being fluent, makes a significant mistake in terms of accuracy because it uses the pronoun "it" where "him" should be used in order to make a coherent passage. DeepSeek offers the most accurate rendering of the source text, making sure that referents are correct as well as imagery and metaphors are kept with the help of such lexemes as "waterskin," "stagnant water," and "grease." It also employs the domestication strategy. There are no MQM errors in the translation.

**Table 3.** A taboo expression quoted from *Woman at Zero Point* (page 44)

ST	HT	Penalties	ChatGPT	Penalties	DeepSeek	Penalties
و أحس لعابة اللزج داخل فمي، فأطرده بلساني إلى فمه مرة أخرى (ص: 44)	I could feel his thick saliva between my lips and with a push of the tongue sent it back into his mouth.	Total words =21 P= $1*0+5*0$ $+10*0/21=$ 00.0 Q= 100- $(10*00.0) =$ 100%	I curled my lips to spit in his face, but he caught my lips with his teeth.	Total words =17 P= $1*0+5*0+1$ $0*1/$ $17=0.588$ Q= 100- $(10*0.588)$ = 94.12%	I feel his viscous saliva inside my mouth, so I thrust it back into his with my tongue.	Total words =18 P= $1*0+5*0+10*$ $0/18=00.0$ Q= 100- $(10*00.0) =$ 100%
	Minor=0 Major=0 Critical=0		Minor=0 Major=0 Critical=1	Mistranslati on	Minor=0 Major=0 Critical=0	

The above example is regarded as an instance of obscenity because it features a

clear bodily reference with the notion of physical intimacy and the protagonist's

repulsion towards the male character. Human translation successfully conveys the message of the original text with all imagery and communicative value being preserved, which leads to zero MQM mistakes. Clearly, the translator utilizes the domestication strategy. Meanwhile, ChatGPT makes many changes to the original phrase, including some events that do not appear in the source text, such as "I curled my lips to spit in his face" and "he caught my lips with his teeth". Such modifications lead to the distortion of events'

sequence and constitute a significant accuracy problem according to the MQM guidelines, leading to the lack of semantic and pragmatic equivalence. Nevertheless, DeepSeek manages to reproduce both the bodily reference and the events' sequence via phrases like "viscous saliva" and "I thrust it back into his mouth" without creating any MQM errors. Concerning the translation strategy, taboo-to-taboo transfer was used by the human translator and DeepSeek, while interpretive mistranslation occurred in ChatGPT.

**Table 4.** A taboo expression quoted from *Woman at Zero Point* (page 48)

ST	HT	Penalties	ChatGPT	Penalties	DeepSeek	Penalties
و يشف من تحته ثدياي من غير مشد ص: (48)	My breasts were showing almost naked under my dress.	Total words =9 P= 11* +5*0+10*0 =0.119/ Q= 100- (10*0.11)= 98.9%	My breasts show through it without a bra.	Total words 8= P= 1*0+5*0+10* 0/ 8=00.0 Q= 100- (10*00.0)= 100%	Beneath it, my breasts show through, unsupported.	Total words 7= P= 1*0+5*0+10* 0/ 7=00.0 Q= 100- (10*00.0)= 100%
	Minor=1 Major=0 Critical=0		Minor=0 Major=0 Critical=0		Minor=0 Major=0 Critical=0	

The above statement in example 4 is considered to be an example of offensiveness because it clearly refers to the woman's body. It further supports the feminist nature of this novel, where females appear to be vulnerable. The human translator has maintained the same message in the target language, but the phrase "من غير مشد" has been skipped, which makes the statement less offensive. Therefore, this is a minor omission error according to the MQM criteria. However, the translator transfers the implicature or the intended meaning accurately to the target readers. On the other hand, ChatGPT has kept the same

source meaning intact through the translation "My breasts show through it without a bra". Likewise, DeepSeek has maintained the source meaning through the phrase "unsupported," which, although less obscene, semantically conveys the source message. Still, AI translation maintains the implicature or the intended meaning in the target text. None of the AI translations contains any MQM errors. Talking about translation strategy, both the AI translations have used the taboo-to-taboo strategy, while human translation is based on the partial omission strategy, thus making it less pragmatic.

**Table 5.** A taboo expression quoted from *Woman at Zero Point* (page 48)

ST	HT	Penalties	ChatGPT	Penalties	DeepSeek	Penalties
حلمة الثدي دائرة سوداء بارزة, ضغط عليها بيده و هو	The nipples standing out in two dark	Total word =21 P= 1*0+5*1+ 10*0/ 21=	My nipple, a prominent dark circle, was pressed by his hand	Total word =17 P= 1*0+5*0+1 0*0/ 17=	In his dream: a dark, circular breast —	Total word =19 P= 1*0+5*0+10* 1/ 19= 0.52

يساعدني بركوب السيارة (ص: 48).	circles. As he helped me into the car. He pressed his arm against them.	0 Q= 100- (10*0)= 100%	as he helped me into the car.	00.0 Q= 100- (10*00.0)= 100%	prominent. He pressed it with his hand while helping me into the car.	Q= 100- (10*0.52)= 94.8%
	Minor=0 Major=0 Critical=0		Minor=0 Major=0 Critical=0		Minor=0 Major=0 Critical=1	

This example is considered to be obscene language as the text includes a direct reference to the female body by means of the noun nipple. Thus, the given example reveals the problem of female objectification and sexual vulnerability described in the novel. As far as the human translation is concerned, it can be said that it perfectly captures the intended meaning of the source text and maintains its pragmatic effect since there is no information loss or distortion in the translation. ChatGPT translation is also faithful, and there are no MQM errors in it since the expression "My nipple, a prominent dark circle" demonstrates lexical precision and does not involve any modification of the original text. Both human translator and ChatGPT employ a domestication strategy. However, the translation produced by

DeepSeek is flawed with one serious mistake, as the translator used the expression "In his dream: a dark, circular breast" in the place of "حلمة الثدي" thus adding some new information to the source text and distorting the meaning of the text. Such semantic distortion is caused by the fact that it changes both the propositional meaning and the pragmatic effect of the text, making it less faithful to the source text. Pragmatically speaking, it is clear that the translation fails to capture the communicative intention of the source text since "In his dream" is hallucinated, and the inaccurate "breast" instead of "nipple" distorts the picture. Concerning the translation strategy, it should be noted that in this case, the human translator and ChatGPT use taboo-to-taboo transfer while DeepSeek makes an interpretive mistranslation.

**Table 6.** A taboo expression quoted from *Woman at Zero Point* (page 18)

ST	HT	Penalties	ChatGPT	Penalties	DeepSeek	Penalties
لم أكن أمسك أي جريدة و أرى فيها صورة لأي رجل منهم حتى أبصق على وجهه (ص 18).	Each time I picked up a newspaper and found the picture of a man who was one of them, I would spit on it.	Total word =22 P= 1*0+5*0+ 10*1/ 22 =00.0 Q= 100- (10*00.0)= 100%	I couldn't hold a newspaper and see the picture of any of those men without spiting on his face.	Total word =19 P= 1*0+5*0+10* 0/ 19 = 0 Q= 100- (10*0)= 100%	I could never hold a newspaper and see any man among them boasting without spiting in his face.	Total word =11 P= 1*0+5*1+10*0/ 11 = 0.454 Q= 100- (0.454*10)= 95.46%
	Minor=0 Major=0 Critical=0		Minor=0 Major=0 Critical=0		Minor=0 Major=1 Critical=0	Added the word "boasting"

The phrase used here is an example of an epithet since it expresses disgust and animosity against a bunch of men through spitting on their pictures. The human translation captures the literal, intended, and derogatory meaning of the source text without any errors in terms of lexemes and grammar, just as ChatGPT does. Thus, there are no MQM errors in both translations. They both use a domestication strategy. However, DeepSeek creates the unneeded addition of "boasting" in its translation that is not present

in the source text and distorts the context in the way that it attributes one of the features not mentioned in the source text. Such an error is considered to be a major error of accuracy since it alters the propositional meaning but keeps the message intact. As far as the translation strategy is concerned, human translation and ChatGPT use taboo-to-taboo transfer, keeping the insult as it is, while DeepSeek uses the combination of direct transfer and addition.

**Table 7.** A taboo expression quoted from *Woman at Zero Point* (page 36)

ST	HT	Penalties	ChatGPT	Penalties	DeepSeek	Penalties
و في الليل يلف ذراعيه و ساقه حوليه، و أترك يده المعروفة تعبت بجسدي كله لا تترك شيئا، كيد جانح لا يرى الطعام من سنين، فإذا به يمسح الصحن و يلعقه لا يترك فيه شيئا. (ص 36)	At night he would wind his legs and arms around me, and let his old, gnarled hand travel all over my body, like the claws of a starving man who has been deprived of real food for many years wipe the bowl of food clean, and leave not a single crumb behind.	Total words =50 P=1*0+5*0+ 10*0/ 50=00 Q= 100-(10* 0)= 100%	At night, he wraps his arms and legs around me, and I let his veined hand roam all over my body, leaving nothing untouched- like the hand of a starving man who hasn't seen food in years, wiping the plate clean and licking it so that nothing remains.	Total words =48 P= 1*0+5*1+10 *0/ 48=00.10 Q= 100-(10* 0.10)= 99%	At night, he coils his arms and leg around me, and I let his calloused hand roam my entire body, sparing nothing—a famished hand that hasn't seen food in years, so it scours the plate and licks it clean, leaving nothing behind.	Total words =43 P= 1*0+5*1+10 *0/ 43= 0.116 Q= 100- (10*0.116)= 98.83
	Minor=0 Major=0 Critical=0		Minor=0 Major=1 Critical=0		Minor=0 Major=1 Critical=0	

In this case, the term used in example 7 is classified as an obscenity since it involves explicit descriptions of physical contact. The translation into human language has captured the pragmatic meaning and style of the source

text, especially the terms "old, gnarled hand," which portray the image of *يده المعروفة*. On the other hand, ChatGPT and DeepSeek have maintained the explicitness in the translation, but each has made one major terminology

mistake by translating the expression as “veined hand” and “calloused hand.” Although these two options do not change the meaning of the text, they have failed to capture the characterizations made in the

source text. Concerning the translation method, all three have used the taboo-to-taboo transfer; however, ChatGPT and DeepSeek have employed lexical substitution, too

**Table 8.** A taboo expression quoted from *Woman at Zero Point* (page 40)

ST	HT	Penalties	ChatGPT	Penalties	DeepSeek	Penalties
فغرز أسنانه في لحم كتفي، و عضني عدة مرات في صدري و بطني و هو يردد: يا مرة يا بنت ال..... (ص: 40)	He sank his teeth into the flesh of my shoulder and bit me several times in the breast, and then over my belly, while he was biting me, he kept on repeating: 'slut, bitch'.	Total word =34 P= 1*0+5*0+1 0*0 / 34=0 Q= 100- (10*0) = 100%	He sank his teeth into the flesh of my shoulder and bit me several times on my chest and stomach, repeating: 'You woman, you daughter of a...!	Total words =27 P= 1*1+5*1+1 0*0 / 27=0.22 Q= 100- (10*0.22) = 97.8%	He sank his teeth into the flesh of my shoulder and bit me repeatedly on my chest and stomach, muttering: 'You wretched woman, you daughter of a...!	Total words =27 P= 1*2+5*2+10* 0 / 27= 0.40 Q= 100- (10*0.40) = 96%
	Minor=1 Major=0 Critical-1		Minor=1 Major=1 Critical-0		Minor=2 Major=2 Critical=0	

The above expression falls under the category of obscenity since it involves the use of both bodily violence and very obscene verbal insults. The human translation has captured both the physicality and the insulting nature of the expression by translating "يا مرة يا بنت ال..." as "slut, bitch," thus capturing the offensive nature of the expression to the target reader. That is, the translator employs the domestication strategy. ChatGPT has maintained the literal sense of the expression but downgraded its insulting nature via literal translation and lexical compensation. DeepSeek has downgraded the insulting nature of the expression even more by using "You wretched woman" in place of the insult and "muttering" in place of "repeating" among other words that have made the translation inaccurate both lexically and

pragmatically. Under MQM standards, these lexical and pragmatic changes explain why the two AI-based translations are inferior compared to the human translation. With regard to the translation technique, a human translator has used cultural adaptation (taboo-to-taboo), ChatGPT has depended on literal translation, while DeepSeek has used euphemization to make a pragmatic deviation from the source text and lose the feminist intensity of the source text.

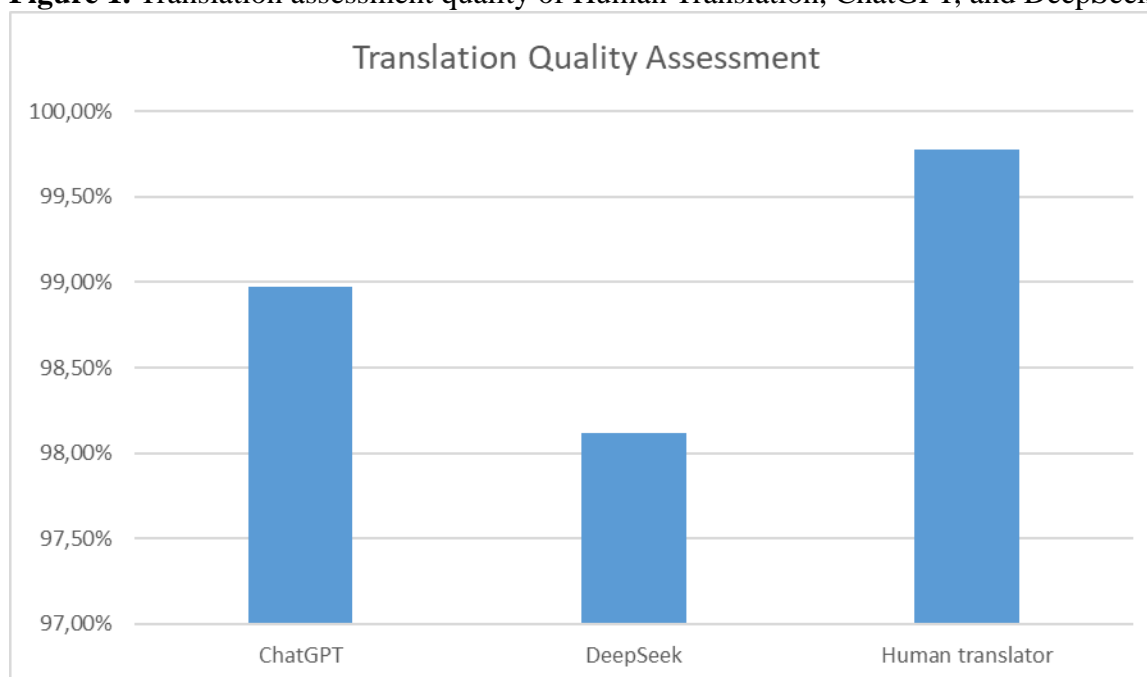
#### 4. Discussion and analysis

Figure presents the overall MQM quality scores calculated from the complete corpus of 50 taboo expressions, revealing only minor quantitative differences among the three translation outputs. The human translation achieved the highest overall quality score (99.78%), followed by ChatGPT

(98.97%) and DeepSeek (98.12%). Although the numerical differences are relatively small, the qualitative analysis demonstrates that the human translator more consistently preserved the pragmatic, cultural, and contextual meanings of taboo expressions, particularly in passages requiring culturally appropriate lexical choices and the preservation of El-Saadawi's feminist discourse. ChatGPT performed competitively, producing highly accurate lexical renderings in several examples, but occasionally introduced grammatical or semantic distortions that

affected pragmatic equivalence. DeepSeek likewise generated fluent translations but showed a greater tendency toward lexical substitution and contextual inaccuracies, resulting in slightly lower MQM scores. Overall, the findings indicate that while contemporary AI systems can achieve translation quality approaching that of professional human translators, human expertise remains superior in preserving the cultural and pragmatic functions of taboo expressions in literary translation.

**Figure 1.** Translation assessment quality of Human Translation, ChatGPT, and DeepSeek



While ChatGPT and DeepSeek delivered highly accurate translations of the source texts into English, the qualitative study revealed the superiority of human translation in conveying pragmatic, cultural, and contextual features of taboo terms. The discrepancies between the three systems did not have much to do with the fluency of translation but had much to do with the choice of vocabulary, interpretation of context, and preservation of the power of culturally sensitive speech. ChatGPT often supplied highly accurate equivalents for words from the source texts and even outperformed human translation in some cases, but at the same time, it often distorted both grammar

and semantics of the examples, affecting their pragmatic quality. DeepSeek usually supplied fluent translations but was prone to lexical substitutions, additions, and inaccurate contextual interpretation.

Further analysis of the translation strategies used shows several distinguishing features between the three translation systems. Firstly, the human translator always applied culturally appropriate equivalents where necessary without losing the pragmatic and ideological function of taboo words. Secondly, ChatGPT was based mainly on the use of direct transfer and managed to provide an accurate translation in most cases; however, there were some problems in

conveying the interpersonal meaning of some sentences. Lastly, DeepSeek preferred using such strategies as lexical replacement and interpretation, which sometimes led to a loss of explicitness or the strength of the text.

Generally speaking, it has been found that modern-day AI technology is capable of achieving a level of quality of translation that is almost the same as that achieved by professionals. Nevertheless, some texts contain several culturally loaded and ideologically charged words and phrases, and they include the novel called *Woman at Point Zero*, which should be professionally translated since only professionals can translate it properly and interpret its content and pragmatics. Consequently, literary translation via AI technology is regarded as an additional option for professional translators.

### 5. Conclusions

The present research examined the efficiency of human translation, ChatGPT, and DeepSeek in translating taboo expressions from Arabic to English in Nawal El-Saadawi's *Woman at Point Zero* by means of the MQM metrics system and Khoshsaligheh and Ameri's (2014) classification of taboo translation strategies. The results indicate that, despite the high MQM scores of the translation tools under consideration, there were certain qualitative distinctions regarding their translation of culturally delicate expressions. The highest overall translation quality score was attained by the human translation (99.78%), proving its greater consistency in preserving the meaning of taboo expressions semantically, pragmatically, and culturally. Even though ChatGPT demonstrated comparable quality in performance (98.97%) and even sometimes offered more accurate lexemes than the human translation in question, it made some grammatical and semantic mistakes in several cases, which negatively influenced the pragmatic equivalence of the translation.

In addition, the qualitative analysis proved that human translator and ChatGPT generally employ a domestication strategy by

making the text accessible and natural for the target readers. The three translations utilized different strategies, such as taboo-to-taboo translation, lexicon replacement, partial deletion, culture-specific transformation, euphemism, or mitigation strategy. The majority of translations maintain the implicature and the face-threatening acts in the target language. This is done as a result of the domestication strategy employed in the translation. This means that an effective translation of taboo terms cannot be simply based on the use of equivalent lexical items. Rather, the cultural and ideological roles played by the source text have to be considered. The human translation, in particular, was able to conserve better the discourse of El-Saadawi's narrative. Although there is some encouraging evidence regarding the success of modern AI programs, the results have indicated that AI must be seen as a helpful tool in translation rather than a substitute for literary translators while translating culturally and ideologically loaded texts. The involvement of humans is necessary to maintain contextuality, pragmatics, and cultural lexicons.

The research is limited to one literary piece, a human translation of it, and only two generative AI tools. As such, the results of the research cannot be generalised to all genres of literature and all AI algorithms. It would be possible in future research to study different literary texts, test other AI translation tools, analyze the translation of figurative language, and assess the efficiency of human post-editing of AI-generated translations of taboo language.

### References

- Abedaal, N. and Al Sarhani, A. (2021). Subtitling Strategies of Swear Words and Taboo Expressions in the Movie "Training Day". *Heliyon*, 4 (7), 1–128. (In English)
- Al-Harashsheh, A. and Sehab, E. (2024). Translation of Paratextual Elements in El-Sa'dawi's *Woman at Point Zero* into English: A Socio-Cultural Perspective. *Journal of Foreign Language Teaching and Translation Studies*, 8 (4), 47–166. (In English)

- Allan, K. and Burrige, K. (2006). *Forbidden words: Taboo and the censoring of language*. Cambridge: Cambridge University Press. (In English)
- Amireh, A. (2000). Framing Nawal El Saadawi: Arab feminism in a transnational world, *Signs: Journal of Women in Culture and Society*, 26 (1), 215–249. (In English)
- Al-Yasin, N. F., and Rabab'ah, G. A. (2019). Arabic audiovisual translation of taboo words in American hip hop movies: A contrastive study, *Babel*, 65 (2), 222–248. (In English)
- Badmus, A. O., Kilani, S. O. and Michael, M. E. (2024). A pragmatic analysis of taboo expressions in Idoma language, *International Journal of Humanities and Communication Studies*, 4 (1), 149–160. (In English)
- Baker, M. (2018). *In other words: A coursebook on translation*, Routledge, UK. (In English)
- Balaa, L. (2018). El Saadawi does not orientalize the other in *Woman at Point Zero*, *Journal of International Women's Studies*, 19 (6), 236–253. (In English)
- Brahimaj, M. R. D. (2024). The use of taboo words and expressions in today's society, *International Journal of Religion*, 5 (10), 2938–2946. (In English)
- Castaldo, A., Castilho, S., Moorkens, J., and Monti, J. (2025). Extending CREAMT: Leveraging large language models for literary translation post-editing, in *Proceedings of Machine Translation Summit XX*, 1, 506–515. (In English)
- Castro, O., and Ergun, E. (2017). Feminist translation studies. *London: Routledge*, 10, 9781315679624. (In English)
- Fitriansyah, F. (2023). An Analysis of Translation Techniques and Quality of Taboo Words in the Gran Torino Movie. *Jurnal Bahasa Sastra dan Studi Amerika*, 29 (1), 13–23. (In English)
- Haider, A. S., Saideen, B. and Hussein, R. F. (2023). Subtitling taboo expressions from a conservative to a more liberal culture: The case of the Arab TV series *Jinn*. *Middle East Journal of Culture and Communication*, 16 (4), 363–385. (In English)
- Guerberof-Arenas, A., and Toral, A. (2022). Creativity in translation: Machine translation as a constraint for literary texts, *Translation spaces*, 11(2), 184–212. (In English)
- House, J. (2015). *Translation quality assessment: Past and present*, Routledge, UK. (In English)
- Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., ... and Awadalla, H. H. (2023). How good are GPT models at machine translation? a comprehensive evaluation, arXiv preprint arXiv:2302.09210. (In English)
- Jay, T. (1992). *Cursing in America*, John Benjamins B.V., Amsterdam. (In English)
- Jay, T. (2009). The utility and ubiquity of taboo words, *Perspectives on psychological science*, 4 (2), 153–161. (In English)
- Khoshsaligheh, M. and Ameri, S. (2014). Translation of taboos in dubbed American crime movies into Persian, *T&i Review*, 4 (2), 25–50. (In English)
- Kilani, S. O., and Michael, M. E. (2024). A Pragmatic Analysis of Taboo Expressions in Idoma Language, *International Journal*, 4 (1). (In English)
- Lestari, R., and Sutrisno, A. (2023). Euphemism of taboo translation in the Big Little Lies Series. *Diglosia: Jurnal Kajian Bahasa, Sastra, Dan Pengajarannya*, 6 (3), 711–722. (In English)
- Olimat, S. N., Mahadin, D. K., Mashaqba, B. and Al-Adwan, A. (2023). Subtitling sensitive Arabic language for English audiences: A case study of Netflix's *Jinn* and *Al-Rawabi* shows, *Ampersand*, 11, 100142. (In English)
- Ljung, M. (2011). *Swearing: A Cross-Cultural Linguistic Study*, Palgrave Macmillan, London. (In English)
- Macken, L. (2024). Machine translation meets large language models: Evaluating ChatGPT's ability to automatically post-edit literary texts. In *Proceedings of the 1st Workshop on Creative-text Translation and Technology*, 65–81. (In English)
- Macken, L., Vanroy, B., Desmet, L. and Tezcan, A. (2022). Literary translation as a three-stage process: Machine translation, post-editing and revision, in *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, 101–110. (In English)
- Nababan, N. and Sumardiono. (2012). Pengembangan model penelitian kualitas terjemahan. *Jurnal Kajian Linguistik dan Sastra*, 24 (1), 39–57. (In English)

Pedersen, J. (2007). How is culture rendered in subtitles? Challenges of Multidimensional Translation, in Proceedings of the Marie Curie Euroconferences MuTra: Challenges of Multidimensional Translation - Saarbrücken 2-6 May 2005, 1–18. Retrieved from <https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-17661> (Accessed 15 June 2026). (In English)

Shormani, M. Q. (2024). Can ChatGPT capture swearing nuances? Evidence from translating Arabic oaths, *arXiv preprint arXiv:2412.02466*. (In English)

Shormani, M. Q. and Al-Samki, A. A. (2025). Translating dialects between ChatGPT and DeepSeek: Yemeni Sana'ani Arabic terms as a case-in-point [version 1; peer review: 2 approved with reservations], *F1000Research* 2025, 14:694. <https://doi.org/10.12688/f1000research.165879.1> (In English)

Simon, S. (1996). Gender in translation. Cultural identity and the politics of transmission, Nueva York y Londres, Routledge, UK. (In English)

Zhao, Y., Zhang, M., Chen, X., Deng, Y., Geng, A., Liu, L., ... and Zhang, Z. (2023). Human evaluation for translation quality of chatgpt: A preliminary study. *HiT-IT 2023*, 282–287. (In English)

Zagood, M. J., Almazrouei, A. H., Alnaqbi, M. S. and Almheiri, F. A. (2022). Translating

Taboos: An Analysis Concepts, Discourses, and Translations, 299. (In English)

### Corpus Material

DeepSeek AI. (2023). GitHub page: <https://github.com/deepseek-ai>

El-Saadawi, N. and Hatata, S. (Trans.). (1983). *Women at Point Zero*, Zed Books, UK. (In English)

OpenAI. (n.d) introducing ChatGPT. Retrieved from: <https://openai.com/blog/chatgpt>

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

*Авторы прочитали и одобрили окончательный вариант рукописи.*

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

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

**Ahmad Al-Harashseh**, PhD, Full Professor of Applied Linguistics, Yarmouk University, Irbid, Jordan.

**Sujood Al-Husban**, MA in Translation, Lecturer, Yarmouk University, Irbid, Jordan.

**Zakaryia Almahasees**, PhD, Assistant Professor of Translation, Yarmouk University, Irbid, Jordan.

## Appendix (A)

**Table 9.** The corpus of the study, 20 taboo expressions quoted from *Woman at Zero Point*

No.	Source Text	Human Translation	ChatGPT	DeepSeek
1	مع أنني عار به الجسد و لا أعرف السباحة(ص 17).	I was naked and knew not how to swim.	Even though I cannot naked and didn't know how to swim.	Even though my body is bare and I cannot swim.
2	و يهرشون دائما تحت إبطهم أو ما بين فخذيهم. (ص: 19)	And constantly scratching under the armpits and between the thighs.	Or they are always scratching under their armpits or between their thighs.	They are always scratching under their armpits or between their thighs.
3	وأنت امرأه معها مطواه أو شفرة موس، و قطعوا قطعة من اللحم بين فخذي. (ص: 19)	Then she brought a woman who was carrying a small knife or maybe a razor blade. They cut of piece of flesh from between my thighs.	She brought a woman carrying a pocketknife or a razor blade, and they cut a piece of flesh from between my thighs.	She brought a woman carrying a dagger or razor blade, and they cut a piece of flesh from between my thighs.
4	و أدركت أنهم كلهم رجال، و نفوسهم شرهة مشوهة، و شهواتهم للمال و الجنس و السلطة لا حدود لها و لا رقابة عليها (صفحة 27)	I discovered that all these rulers were men. What they had in common was an avaricious and distorted personality, a never-ending appetite for money, sex and unlimited power.	And I realized that they were men, and their souls were greedy and distorted, and their desires for money, sex, and power were limitless and unchecked.	And I realized that they, like all men, have voracious and distorted souls, and their cravings for money, sex, and power are boundless and unchecked.
5	حينما يجف الثقب أتركه يقبلني، و أحس الورم الكبير فوق شفتي ووجهي مثل الكيس أو القربة المليئة بالماء الراكد أو الدهن( ص 36)	When the hole was dried up, I let him kiss me. I could feel the swelling on my face and lips like a small purse, or a water skin, full of a stagnant greasy fluid.	“When the sore dries up, I let <u>it</u> kiss me, and I feel the large swelling over my lip and face like a sack or a skin filled with stagnant water or fat.”	When the hole dries, I let him kiss me, and I feel the large swelling over my lip and face like a waterskin filled with stagnant water or grease.
6	و أحس لعابة اللزج داخل فمي، فأطردته بلساني إلى فمه مرة أخرى (ص: 44)	I could feel his thick saliva between my lips and with a push of the tongue sent it back into his mouth.	I curled my lips to spit in his face, but he caught my lips with his teeth.	I feel his viscous saliva inside my mouth, so I thrust it back into his with my tongue.

7	وكل العيون الوقحة التي تعزيني ولا تغض البصر (ص: 19)	Like the spit of all the brazen eyes that underdressed me and examined my nakedness with a slow insolence.	And all the shameless eyes that strip me naked and do not look away.	And all the brazen eyes that strip me bare yet refuse to lower their gaze.
8	و فستاني الخفيف المبلل بالمطر يلتصق بجسدي. (ص: 48)	And my dress was clinging to my body, soaked in rain.	And my light rain-soaked dress clings to my body.	My thin dress, drenched by the rain, clings to my body.
9	و يشف من تحته ثدياي من غير مشد (ص: 48)	My breasts were showing almost naked under my dress.	My breasts show through it without a bra.	Beneath it, my breasts show through, unsupported.
10	حلمة الثدي دائرة سوداء بارزة، ضغط عليها بيده و هو يساعدي بركوب السيارة (ص: 48).	The nipples standing out in two dark circles. As he helped me into the car. He pressed his arm against them.	My nipple, a prominent dark circle, was pressed by his hand as he helped me into the car.	In his dream: a dark, circular breast — prominent. He pressed it with his hand while helping me into the car.
11	أسفل بطني بين فخذني. (ص: 43).	Below it to the place between my thighs.	Below my stomach, between my thighs.	Below my abdomen, between my thighs.
12	و أكور شفتي لأبصق في وجهه، فإذا به يمسك شفتي بأسنانه (ص: 44)	In answered I would purse my lips and prepare to spit in his face, bur he would start biting them with his teeth.	I curl my lips to spit in his face, but he catches my lips between his teeth- have mercy on him.	I purse my lips to spit in his face, but suddenly he clamps them shut with his teeth
13	فرايتها راقدة عارية و إلى فوزي. جوارها (ص: 46)	And found her laying necked with Fawzy at her side.	I saw her lying naked, with Fawzi beside her.	Then I saw her lying naked, with Fawzi beside her.
14	لم أكن أمسك أي جريدة و أرى فيها صورة لأي رجل منهم حتى أبصق على وجهه (ص: 18)	Each time I picked up a newspaper and found the picture of a man who was one of them, I would spit on it.	I couldn't hold a newspaper and see the picture of any of those men without spiting on his face.	I could never hold a newspaper and see any man among them boasting without spiting in his face.
15	و في غير أيام الجفاف أبعد عنة شفتي و أبعد أنفي؛ فهي رائحة تشبه رائحة الكلب الميت تخرج من الثقب. (ص 36)	But on days when it was not dry I would turn my lips and face a way to avoid the odour of dead dogs which emanated from it.	And on days when it isn't dry, I turn my lips and nose away from him; it's a smell like that of a dead dog coming from the sore.	And on days without drought, I pull my lips and nose away from him-a stench like that

				of a dead dog seeps from the hole.
16	أترفعين صوتك على صوتي يا بنت الشوارع يا ساقطة؟! (ص: 40)	How dare you raise your voice when you're speaking to me, you street walker, you low woman?!	Are you raising your voice above mine, you street girl, you slut?!	How dare you raise your voice over mine, you street whore, you slut?!
17	و في الليل يلف ذراعيه و ساقه حولي، و أترك يده المعروقة تعيث بجسدي كله لا تترك شيئاً، كيد جائع لا يرى الطعام من سنين، فإذا به يمسح الصحن و يلغقه لا يترك فيه شيئاً. (ص: 36)	At night he would wind his legs and arms around me, and let his old, gnarled hand travel all over my body, like the claws of a starving man who has been deprived of real food for many years wipe the bowl of food clean, and leave not a single crumb behind.	At night, he wraps his arms and legs around me, and I let his veined hand roam all over my body, leaving nothing untouched-like the hand of a starving man who hasn't seen food in years, wiping the plate clean and licking it so that nothing remains.	At night, he coils his arms and leg around me, and I let his calloused hand roam my entire body, sparing nothing—a famished hand that hasn't seen food in years, so it scours the plate and licks it clean, leaving nothing behind.
18	و انقض علي ككلب مسعور. (ص: 37)	He leapt on me like a mad dog.	He pounced on me like a rabid dog.	He lunged at me like a rabid dog.
19	فغرز أسنانه في لحم كتفي، و عضني عدة مرات في صدري و بطني و هو يردد: يا مرة يا بنت ال..... (ص: 40)	He sank his teeth into the flesh of my shoulder and bit me several times in the breast, and then over my belly, while he was biting me, he kept on repeating: 'slut, bitch'.	He sank his teeth into the flesh of my shoulder and bit me several times on my chest and stomach, repeating: 'You woman, you daughter of a...!'	He sank his teeth into the flesh of my shoulder and bit me repeatedly on my chest and stomach, muttering: 'You wretched woman, you daughter of a...!'
20	تغوص بين ثدي. (ص: 43)	Bury themselves between my breasts.	He buries his face between my breasts.	He sinks between my breasts.

УДК 811.161.1'367.6:811.581'367.6(045)

DOI: 10.18413/2313-8912-2026-12-2-0-2

Баркович А. А.<sup>1</sup>  
Ван Ц.<sup>2</sup>

Квалитативная и квантитативная специфика  
корпусной репрезентации частеречной  
категориальности русского и китайского языков

<sup>1</sup> Белорусский государственный университет  
пр-т Независимости, 4, Минск, 220004, Беларусь  
E-mail: [barkovichaa@gmail.com](mailto:barkovichaa@gmail.com)  
ORCID: 0000-0001-8469-8431

<sup>2</sup> Белорусский государственный университет  
пр-т Независимости, 4, Минск, 220004, Беларусь  
E-mail: [qing.wang@mail.ru](mailto:qing.wang@mail.ru)  
ORCID: 0009-0005-2248-256X

Статья поступила 30 сентября 2025 г.; принята 15 июня 2026 г.;  
опубликована 30 июня 2026 г.

**Аннотация:**

**Актуальность.** Данная статья посвящена анализу корпусной репрезентации частей речи в русском и китайском языках как объективному и актуальному ресурсу лингвистических исследований. Опыт обобщения в работе широкого спектра данных и метаданных является релевантным как для осмысления теоретических основ описания языка, так и для практико-ориентированного использования в прикладных междисциплинарных разработках; в дальнейшем он может быть продуктивно имплементирован в разноаспектную исследовательскую практику и развит. Научная значимость работы определяется тем, что корпусно-верифицированное сопоставление частеречных систем типологически удалённых языков позволяет не только уточнить статистические параметры их функционирования, но и аргументированно поставить вопрос о лингвистической состоятельности дискуссионных категорий – предикатива в русском языке и дифференцирующего слова в китайском.

**Проблематика.** И для русского, и для китайского языков проблематика совершенствования лингвистических категорий оказывается одинаково актуальной, что подтверждается модернизацией созданных для них частеречных систем в реалиях корпусной репрезентации. Такого рода данные позволяют интерпретировать частеречную категориальность как идентифицирующий и квалифицирующий признак метаязыковой практики.

**Методы.** Полученные посредством интегрированных в лингвистические корпуса компьютерных программ данные позволяют на материале принципиально стабильной институции частей речи преодолеть шаблонность и саморефлексию традиционных лингвистических описаний языка. В исследовании был задействован комплексный методологический подход, в том числе, квалитативный и квантитативный анализ, метод корпусных исследований.

**Результаты.** Подтверждена целесообразность метаязыкового позиционирования частеречной категориальности как лингвистической универсалии. В статье – по данным «Национального корпуса русского языка» и «Онлайн-корпуса» китайского

языка – рассмотрены квалитативные и квантитативные особенности функциональности языковых единиц в контексте их частеречной принадлежности. **Выводы.** Полученные корпусные данные о шаблонном реальноречевом функционировании единиц того или иного частеречного кластера в типологически далеких языках убедительно свидетельствуют о целесообразности классификации частей речи как метаязыкового инструментария. В свою очередь, проблематика развития частеречных систем русского и китайского языков является не менее лингвистически актуальной, что подтверждается насущностью выделения новых разрядов частей речи и их фактическим инкорпорированием в корпусную аннотацию. В данной связи показательно, что на основании корпусной репрезентации подтверждена лингвистическая состоятельность таких категорий, как предикатив в русском языке и дифференцирующее слово в китайском. Полученные результаты относятся прежде всего к высокочастотному сегменту лексики обоих языков. Они могут быть применены в практике корпусной аннотации, сопоставительной грамматике и преподавании русского и китайского языков как иностранных.

**Ключевые слова:** Корпусная репрезентация; Часть речи; Квантитативный анализ; Квалитативный анализ; Русский язык; Китайский язык; Частеречная категориальность

**Информация для цитирования:** Баркович А. А., Ван Ц. Квалитативная и квантитативная специфика корпусной репрезентации частеречной категориальности русского и китайского языков // Научный результат. Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 2. С. 25–60. DOI: 10.18413/2313-8912-2026-12-2-0-2

УДК 811.161.1'367.6:811.581'367.6(045)

DOI: 10.18413/2313-8912-2026-12-2-0-2

Aliaksandr A. Barkovich<sup>1</sup>  
Qing Wang<sup>2</sup>

Qualitative and quantitative specificity of corpus  
representation of part-of-speech categoriality in Russian  
and Chinese

<sup>1</sup> Belarusian State University,  
4 Nezavisimosty Ave., Minsk, 220004, Belarus  
E-mail: [barkovichaa@gmail.com](mailto:barkovichaa@gmail.com)  
ORCID: 0000-0001-8469-8431

<sup>2</sup> Belarusian State University,  
4 Nezavisimosty Ave., Minsk, 220004, Belarus  
E-mail: [qing.wang@mail.ru](mailto:qing.wang@mail.ru)  
ORCID: 0009-0005-2248-256X

*Received 30 September 2025; accepted 15 June 2026; June 30 March 2026*

**Abstract:**

**Relevance.** This article analyzes corpus-based representations of parts of speech in Russian and Chinese as an objective and relevant resource for linguistic research. The experience of generalization of a wide range of data and metadata is relevant both for understanding the theoretical foundations of language description and for practice-

oriented use in applied interdisciplinary research. It can be productively implemented and further developed in multifaceted research. The scientific significance of the work is determined by the fact that a corpus-verified comparison of part-of-speech systems of typologically distant languages allows not only to clarify the statistical parameters of their functioning, but also to argue for the question of the linguistic validity of the discussion categories – the predicative in Russian and the differentiating word in Chinese.

**Problems.** Improving linguistic categories is equally relevant for both Russian and Chinese, as evidenced by the modernization of the parts-of-speech systems created for them in the context of corpus representation. Such data enables us to interpret part-of-speech as a feature that identifies and qualifies metalinguistic practice.

**Methods.** Data obtained through computer programs integrated into linguistic corpora enable us to overcome the commonplaceness and self-reflective nature of traditional linguistic descriptions of language using material from a fundamentally stable system of parts of speech. The study employed a comprehensive methodological approach, including qualitative and quantitative analysis and corpus-based technique.

**Results.** The feasibility of the metalinguistic positioning of part-of-speech categoricity as a linguistic universal is confirmed. Using data from the Russian National Corpus and the Online Corpus of Chinese, this article examines the qualitative and quantitative features of the functionality of linguistic units in the context of their part-of-speech affiliation.

**Conclusions.** The obtained corpus data on the units' real-speech functioning stereotypeness of a particular part-of-speech cluster in typologically distinct languages convincingly demonstrates the feasibility of classifying parts of speech as a metalinguistic tool. In turn, the development of part-of-speech systems in Russian and Chinese is no less linguistically pressing, as evidenced by the urgency of identifying new categories of parts of speech and their actual incorporation into the corpus annotation. In this regard, it is significant that the linguistic validity of such categories as the predicative in Russian and the differentiating word in Chinese was confirmed based on corpus representation. The knowledge gained primarily relates to the high-frequency segment of the vocabulary of both languages. The obtained results can be applied in the practice of corpus annotation, comparative grammar and teaching Russian and Chinese as foreign languages.

**Keywords:** Corpus representation; Part of speech; Quantitative analysis; Qualitative analysis; Russian; Chinese; Part-of-speech categorization

**How to cite:** Barkovich, A. A., Wang, Q. (2026). Qualitative and quantitative specificity of corpus representation of part-of-speech categoricity in Russian and Chinese, *Research Result. Theoretical and Applied Linguistics*, 12 (2), 25–60. DOI: 10.18413/2313-8912-2026-12-2-0-2

## 1. Введение

Современная лингвистика, благодаря компьютерным технологиям, приобрела новые возможности в верифицированной репрезентации структуры и функциональности языка как феномена (Barkovich, 2019). Это предполагает, в том числе, корпусную интерпретацию фундаментальных лингвистических

институций. Можно по-разному относиться к их обоснованности, но они созданы и востребованы: «Очевидная безграничность межъязыкового разнообразия породила два крайних подхода к языковым универсалиям. С одной стороны, следуя Аристотелю, Соссюру и Блумфилду в русле вариативности, можно идеализировать

многообразии, вплоть до полного отказа от универсалий. С другой стороны, разделяя увлеченность Платона, Соссюра и Хомского единообразием, можно игнорировать очевидные несоответствия как поверхностные, эфемерные, эпифеноменальные или воображаемые – принадлежащие эйдосу, языку, компетенции – и постулировать чрезвычайно сложные и абстрактные (или формальные) универсалии» (Givón, 1984 : 24).

К общепризнанным и основополагающим универсалиям языка, безусловно, относится и частеречная система. Трудно переоценить актуальность референтных данных о ней в изучении природы языка, особенно в формализации получаемых данных. В развитии системы знаний о языке соответствующие исследования играют важную роль, значимость которых в научной интерпретации современной коммуникации по-прежнему высока. Традиционно фундаментальной в обеспечении стабильной метаязыковой институциональности является частеречная система (Алпатов, 2018; Розенталь и др., 2010; Сунь; Ли, 1993 и др.). При этом, она призвана отражать логику функциональности весьма динамичного объекта моделирования, языка. Конечно, «... каждый язык имеет свою собственную схему» частей речи, характеризующуюся определенными «количеством, природой и необходимыми границами» (Sapir, 1921 : 119). Вместе с тем, изучение частеречной специфики каждого языка способствует общей систематизации релевантных знаний, как каждый фрагмент мозаики улучшает ее визуализацию. При этом сопоставление метаданных о типологически удаленных языках имеет особую ценность. Изучение лингвистической специфики разноструктурных языков является продуктивным, поскольку, пусть при этом анализируются и типологически достаточно далекие языковые системы, –

но они функционируют в общих для всех языков коммуникативных обстоятельствах и являются элементами единого и все более взаимосвязанного интеръязыкового континуума. Более того, поиск общих закономерностей развития типологически далеких языков очевидно целесообразен и по причине высокой вероятности того, что они могут быть экстраполированы и на описание множества более близких языковых систем.

Первые шаги по систематизации постепенно накапливающихся знаний о частях речи были сделаны античными мыслителями. Изначально в фокусе их внимания, естественно, оказался древнегреческий язык; позже внимание ученых концентрировалось на латыни, как языке науки, актуальном весьма протяженный период времени. Позже влияние античных грамматиков воплотилось в созданных по их образцу грамматических системах европейских языков – в них было инкорпорировано и адаптированное к их специфике учение о частях речи. По этому же пути пошли и грамматисты многих неевропейских языков, в том числе, – пусть и с некоторым опозданием – китайские ученые. Подобным же образом зародилась частеречная концепция для русского языка, ориентировавшаяся на уже апробированные шаблоны адаптации классической модели в западноевропейских языках. В русском языке части речи были системно идентифицированы М.В. Ломоносовым в его труде «Российская грамматика», впервые изданном в 1755 г. (Ломоносов, 1855). Для китайского языка система частей речи, как значимая новация, была представлена в работе «Объяснение правил письменного языка господина Ма» («马氏文通»), изданной в 1898 г. (Ma, 1983). Частеречная категориальность в описаниях русского и китайского языков прошла долгий путь становления и является до настоящего времени

дискуссионной. Доступные сегодня посредством корпусных методик данные позволяют приблизить функциональность частеречной системы к современным реалиям коммуникации.

Как металингвистическое представление о классах языковых единиц, *частеречная категориальность* характеризуется рядом семантически, формально и функционально значимых кодифицирующих признаков. Это не означает их полного подобия: в разных языках, особенно разноструктурных, совокупность таких признаков может быть разной, ее параметры и организация могут обладать выразительной уникальностью. В частности, для русского языка характерна флективность и развитая морфология, отражающая специфику многочисленных форм рода, числа, падежа, времени и т. д. В свою очередь, в контексте китайского языка важнейшую нагрузку несет синтаксическое позиционирование языковых единиц, по сути, организующее семантическую архитектуру текста. Лингвистически обоснованная дифференциация языковых единиц типичным образом оказывается дискуссионной на статистически репрезентативном материале: «В этом континууме нет четкой дифференциации ни по одной категории» (Ху, 2018 : 208). Тем не менее, частеречная система совершенно определенно является универсалией, свойственной научным представлениям о возможностях лингвистического моделирования языка как феномена, в целом.

Научная новизна настоящего исследования обеспечена как теоретическим обоснованием синтетического, количественно и качественно сбалансированного подхода к анализу репрезентативных языковых данных, так и практической реализацией методологически инновационной логики комплексного лингвистического исследования. В работе предложена методика формирования пропорционально

сопоставимых эталонных корпусов для изучения типологически удалённых языков на базе материала, изначально существенно различающегося по объёму. В соответствии с данным подходом, в процессе исследования было выполнено системное количественно-качественное сопоставление частеречных систем русского и китайского языков, обеспеченное референтными эталонными корпусами. На основе корректно агрегированных и интерпретированных корпусных данных была сформирована аргументационная база для единообразной параллельной оценки частеречного потенциала предикатива в русском и дифференцирующего слова в китайском языках.

## 2. Методы и материал

Практически любой естественный язык требует сложноорганизованной метаязыковой надстройки, что нередко обуславливает ее существенную вариативность. В данной связи, очевидна уместность и реалистичность корпусно-обеспеченного подхода к оценке фундаментальных грамматических проблем – в востребованном на сегодня дескриптивно-прескриптивном формате. Присущая ему достоверная фактография оказывается не только количественно, но и качественно продуктивной. Конечно, она не всегда подтверждает складывавшиеся долгие годы теоретические стереотипы, но зато позволяет уверенно решать прикладные задачи метаописания языка – особенно актуальные в контексте тотальной информатизации коммуникации. Вместе с тем, корпусный подход никак не отменяет сложившиеся шаблоны лингвистической практики: «Интуиция может быть чрезвычайно хорошим проводником при работе с корпусом. Проще говоря, интуиция лингвиста может предоставить ему основу для обоснованных догадок, позволяющую быстро и эффективно взаимодействовать с корпусом. Аналогичным образом, уже существующие теоретические подходы

могут помочь корпусному лингвисту» (McEnery and Hardie, 2012 : 161).

Количественная характеристика языковой специфики является лингвистически валидной и верифицируемой (Баркович, Ван, 2015). Статистика частотности языковых единиц, полученная из репрезентативных корпусов, таких как «Национальный корпус русского языка» («НКРЯ») и «Онлайн-корпус» китайского языка обладает высокой убедительностью для решения широкого круга лингвистических задач (НКРЯ; Онлайн-корпус). В частности, фактологически обеспеченный анализ частей речи по корпусным данным позволяет сформировать надежный фундамент для оценки их релевантности существующим грамматическим стереотипам (см., напр., Цао, 2011; Национальный корпус; Савчук и др., 2024 и др.). Качественная интерпретация полученной из репрезентативных корпусных источников статистики органично дополняется метаязыковым анализом эмпирически убедительных первичных данных. Такого рода методологический синкретизм полностью оправдан при рассмотрении фундаментальных основ языкознания, позволяя синтезировать новое знание. Частеречная категориальность – традиционно значимая лингвистическая база всех грамматических надстроек языка, что подтверждается ее интеграцией в метаописания таких типологически удаленных языков как русский и китайский. Корпусно-обеспеченный анализ современной речевой практики – целесообразная реализация современных возможностей лингвистической практики, в том числе, углубленного сопоставительного изучения разноструктурных языков.

Цель данного исследования – выявление и описание лингвистически релевантной качественной и количественной специфики корпусной репрезентации частеречной

категориальности русского и китайского языков в высокочастотном регистре.

Объектом исследования являются части речи русского и китайского языков, рассматриваемые как на уровне их качественных характеристик, так и на уровне их количественных показателей по данным корпусов.

Релевантными в данной связи задачами являются следующие:

1) обобщить сложившиеся в языкознании теоретические представления о сущности и особенностях систем частей речи для русского и китайского языков;

2) посредством корпусного инструментария получить объективную количественную информацию о фактическом распределении единиц русского и китайского языков по частеречным разрядам;

3) охарактеризовать качественную корреляцию распределения языковых единиц по частям речи с их функциональной спецификой в речевой практике;

4) соотнести данные о статистической дифференциации единиц русского и китайского языков разной частеречной принадлежности с имеющимися качественными обобщениями;

5) на основании корпусных данных оценить модификационный потенциал частеречных систем русского и китайского языков, в том числе, их метаязыковую сбалансированность.

Цель и задачи исследования обусловили задействование сбалансированного метаязыкового подхода с применением комплекса методов: обобщения, моделирования, структурного, контекстуального, качественного, количественного, компонентного, корпусного анализа, сопоставления.

Эмпирической базой исследования послужили следующие источники материала: русскоязычная специфика выявлялась по «Основному корпусу»

«Национального корпуса русского языка», китайскоязычная – по «Корпусу современного китайского языка» («现代汉语语料库») «Онлайн-корпуса». Оба ресурса являются достаточно репрезентативными: объем привлеченного в качестве базового русскоязычного материала превысил 389 млн. словоупотреблений, объем китайскоязычного – превысил 20 млн. словоупотреблений. Частеречная разметка в вышеупомянутых корпусах различается: «НКРЯ» оснащен полноценной морфологической аннотацией статистического характера; тогда как «Онлайн-корпус» воплощает преимущественно функциональный подход по модели, основанной на правилах, – что обусловлено типологической спецификой китайского языка, в частности, отсутствием в нем морфологического словоизменения. Жанровый состав корпусов также не является идентичным. Тем не менее, качество аутентичного материала задействованных корпусов позволило решить поставленные в исследовании задачи. Количественно объем «НКРЯ» существенно превышает объем «Онлайн-корпуса». Это обстоятельство было учтено в ходе исследования: целесообразное сопоставление проводилось на основе полностью сопоставимых данных эталонных корпусов, сформированных по единому шаблону репрезентации материала. Изначальная диспропорция объемов анализируемого материала в задействованных корпусах была нивелирована путем агрегации эталонных корпусов, составленных из 500 наиболее востребованных единиц для каждого языка. Данный порог был выбран как оптимальный с точки зрения баланса между репрезентативностью и управляемостью материала: он охватывает высокочастотный сегмент лексики, обеспечивающий анализ значимых статистических закономерностей, и, вместе с тем, допускает детальный

квалитативный анализ каждой единицы. Оба сформированных эталонных корпуса использовались как сопоставимые по объёму и принципам отбора базы данных, которые были проанализированы в соответствии с задачами исследования. В рамках исследования единицами материала являлись лексемы русского языка и сегментированные слова ( $\vec{w}$ ) китайского языка, идентифицируемые с учётом их частеречной разметки. Омонимичные и полифункциональные единицы (особенно применительно к китайскоязычному материалу) анализировались с учётом контекста и их преобладающей синтаксической функции в текстах выборки. Прошедшие автоматическую сегментацию и частеречную разметку омонимичные и полифункциональные единицы учитывались отдельно для каждого частеречного значения в соответствии с разметкой корпуса. Так, если одна и та же графическая форма аннотировалась в качестве репрезентанта разных частей речи, такие вхождения фиксировались как отдельные частеречные реализации. Для русскоязычного материала более детализированная и репрезентативная дополнительная выборка не понадобилась. Для углубленного анализа функциональности единиц китайского языка дополнительно на базе «Онлайн-корпуса» была сформирована расширенная выборка из 16 254 наиболее частотных единиц. Такой порог выборки обеспечил возможность анализа размеченных как частеречные классы совокупностей языковых единиц в составе не менее 50 контекстов. Показатель в 16 254 единиц был установлен экспериментальным путём как минимальное значение, при котором корпус-менеджер «Онлайн-корпуса» сформировал репрезентативную выборку, включившую все частеречные классы, в том числе не представленные либо недостаточно представленные в базовой выборке из 500 единиц. При этом, полученный в результате статистически

актуальный набор единиц с полной частеречной разметкой позволил избежать избыточной корпусной выборки, которая могла бы затруднить целесообразный анализ материала. Данная ситуация обусловлена большим знаковым разнообразием китайскоязычной графики, но на достоверность выводов для обоих языков не повлияла. Опционально идентичность исследованных языковых единиц сверялась с «Толковым словарем русского языка» и «Словарем современного китайского языка» (Ожегов и Шведова, 2006; Словарь современного китайского языка, 2016).

В целом, реализованная исследовательская стратегия позволила обеспечить необходимый уровень репрезентативности материала и объективно охарактеризовать специфику лингвистической интерпретации частеречной принадлежности языковых единиц русского и китайского языков. Полученные результаты позволили не только охарактеризовать совокупность референтных языковых единиц, но и задействовать их метаописание для формирования аргументированной позиции по возможной модернизации частеречных систем русского и китайского языков.

### 3. Результаты

Для того чтобы всесторонне изучить системы частей речи в русском и китайском языках, в исследовании потребовалось не только тщательно проанализировать квантитативные данные, извлечённые из языковых корпусов, но и обобщить полученные данные в квалитативном ключе, опираясь на имеющуюся теоретическую базу. Задействование встроенного в корпусные ресурсы аналитического инструментария полностью отвечает современному тренду лингвистики: «Поскольку тексты становятся онлайн доступными во всё возрастающих объемах и на всё большем количестве языков, растёт потребность в надёжных методах обработки, которые

могут анализировать текст без дорогой и трудоёмкой адаптации к новым областям и жанрам» (Schütze, 1995 : 141). И «НКРЯ», и «Онлайн-корпус» в достаточной мере обеспечены актуальными на сегодняшний день аутентичными образцами речевой продукции, что позволяет оценить текущие реалии языкового функционирования. Вместе с тем, как показало проведенное исследование, статистика требует интерпретации и лишь в этом случае может обладать верифицированной научной ценностью, будучи совместимой с уже имеющимися теоретическими наработками. Реализованный в данном исследовании подход позволил получить новые сведения о том, как функционируют языковые единицы, как они взаимодействуют между собой и какими метаязыковыми свойствами обладают. Исследование дало возможность на статистически убедительном материале охарактеризовать распределение языковых единиц по частям речи в каждом из двух языков и сделать выводы относительно перспектив модификации частеречных систем для данных языков. В фокусе исследования оказались категории предикатива в русском языке и дифференцирующего слова в китайском языке, состоятельность частеречного потенциала которых была, в целом, подтверждена в синтетическом квантитативно-квалитативном аспекте.

#### 3.1. Квалитативная специфика частей речи русского и китайского языков

Обеспеченность языковой практики тем или иным набором частей речи убедительно свидетельствует о присутствии ее разновидностям определенном спектре модальности. Соответствующая квалитативная специфика, будучи референтной качественным аспектам языкового функционирования, в равной мере присуща и русскому, и китайскому языку, что позволяет ее интерпретировать в контексте данного исследования параллельно. Корпусные данные

ориентированы на языковую практику, которая и должна служить основанием широких метаязыковых обобщений: «В концепции, ориентированной на языковую практику, где грамматика напрямую основана на языковом опыте, нет типов данных, которые исключаются из рассмотрения, поскольку они считаются отражающими реализацию...» (Bybee, 2010 : 11). В данном аспекте, например, в случае насыщенности речи существительными и прилагательными, целесообразно предполагать статичность и описательность ее содержания. В свою очередь, насыщенность текстов глагольными формами позволяет уверенно констатировать динамику и хронологическое развитие содержания речевых фрагментов.

Рассмотрим, как это может быть проиллюстрировано и интерпретировано применительно к проблематике данного исследования. В частности, характерным для **модальности действительности / недействительности** оказывается наличие в контексте русскоязычных текстов определенных **прилагательных** или **наречий**. Так, фактически в русском языке соответствующая модальность зачастую маркируется единицами, прямо указывающими на оценку субъектом 'реальности' / 'нереальности' описываемой им ситуации. Например, прилагательные, специфичные для таких контекстов, можно условно разделить на две группы. Прилагательные, характерные модальности действительности подчёркивают 'реальность', 'достоверность', 'фактичность' описываемых событий: *реальный, действительный, подлинный, истинный, фактический, достоверный* и др. Прилагательные **модальности недействительности** указывают на 'нереальность', 'иллюзорность', 'вымышленность' описываемых событий: *мнимый, кажущийся, иллюзорный, фиктивный, ложный, воображаемый*, и др. Многие из таких прилагательных

имеют антонимичную пару. Наречия модальности действительности / недействительности, подобно соответствующим прилагательным, также, обладают ярко выраженной семантикой: *действительно, истинно, фактически* и др., с одной стороны, и *якобы, словно, вряд ли* и др. – с другой стороны. Факультативным маркером такой модальности выступают специфичные для оцениваемых ситуаций **глаголы**. Например, в русском языке:

*Преодолеть инстинкт, заставляющий животных спастись бегством от истинной или мнимой опасности, невозможно.<sup>1</sup>*

*Она кажется слишком правдоподобной, и новички вряд ли бы стал интересоваться, существует ли на самом деле Любезность.*

Для китайского языка характерен аналогичный набор языковых средств. Здесь данную функцию выполняют преимущественно наречия (**副词**) и прилагательные (**形容词**), причём граница между ними зачастую размыта, а их частеречная идентичность зависит от контекста. Так, для формирования модальности действительности ключевыми являются следующие языковые единицы: *确实* ('действительно', 'в самом деле'), *的确* ('воистину', 'точно'), *真实* ('истинно'), *果然* ('и в самом деле', 'как и ожидалось'). О модальности недействительности свидетельствуют такие языковые единицы, как *似乎* ('кажется', 'будто'), *好像* ('как будто', 'словно'), *仿佛* ('словно', 'как будто'), *虚假* ('ложно'). При переводе данные единицы китайского языка в большинстве случаев передаются наречиями русского языка. Например:

<sup>1</sup> Здесь и далее русскоязычные примеры приводятся по данным «НКРЯ».

现在越来越多的材料证明，美洲大陆确实是在公元五世纪的时候被中国人首先发现的。<sup>2</sup> (Всё больше материалов свидетельствует о том, что американский континент действительно был впервые открыт китайцами в V веке нашей эры).

他似乎不知道这件事，但眼神出卖了他。(Будто он ничего не знал об этом, но взгляд его выдал).

那位收费的人看见这情况，似乎也有点怜悯之情，过来看了看孩子，说：“病得确实不轻！鼻子都有点扇了！”(Кассир, увидев это, словно проникся состраданием, подошёл посмотреть на ребёнка и сказал: “Болен действительно серьёзно – даже нос начал увеличиваться!”).

В последнем примере примечательно сочетание в одном контексте языковых единиц **确实** и **似乎** – выступающих как наречия. Здесь маркеры модальности действительности и модальности недействительности соседствуют, создавая сложную модальную семантику.

Важное типологическое различие между русским и китайским языками состоит в том, что в китайском модальность действительности / недействительности выражается исключительно лексически – через наречия и прилагательные, тогда как в русском языке существенным является частое присутствие в соответствующих контекстах глаголов в сослагательном наклонении. Это делает участвующие в формировании модальности действительности / недействительности прилагательные и наречия особенно значимыми для формирования модальности в китайской речи.

<sup>2</sup> Здесь и далее китайскоязычные примеры приводятся по данным «Онлайн-корпуса».

Подобным образом, сопровождаются специфичной глагольной лексикой контексты, которым присуща **потенциальная модальность** (с семантикой ‘возможности’ / ‘невозможности’); **дебитивная модальность** (с семантикой ‘необходимости’ и ‘долженствования’); **гипотетическая модальность** (с семантикой ‘предположительности’); **интенциональная модальность** (с семантикой ‘намерения’); **императивная модальность** (с семантикой ‘побудительности’); **оптативная модальность** (с семантикой ‘желательности’). Идентифицируемые – в плане присутствия в них единиц определенной частеречной категориальности – конструкции, содержащие глаголы, характерны для множества контекстов, фактически модально маркированных тем самым. Это, в равной мере, присуще и русскоязычным, и китайскоязычным контекстам. Например, в русском языке:

**Необходимо отметить, что несмотря на некоторые опасения со стороны правоохранительных органов, массовое мероприятие прошло очень спокойно, никаких серьезных инцидентов зафиксировано не было.**

**В целом можно сказать, что новые банковские системы денежных переводов работают по тем же принципам, что и гранды рынка переводов MG и WU.**

В китайском языке:

**基本民生的意义应该也包含扩大社会保障, 比如提高医疗费用报销比例, 扩大失业保险的覆盖面或者工伤保险覆盖面等。** (А обеспечение основного народного благосостояния должно включать в себя расширение социальной защиты, например, повышение оплаты гражданских медицинских расходов, расширение страхования от безработицы или производственного травматизма).

**据了解, 要将壁画完整地搬到电脑里, 前期需要采集海量信息, 后期数据拼接**

、整合、存储等一系列工作都需要人工手段完成。(Как стало известно, для того, чтобы перенести полную фреску в компьютер, прежде всего, необходимо собрать огромный объем реальных данных, затем специалистам нужно проделать работу по соединению элементов фрески и сохранению соответствующих данных в ручном режиме).

Для персонализированной речи свойственна «местоименная» насыщенность речевых фрагментов – при описании субъектности коммуникантов, характеристике их отношения к тому или иному действию или его признаку. Содержание речевых фрагментов может дифференцировано проецироваться посредством всего спектра местоимений. Для русского языка это, прежде всего личные местоимения, с помощью которых презентуется, в частности, позиция адресанта – посредством единиц *я* и *мы* (местоимения первого лица); позиция адресанта – посредством единиц *ты* и *вы* (местоимения второго лица); позиция наблюдателя – посредством единиц *он*, *она*, *оно* и *они* (местоимения третьего лица). Для китайского языка для соответствующих коммуникативных ситуаций актуален аналогичный перечень языковых средств: это местоимения *我* ('я') и *我们* ('мы'), как элементы категории местоимений первого лица, или *第一人称代词*; это *你* ('ты') и *您* ('вы') / *你们* ('вы'), как элементы категории местоимений второго лица, или *第二人称代词*, и местоимения *他* ('он'), *她* ('она'), *它* ('оно') и *他们*, *她们*, *它们* ('они'), как элементы категории местоимений третьего лица, или *第三人称代词*.

Например, в русском языке:

Из тех, кто может свою судьбу перевернуть и из тех, кто плывет по течению. Вот ты смог и я смог. А Коля тот, у него даже мечты нет!

А ты думаешь, они умеют? – хмыкнул Двинятин.

В китайском языке:

他说:

“我特别喜欢中华传统文化, 和中国朋友一起过春节非常开心, 我还亲手写了福字给他们拜年, 祝他们幸福安康诸事顺遂。” («Мне очень нравится китайская традиционная культура. Я неописуемо рад возможности отпраздновать Праздник Весны с китайскими друзьями. Чтобы поздравить китайских друзей с Новым годом, я написал китайский иероглиф «Фу», который должен принести им счастье, благополучие и удачу», – сказал юный житель Благовещенска).

你明天有事吗? 你能跟我们两一起去

◦ (У тебя завтра есть дела? Ты можешь пойти вместе с нами?).

В последнем примере использовано два местоимения: дважды *你* ('ты') и единожды *我们* ('мы'). При этом, здесь единица *我们* не включает в группу субъектов сообщения адресата: адресант под сочетанием иероглифов *我们* имеет в виду себя и постороннего человека (не собеседника). По этой причине используется именно комбинация иероглифов *我们*, а не, например, сочетание *咱们*, использование которого подразумевало бы включенность собеседника в состав субъектов, обозначаемых местоимением как *мы*.

**Временная** актуализация текста, в частности, при его ассоциации с определенными временными ориентирами, сопровождается своими специфическими наречиями и предлогами. Типичным образом описываемые события привязываются к моменту речи, а востребованными маркерами временной отнесенности коммуникационной ситуации выступают, в частности, такие единицы русского языка как *после*, *сейчас*, *до* и т. д. В китайском языке типичными в

подобных ситуациях языковыми единицами являются *已经* ('уже'), *刚刚* ('как только'), *当* ('в', 'на'), *从* ('с', 'от') и др. Например, в русском языке:

*Записать, не откладывая. Нет, потом, не сейчас. А иногда, ты знаешь, я просто боюсь начать писать о тебе.*

*Когда соус закипит, в него влить картофельный крахмал, разведенный мадерой (при вливании нужно размешивать соус лопаткой во избежание образования комков), и довести до кипения, после чего процедить через частое сито.*

В китайском языке:

*等我赶到医院的时候,他已经做完手术了,正在恢复室里休息。(Когда я добрался до больницы, ему уже сделали операцию и он отдыхал в палате восстановления).*

*她刚刚走出门口,电话就响了。*

*(Как только она вышла за дверь, тут же зазвонил телефон).*

В первом из вышеприведенных контекстов языковая единица *已经* решает задачу временной актуализации, формируя таксисное отношение предшествования: действие (завершение операции) совершилось раньше, чем наступил момент отсчёта времени (приезд субъекта). Далее уже частица *了* как показатель завершённости в постпозиции к глаголу конкретизирует это значение. Во втором контексте языковая единица *刚刚* также актуализирует событие – как только что завершившееся; далее частица *就* вводит следующее событие. В обоих контекстах наречия выступают как типичные средства временной актуализации. При этом, их использование компенсирует отсутствие в китайском языке глагольного времени как грамматической категории: там, где в русском языке используются видо-временные формы глагола, в китайском

задействуются лексические средства наподобие *已经* и *刚刚* – в сочетании с частицами.

**Локальная** актуализация той или иной коммуникационной ситуации может быть идентифицирована по фактам задействования в речи определенного ассортимента **наречий** и **предлогов**. Сложно представить, что носитель языка осознанно имплементирует в речевой поток какие-либо языковые единицы на основе их частеречной принадлежности, выбор языковых средств коммуникантами происходит интуитивно – у них, обычно, просто нет времени на подобные изыски. Тем не менее, сам факт использования в речи единиц определенного свойства, в том числе, соотносимо их ассоциированности с какой-либо частью речи достаточно красноречиво свидетельствует о выраженной таким образом модальности сообщения. В русском языке типичным образом о локальной актуализации содержания речевого фрагмента позволяет судить достаточно компактный набор референтных единиц, таких как *вверху*, *внизу*, *далеко*, *близко*, *здесь*, *там* и т. д. Для китайского языка не менее характерны свои языковые маркеры, принадлежащие, по определению, к узкому кругу наречий или предлогов – такие как *下面* ('под, внизу'), *上面* ('на, над, вверху'), *那里* ('там'), *这里* ('здесь') и др. Например, в русском языке:

*Там они вершат – здесь они отдыхают, совершенно инкогнито, в том смысле, что их никто и не пытается узнать.*

*Не знаю, близко ль, далеко ль, не знаю, В какой стране и при луне какой, Веселая, забытая, родная, Звучала ты, как песня за рекой.*

В китайском языке:

*伍月这时还算懂事,马上配合他说:*

*"他明天一早的飞机,要到西安参加书市。"*

但在桌子下面踢了严守——脚。严守吓了一跳，急忙把脚收了回来。(У Юэ благоразумно сразу же подыграла ему: «Он летит завтра ранним утром в Сиань, на книжную ярмарку». Но под столом она пнула Янь Шоуи. Тот вздрогнул и торопливо отдернул ногу назад).

大厦前脸有个院落，院落门口有座牌坊；沿着牌坊，拉着警戒线；警戒线处，有五六个门卫守着，不让人进。(Перед зданием находился дворик с аркой на входе, вдоль которой тянулась заградительная лента; у ограждения стояло человек пять-шесть охранников, никого не пропуская).

В первом примере присутствует 到 – предложный показатель направления ('в'); 在 – предложный показатель места ('под') дополнен локативным показателем со значением 'внизу' – 下面. Во втором – языковая единица 沿着, совмещающая функции наречия и предлога, вводит обстоятельство места с оттенком направленности ('вдоль чего-либо'); показатель 着 из ее состава указывает на 'длительное, протяжённое действие или положение'.

Таким образом, качественное обобщение языковых данных о присутствии тех или иных частей речи в тексте объективно свидетельствует о характере и специфике модальности и может быть описано посредством лингвистически корректных моделей.

### 3.2. Квантитативная специфика частей речи русского и китайского языков

При всей результативности качественного анализа речи посредством корпусной методики, не менее лингвистически значимым оказывается её квантитативный анализ на материале лингвистических корпусов. Корпусные данные о функциональности тех или иных частей речи обладают вполне определенной ценностью для уточнения самых разнообразных параметров и структуры языкового материала:

«Квантитативный анализ частей речи позволяет выявить не только стилистические, но и диахронические и жанровые закономерности» (Francis and Kučera, 1982 : 198). Они актуальны как при метаязыковом исследовании отдельных текстов, так и при оценке всего спектра языковых отношений на базе речевой практики. Как объективная проекция языковой системы, речь, в конечном итоге, отражает все действительно релевантные языку как объекту научного представления возможности его рефлексии. Извечная проблема заключается в том, чтобы их выявить и интерпретировать. Корпусный формат в данной связи является апробированным и надежным инструментарием: посредством его уже решались достаточно сложные метаязыковые задачи (Захаров, Богданова, 2011; Partee, 1987; Gries, 2009; Wang et al., 2025 и др.). Наряду с широким кругом насущных лингвистических задач, корпусы, благодаря так называемой POS-разметке, как компоненту морфологической аннотации, референтны и в статистической репрезентации закономерностей строения и функционирования системы частей речи. Ввиду их высокого типологического своеобразия, в контексте данного исследования системы частей речи для русского и китайского языков на базе корпусного материала целесообразно рассматривать дифференцированно.

#### 3.2.1. Интерпретация квантитативной специфики данных русского языка

Для получения лингвистически специфицированных данных о сравнительной статистике встречаемости языковых единиц русского языка в «Основном корпусе» «НКРЯ» был сформирован соответствующий запрос относительно 500 наиболее востребованных языковых единиц. Наличие в данном ресурсе разноплановой разметки изначально позволяло рассчитывать на интерпретацию

полученных данных в аспекте их грамматической идентичности – морфологическая разметка является базовой и основной практически для всех лингвистически компетентных корпусных ресурсов. Опираясь на верифицированную репрезентативность корпусных данных «НКРЯ», охватывающих широкий спектр речевой практики, из полученных данных был сформирован эталонный корпус языковых единиц. Анализ полученных

данных проводился по двухаспектной методике: оценивалась как востребованность лексем той или иной части речи, так и востребованность словоупотреблений единиц с привязкой к их частеречной идентичности. Данные были представлены как в абсолютном (в целых числах), так и в относительном (процентном), или удельном, выражении; перечень частей речи был представлен в виде рейтинга (Таблица 1).

**Таблица 1.** Сравнительная частотность единиц эталонного корпуса, представляющих разные части речи русского языка

**Table 1.** Comparative frequency of units in the reference corpus representing different parts of speech in the Russian language

Перечень наиболее распространенных в эталонном корпусе частей речи	Статистика выявленных лексем в абсолютном выражении	Статистика выявленных лексем в процентном выражении	Статистика выявленных словоупотреблений в абсолютном выражении	Статистика выявленных словоупотреблений в процентном выражении
существительное	161	32,53 %	27 481 950	13,47 %
глагол	99	20,00 %	23 250 959	11,40 %
наречие	57	11,51 %	12 661 920	6,20 %
прилагательное	48	9,70 %	7 167 531	3,51 %
местоимение	39	7,88 %	43 790 885	21,46 %
предлог	29	5,86 %	39 875 883	19,54 %
частица	20	4,04 %	14 341 685	7,03 %
союз	17	3,43 %	30 433 604	14,92 %
числительное	14	2,83 %	2 685 379	1,32 %
предикатив	5	1,01 %	1 297 353	0,64 %
вводное слово	5	1,01 %	715 656	0,35 %
междометие	1	0,20 %	330 245	0,16 %
ИТОГО	495*	100 %	204 033 050	100 %

\* Из общего объема наиболее частотных единиц русского языка (500) были исключены цифры (5).

Согласно данным эталонного корпуса наивысшие показатели частотности – как совокупность единиц определенной части речи – продемонстрировали **существительные** русского языка. Их совокупный удельный вес среди всего количества идентифицированных корпус-менеджером «НКРЯ» языковых единиц составил 32,53 %, что в абсолютном

выражении составило 161 лексему. Наиболее употребительной единицей из их числа оказалась лексема **год**, ее частотность составила 1 111 801 словоупотреблений – около 4 % от общего количества существительных. Например:

*Даже если на знакомство с каждым из этих ресурсов Сети затрачивать не более минуты, то на просмотр пятой*

части доступных русских серверов уйдет целый год рабочего времени.

В 1996 году повесть была опубликована в журнале «Октябрь».

Если принимать в расчет общее количество соответствующих определенной части речи словоупотреблений, самой востребованной из них следует признать *местоимение* – с показателем в 21,46 %. Из ассортимента местоимений наиболее востребованной единицей оказалось *я* – с показателем в 4 998 377 случаев использования, что составило более 11 % от общего объема соответствующей совокупности. Например:

*Поэтому считаю, что бороться надо на всех фронтах, в том числе и личными примерами – я, например, тоже не курю, никогда не курил, выступаю за здоровый образ жизни и всех к этому призываю.*

*Я не раз приводил аргументы, что он своими поправками пытается спихнуть груз ответственности на работников...*

Согласно полученным данным, была выявлена речевая реализация и пока не являющихся «официальными» частями речи предикатива и вводного слова. Наличие отдельных типов разметки в корпусе подтверждает реальность их потенциального позиционирования как частей речи. Языковые единицы вышеупомянутых категорий обнаруживаются в эталонном корпусе, являясь значительно более распространенными, чем, например, входящие в «академический» перечень частей речи русского языка междометия.

Однако очевидно, что необходимо принимать в расчет не только количественные показатели той или иной категории языковых единиц. При качественном рассмотрении функциональности в речи, в частности, *вводного слова*, выявляется его существенная парадигматически-синтагматическая ограниченность. Так, вводные слова совпадают по форме с

единицами других частей речи и не случайно обособляются на письме запятыми. Согласно сложившимся представлениям, вводным словом называется «слово, не связанное формально-грамматически с другими словами в предложении, не являющееся членом предложения и выражающее, например, отношение говорящего к высказываемому, отношение данной мысли к предыдущей или последующей речи или указывающее на то, что говорящий считает высказанную мысль чужой» (Ахманова, 1966: 69). Не обладая собственной содержательной значимостью, вводное слово является своеобразным комментарием к содержанию фрагмента речи, отражает эмоциональное состояние и прагматические установки говорящего. Фактически, оно не несет какой-либо существенной грамматической нагрузки в предложении – той среде, в контексте которой типичным образом проявляется грамматическая идентичность языковой единицы как части речи. Вся совокупность количественно-качественных характеристик вводного слова не дает достаточных оснований констатировать его соответствие частеречному статусу.

В свою очередь, проведенный анализ речевой функциональности *предикатива* достаточно убедительно подтверждает потенциал его частеречной категориальности. Представленность предикатива оказывается вполне репрезентативной: в количественном измерении предикатив занимает девятое место в общем рейтинге размеченных «НКРЯ» словоупотреблений – как репрезентантов частеречных категорий языковых единиц. В числе 500 наиболее частотных лексем в эталонном корпусе обнаруживается небольшой ассортимент предикативов – только пять: *можно, нельзя, нет, мало, нужно*. Например:

*Неужели нужно выделять кого-то из длинного ряда исполнителей, подчеркивать, как далась ему роль, или*

вообще сравнивать образы с историческими портретами?

Времени то *нет*, да и вообще понимаешь ли, эмансипация.

Вместе с тем, статистика говорит о высокой востребованности предикатива в речи: всего пять предикативов в количественном измерении дали суммарно 1 297 353 единицы в рейтинге частотности. Это, само по себе, веское основание для констатации существенного частеречного потенциала предикатива. При этом, качественный анализ речевой функциональности контекстов с присутствием предикатива не менее убедительно свидетельствует о его синтагматической состоятельности и парадигматической идентичности. По совокупности качественно-количественных характеристик подтверждается, в целом, обоснованность позиционирования предикатива как части речи. В грамматическом плане предикатив отличается от наречий прежде всего своей синтаксической функцией: он выступает исключительно в роли главного члена безличного предложения и не может выступать в позиции обстоятельства, типичной для наречий. В отличие от кратких прилагательных, предикативы не имеют полной формы и не изменяются по родам. От «категории состояния» в

традиционной трактовке предикатив отличается тем, что включает единицы, выражающие не только состояние (*нельзя, можно*), но и оценочную квалификацию (*нужно, мало*) и бытийную констатацию (*нет*). Устойчивость данных синтаксических и семантических признаков подтверждает правомерность выделения предикатива в самостоятельный частеречный класс.

### 3.2.2. Интерпретация количественной специфики данных китайского языка

В запросе корпусному менеджеру «Онлайн-корпуса» по поиску референтных данных в подкорпусе «Корпус современного китайского языка» целесообразным оказалось задание порога репрезентативности количественной выборки в 2 000 словоупотреблений. Такое ограничение позволило агрегировать действительно статистически значимые языковые единицы в релевантной «Статистической таблице по частотности» («词频统计表»). В свою очередь, полученные таким путем данные послужили базой эталонного корпуса для решения задач исследования: он был сформирован на базе 500 наиболее частотных языковых единиц корпусного ресурса (Таблица 2).

**Таблица 2.** Сравнительная частотность единиц эталонного корпуса, представляющих разные части речи китайского языка

**Table 2.** Comparative frequency of units in the reference corpus representing different parts of speech in the Chinese language

Перечень наиболее распространенных в эталонном корпусе частей речи	Статистика выявленных лексем в абсолютном выражении	Статистика выявленных лексем в процентном выражении	Статистика выявленных словоупотреблений в абсолютном выражении	Статистика выявленных словоупотреблений в процентном выражении
существительное	159	32,32 %	838 322	17,62 %
глагол	128	26,02 %	926 571	19,48 %
наречие	47	9,55 %	421 379	8,86 %
местоимение	37	7,52 %	465 729	9,79 %
предлог	30	6,10 %	365 330	7,68 %

союз	29	5,89 %	279 053	5,87 %
прилагательное	25	5,08 %	151 956	3,19 %
числительное	16	3,25 %	173 406	3,65 %
частица	15	3,05 %	1 057 440	22,23 %
счетное слово	6	1,22 %	77 637	1,63 %
ИТОГО	492**	100 %	4 756 823	100 %

\*\* Из общего объема наиболее частотных единиц китайского языка (500) были исключены цифры (8).

Корпусный анализ статистически репрезентативного материала позволил получить лингвистически информативные данные о распространенности тех или иных частей речи в китайском языке. Данные о специфике устной и письменной речи могут отличаться, но в корпусном масштабе и устная, и письменная речь достаточно объективно отражают метаязыковую идентичность языка, являясь надежным материалом для лингвистических обобщений. Полученные данные о фактическом речевом функционировании китайского языка не противоречат статистике русского языка – здесь также самыми распространенными являются языковые единицы, выполняющие номинативную функцию. Для китайского языка анализ речевой обусловленности языковых единиц особенно актуален: одна и та же единица может использоваться, например, и как существительное, и как глагол. Основанная на синтаксической функции языковых единиц корпусная разметка успешно решает возникающие при лингвистическом анализе дилеммы подобного рода. Так, по данным эталонного корпуса наиболее востребованы в китайском языке единицы в роли **существительных**, общий объем которых составил 32,32 % от общего количества исследованных языковых единиц. Этот показатель весьма близок данным русской речи – 32,53 %. Однако в китайском языке обнаруживаются свои приоритеты словоупотребления: здесь самым распространенным существительным оказалась единица **中** ('середина'), использованная 40 105 раз. Например:

他站在屋子的**中间**。(Он стоял посередине комнаты).

Помимо того, что иероглиф **中** может выступать в роли разных частей речи, его функциональность как частицы тоже достаточно разнообразна. Например, такая частица может быть локативным показателем (например, 'внутри') или аспектуальным показателем (например, 'в процессе'):

他把这件事深深地埋在**心中**·从不对人提起。(Он глубоко похоронил это в сердце и никогда никому не рассказывал).

谈判目前仍在**进行中**·双方尚未达成协议。(Переговоры в настоящее время всё ещё **продолжаются**, стороны пока не достигли соглашения).

В первом примере **中** – постпозитивный локативный показатель после существительного **心** ('сердце') – переводится предлогом *в*. Во втором примере **中** – постпозитивный аспектуальный показатель после единицы **进行** ('проводить', 'вести') – переводится глаголом *продолжаются*.

При этом многие содержащие единицу **中** контексты при переводе требуют достаточно сложных переводческих трансформаций. Например:  
 值得一提的是,在满洲里口岸此次**进口煤猛增中**,中国国有企业成为主力军。  
 (Стоит отметить, что на фоне стремительного увеличения объёмов ввозимого через Маньчжурию угля

главными потребителями стали китайские госпредприятия).

Отличается от реалий русского языка и китайскоязычная статистика о распространенности частей речи по количеству словоупотреблений. Здесь наиболее распространены *частицы* – в русском языке местоимения. В китайской речи частицы заняли 22,23 % от общего объема словоупотреблений. Наиболее распространенной единицей оказалась *的* – она была использована 744 863 раз. Она может определять, в том числе, специфику ‘принадлежности’. Например:

*这是我的书。* (Это моя книга).

В данном контексте частица *的* указывает на принадлежность предмета (книги) и точно соответствует русскому притяжательному местоимению *мой / моя / моё*, например в модели «чей-либо + существительное». В отличие от русского языка, где принадлежность выражается флексией или отдельным словом, в китайском языке *的* – единственный универсальный показатель соответствующих отношений, – не изменяющийся ни по роду, ни по числу, ни по падежу. Вообще, частица *的* – одна из самых многофункциональных единиц китайского языка. Она может выступать в нескольких совершенно разных ролях: как структурный элемент (атрибутивная связка): *我的书* – ‘моя книга’; как номинализатор: *红的* – ‘то, что красное’; как финализатор (в конце предложения для акцентирования): *他会来的* – ‘он придёт точно’; как локализатор, выделяя обстоятельство в составе конструкции *是...的*: *他是昨天来的* – ‘он приехал вчера’.

Подобные ситуации достаточно типичны в самых разных языках: «В языках, в которых независимые и зависимые формы различаются по парадигматической структуре, парадигмы независимых маркеров, как правило, богаче парадигм зависимых маркеров» (Siewierska, 2004 : 112). Характерно, что

такие содержащие *的* контексты, как например следующий, не могут переводиться на русский язык дословно:

*用于测试无人交通的路段位于北京中关村北部地区。* (Основной участок для проверки систем автономного вождения расположен в пекинском районе Чжунгуаньцунь).

Согласно информации, содержащейся в «Словаре современного китайского языка», являющегося, в свою очередь, официальным нормативным изданием Института языкознания Академии общественных наук КНР, в китайском языке 12 частей речи. Однако в сделанной по данным «Онлайн-корпуса» выборке – для наиболее востребованных 500 лексем – корпусным менеджером были идентифицированы единицы 10 частей речи. Вместе с тем, этот показатель характеризует лишь базовую выборку из 500 наиболее частотных единиц. Из числа «академически признанных» частей речи в выборку эталонного корпуса не попали звукоподражательные слова и междометия. Это не значит, что их нет в речи: для них выполнена разметка и они находятся в корпусных текстах при расширенном поиске. Расширенная выборка из 16 254 единиц (см. ниже) демонстрирует иное распределение частей речи, включая категории, отсутствующие в базовой выборке.

Например, корпус содержит контексты со следующими *междометиями* *哎呀* (‘Ах!’) и *哎哟* (‘Ой!’):

*哎呀，原来是你！* (Ах, оказывается, это ты!).

*哎哟！我肚子好疼！* (Ой! Как у меня болит живот!).

*到了咖啡厅，喝了一杯咖啡，严守一似乎突然想起什么，看了看表：*

*“哎哟，都十一点半了，我下午一点还得录像，该化妆去了。”* (В кафе они выпили по чашке кофе и тут Янь Шоуи, словно опомнившись, взглянул на часы: *Ой*, уже

половина двенадцатого. У меня в час съемка, еще нужно забежать в гримерку).

Междометие *哎呀* (āiyā) выражает 'неожиданное узнавание', 'приятное удивление'. Его ближайший русский эквивалент – *Ах!* Междометие *哎哟* (āiyō) выражает 'боль', 'физическое страдание' или 'досаду'. Его ближайший русский эквивалент – *Ой!* Строго говоря, в китайском языке нет полных аналогов *Ах!* и *Ой!* и функцию этих языковых единиц распределяют между собой междометия *啊* (a), *哦* (o), *哎呀* (āiyā) и *哎哟* (āiyō) – в зависимости от передаваемой эмоции: 'удивления', 'боли', 'восхищения' или 'узнавания'.

**Звукоподражательные слова** при увеличении порога выборки также в корпусе находятся. Например, это единицы *呼啦呼啦* (hūlāhūlā), *啪* (pā), *哐当* (kuāngdāng):

*冷风呼啦啦扫动着满地的枯叶纸屑，*

*更显出景象格外的凄冷。* (Холодный ветер с шелестом-хлопаньем «шурх-шурх» мёл по земле сухие листья и обрывки бумаги, отчего картина выглядела ещё более мрачной и холодной).

*啪！鞭子一挥，马儿立刻飞奔起来。*

(«Хлоп!» – кнут взвился, и лошадь тут же понеслась вперёд).

*“哐当”一声巨响，这么静的夜，我身体陡然一惊，心头提到嗓子边了。*

(Раздался оглушительный «бум!», и в такой тихой ночи я вздрогнул всем телом, аж сердце чуть не выпрыгнуло из груди).

*他把门 “哐当”*

*一声关上，才听到屋里传来沈雪像狼一样的嚎叫，接着是她痛哭的声音。* (Он захлопнул дверь – «бабах!» – и лишь тогда услышал, как из комнаты донёся вой Шэнь Сюэ, похожий на волчий, а следом – звуки её горького плача.)

Звукоподражательное слово *呼啦呼啦* выражает 'шум ветра', 'хлопанье флага', 'шелест крыльев', 'повторяющийся

хлопок' – передаёт протяжный повторяющийся звук. По-русски естественнее всего оно передается как *шурх-шурх*, *хлоп-хлоп* или *ишиш*. Звукоподражание *啪* означает 'резкий хлопок', 'выстрел', 'удар рукой' или 'падение предмета' – звук резкого однократного удара. По-русски: *хлоп*, *щёлк* или *бах*. Звукоподражательное слово *哐当* имитирует 'глухой удар по металлу или дереву', 'однократный резкий грохот', например, звук захлопывания двери или падения металлического предмета. По-русски ближе всего: *бабах* или *бум*. В последнем примере звукоподражательное слово *哐当* сопровождается единицей *一声*, означающей 'один звук' 'один раз', то есть вся конструкция *“哐当”一声* означает буквально 'с одним звуком «бабах»'. И это типичная для китайского языка модель формирования структуры предложения: *глагол + звукоподражательное слово + 一声 + результат действия*.

Вместе с тем, спектр аннотированных, но не попавших в массив наиболее востребованных категорий единиц китайского языка (для 500 единиц), не исчерпывается междометиями и звукоподражательными словами. Подобным им образом, для текстов корпуса предусмотрена аннотация такого знакового для китайского языка разряда лексики, как **дифференцирующее слово**. В качественном аспекте функциональность относящихся к данной категории языковых единиц уникальна и их использование в речи носит системный характер. В том числе, они идентифицируются в контекстах «Онлайн-корпуса». Например, здесь встречается дифференцирующее слово *女* ('женский'):

*教育女职工正确处理家庭关系上海闸北发电厂把教育女职工正确处理家庭关系作为建设精神文明的一个重要内容，对这方面的先进人物和先进事迹，进行广泛宣*

传·树立榜样。  
electrostation «Чжабэй»  
рассматривает обучение **сотрудниц**  
правильному выстраиванию семейных  
отношений как важную часть духовно-  
нравственного воспитания. Передовые  
**работницы** и их достижения в этой  
области широко освещаются, а их  
примеры служат образцом для  
подражания).

Данный фрагмент текста с разметкой,  
будет выглядеть так:

教育/v 女/f 职工/n 正确/a 处理/v  
家庭/n 关系/n 上海闸北发电厂/ni 把/p  
教育/v 女/f 职工/n 正确/a 处理/v 家庭/n  
关系/n 作为/v 建设/v 精神文明/n 的/u  
一个/mq 重要/a 内容/n , /w 对/p 这/r  
方面/n 的/u 先进/a 人物/n 和/c 先进/a  
事迹/n , /w 进行/v 广泛/a 宣传/v , /w  
树立/v 榜样/n 。/w.

Отличаясь от близкого к нему по  
форме прилагательного собственной  
функциональной идентичностью,  
дифференцирующее слово значимо как  
класс языковых единиц при рассмотрении  
многих теоретических и практических  
вопросов языкознания. Выделение такого  
рода единиц средствами корпусной  
разметки достаточно показательно. В  
текстах «Онлайн-корпуса» они находятся  
как специфическая категория, благодаря  
тегу *f*, позволяющему уверенно отличать  
дифференцирующие слова от близких к  
ним по атрибутивной значимости  
прилагательных. Так, в разметке  
приведенного выше предложения  
дифференцирующее слово **女** ('женский')  
как раз и снабжено специальным тегом *f*,  
который обеспечивает определяемому им  
иероглифу **职工** ('сотрудник' / 'работник')  
– помеченному тегом *n* как  
существительное – соответствующую  
коннотацию. Присвоенная данной  
языковой единице уже как **女职工**

('сотрудница' / 'работница') итоговая  
кодировка имеет вид **女/f 职工/n**.  
Аналогичным образом ведут себя и другие  
единицы данного класса. Так, например,  
дифференцирующее слово **野生** ('дикий') в  
сочетании **野生动物** ('дикое животное')  
идентифицировано как **野生/f** (в  
совокупности – **野生/f 动物/n**); единица  
**初级** ('начальный') в сочетании  
иероглифов **初级阶段** ('начальная стадия')  
получает разметку **初级/f** (в совокупности  
– **初级/f 阶段/n**) и т. д. Во всех подобных  
случаях тег *f* последовательно  
присваивается определяемому  
компоненту, что подтверждает системность  
данной категории в разметке корпуса.

В общем, в «Онлайн-корпусе  
китайского языка» используются  
следующие теги: *n* – 'нарицательное  
существительное'; *nt* – 'темпоратив'; *nd* –  
'существительное со значением  
направления (или директив)'; *nl* –  
'локатив'; *nh* – 'собственное имя'; *nhf* –  
'фамилия'; *nhs* – 'имя'; *ns* –  
'географическое название'; *nn* – 'название  
нации'; *ni* – 'название организации или  
учреждения'; *nz* – 'другое собственное  
имя'; *v* – 'глагол'; *vd* – 'глагол  
направления'; *vl* – 'глагол-связка'; *vi* –  
'модальный глагол'; *a* – 'прилагательное'; *f* –  
'дифференцирующее слово'; *m* –  
'числительное'; *q* – 'счетное слово'; *d* –  
'наречие'; *r* – 'местоимение'; *p* – 'предлог';  
*c* – 'союз'; *u* – 'частица'; *e* – 'междометие';  
*o* – 'звукоподражательное слово'; *j* –  
'аббревиатура'; *w* – 'знак препинания'; *ws* –  
'некитайский символ'; *wi* – 'другой  
неизвестный символ'. В частности,  
дифференцирующее слово здесь  
присутствует как категория метаописания  
и обозначается символом *f*, позволяя  
отграничить данные единицы от,  
например, прилагательных, обозначаемых  
символом *a*. Фактически, данный тегсет  
позволяет выполнять аннотацию

китайскоязычных текстов достаточно полно, что может быть проиллюстрировано следующими примерами.

服装<sup>n</sup> 裁剪<sup>v</sup> 是<sup>vl</sup> 一<sup>m</sup> 项<sup>q</sup> 较<sup>d</sup> 复杂<sup>a</sup> 的<sup>u</sup> 技术<sup>n</sup> , /w 初学者<sup>n</sup> 一般<sup>a</sup> 应从<sup>v</sup> 简单<sup>a</sup> 裁剪<sup>v</sup> 开始<sup>v</sup> , /w 可以<sup>vu</sup> 先<sup>d</sup> 学<sup>v</sup> 剪裁<sup>v</sup> 女<sup>f</sup> 裤<sup>n</sup> 。 /w (Кройка одежды является достаточно сложной технологией, как правило, начинают с простого, например с кройки **женских штанов**).

会上<sup>nl</sup> , /w 全国人大<sup>j</sup> 内务<sup>n</sup> 司法<sup>v</sup> 委员会<sup>n</sup> 副<sup>f</sup> 主任委员<sup>n</sup> 邹瑜作<sup>nh</sup> 了<sup>u</sup> 内务<sup>n</sup> 司法<sup>v</sup> 委员会<sup>n</sup> 关于<sup>p</sup> 最高人民法院<sup>ni</sup> 工作<sup>n</sup> 报告<sup>n</sup> 审议<sup>v</sup> 结果<sup>n</sup> 的<sup>u</sup> 报告<sup>n</sup> , /w 关于<sup>p</sup> 最高人民法院<sup>ni</sup> 工作<sup>n</sup> 报告<sup>n</sup> 审议<sup>v</sup> 结果<sup>n</sup> 的<sup>u</sup> 报告<sup>n</sup> 。 /w (На совещании **заместитель председателя Комиссии по внутренним делам и юстиции Всекитайского собрания народных представителей** Зоу Юй делал доклады о результатах рассмотрения Комиссией по внутренним делам и юстиции деятельности Верховного народного суда и о результатах деятельности Верховной народной прокуратуры).

В вышеприведенных примерах единицы 女 ('женский') и 副 ('заместитель') непосредственно сочетаются с существительными 裤 ('штаны') и 主任委员 ('председатель').

Характерно, что при позиционировании языковых единиц, обозначающих признаки предмета, – их местоположение в сочетании с дифференцирующими словами оказывается инвариантным – перед ними. Однако, при достаточно типичном добавлении частицы 的 место дифференцирующего слова в предложении

может становиться более свободным. В приведенных ниже примерах 复合 'сложный' и 高级 'высокоорганизованный' являются дифференцирующими словами. И в сопровождении данной частицы в предложении их необходимо размещать перед прилагательными – 总 ('общий') и 特殊 ('специфичный'), соответственно.

这<sup>r</sup> 段<sup>q</sup> 话<sup>n</sup> 的<sup>u</sup> 意思<sup>n</sup> , /w 即是<sup>v</sup> 说<sup>v</sup> , /w 历史<sup>n</sup> 是<sup>vl</sup> 社会<sup>n</sup> 的<sup>u</sup> 诸<sup>r</sup> 个人<sup>n</sup> 之<sup>u</sup> 意识<sup>n</sup> 的<sup>u</sup> 活动<sup>v</sup> 创造<sup>v</sup> 出来<sup>vd</sup> 的<sup>u</sup> , /w 历史<sup>n</sup> 的<sup>u</sup> 必然<sup>n</sup> , /w 即是<sup>v</sup> 诸<sup>r</sup> 个人<sup>n</sup> 的<sup>u</sup> 活动<sup>v</sup> 错综<sup>n</sup> 复合<sup>f</sup> 的<sup>u</sup> 总<sup>a</sup> 结果<sup>n</sup> 。 /w (Смысл данного высказывания заключается в том, что история обусловлена сознательной общественной деятельностью человека, а историческая необходимость представляет собой **сложный общий результат** деятельности человека).

其次<sup>c</sup> , /w 人类<sup>n</sup> 社会<sup>n</sup> 同<sup>p</sup> 自然界<sup>n</sup> 一样<sup>a</sup> , /w 也是<sup>v</sup> 物质<sup>n</sup> 运动<sup>n</sup> 形式<sup>n</sup> , /w 不过<sup>c</sup> 它<sup>r</sup> 又<sup>d</sup> 是<sup>vl</sup> 一<sup>m</sup> 种<sup>q</sup> 更<sup>d</sup> 高级<sup>f</sup> 的<sup>u</sup> 特殊<sup>a</sup> 的<sup>u</sup> 物质<sup>n</sup> 运动<sup>n</sup> 形式<sup>n</sup> 。 /w (Во-вторых, человеческое общество, как и природа, является формой движения материи, представляя более **высокоорганизованную специфичную** форму движения материи).

В грамматическом плане дифференцирующие слова обнаруживают ряд признаков, отличающих их как от прилагательных, так и от существительных в атрибутивной функции. В отличие от прилагательных, дифференцирующие слова не имеют степеней сравнения и не могут сочетаться с соответствующими наречиями (很, 非常 – 'очень'). В отличие от существительных в атрибутивной функции, они не требуют наличия структурной частицы 的 в препозиции к

определяемому слову. Дифференцирующие слова образуют закрытую совокупность языковых единиц, что является характерным признаком самодостаточной грамматикализованной категории. От существительных, выступающих в роли компонентов сложных слов – например, 钢笔 ('стальная ручка' как 'авторучка'), где иероглиф 钢 размечен тегом *n* как существительное – дифференцирующие слова отличаются тем, что сохраняют синтаксическую автономность. Они могут выступать определениями при разных существительных – в регулярной, продуктивной модели, например, 女装 ('женская одежда'), 女厕所 ('женский туалет'), 女演员 ('женщина-актриса'). При этом, компоненты сложных слов образуют с определяемым компонентом лексикализованное единство с непродуктивной, идиоматичной сочетаемостью. Эта регулярная продуктивная сочетаемость при отсутствии признаков прилагательного (степени сравнения, сочетаемость со степенными наречиями) и отсутствии признаков существительного в атрибутивной функции убедительно обуславливает выделение дифференцирующего слова в самостоятельный частеречный класс.

Статистическая значимость языковых единиц той или иной части речи зависит от множества факторов, которые не всегда могут быть учтены в какой-то универсальной схеме. Однако масштабируемость корпусных данных позволяет исследовать речевую практику на практически неограниченную глубину – в пределах собранной в корпусе совокупности текстов.

Экспериментальным путем целесообразный порог может быть найден даже без использования специальных программных фильтров. Логично предположить, что имплементация специальных тегов уровня частеречной разметки для дифференцирующего слова обусловлена его фактической значимостью для сбалансированной и непротиворечивой аннотации речи: для разработчиков спецификации корпусного менеджера «Онлайн-корпуса» функциональность дифференцирующего слова в речи очевидна. Для подтверждения данного предположения был проведен соответствующий эксперимент. Благо, что корпусная выборка может корректироваться и в программном обеспечении «Онлайн-корпуса» такая опция предусмотрена. Так, при обращении к описанной выше расширенной выборке из 16 254 единиц анализируемого подкорпуса, в выдаче оказывается более широкий спектр языковых единиц – по сравнению с данными эталонного корпуса. В перечне уже оказываются и междометия, и звукоподражательные слова, и, наконец, дифференцирующие слова: соответственно, 7 226 междометий, что составило 0,08 % относительной частотности; 2 176 звукоподражательных слов, что составило 0,03 % относительной частотности, и 2 794 дифференцирующих слов, что, также, составило 0,03 % относительной частотности (Таблица 3). Подобная избирательность речевой практики, на самом деле, вполне ординарна: в первом корпусе (достаточно репрезентативном «Брауновском корпусе») его создатели с удивлением не обнаружили множество ожидаемой лексики, например, лексемы *smile* (англ. 'улыбка').

**Таблица 3.** Сравнительная частотность единиц «Онлайн-корпуса», представляющих разные части речи китайского языка на материале расширенной выборки из 16 254 единиц

**Table 3.** Comparative frequency of units in the Online Corpus representing different parts of speech in Chinese, based on an expanded sample of 16,254 units

Перечень наиболее распространенных частей речи	Статистика выявленных словоупотреблений в абсолютном выражении	Статистика выявленных словоупотреблений в процентном выражении
существительное	2 573 011	29,35 %
глагол	2 299 497	26,23 %
наречие	630 701	7,20 %
местоимение	536 815	6,13 %
предлог	388 659	4,43 %
союз	327 045	3,73 %
прилагательное	548 764	6,26 %
числительное	220 321	2,51 %
частица	1 073 898	12,25 %
счетное слово	155 194	1,77 %
дифференцирующее слово	2 794	0,03 %
междометие	7 226	0,08 %
звукоподражательное слово	2 176	0,03 %
<b>ИТОГО</b>	<b>8 766 101</b>	<b>100 %</b>

Таким образом, демонстрируя значимую статистику сопоставимого с единицами из «официального» перечня частей речи словоупотребления, в частности для разрядов междометия и звукоподражательного слова, разряд дифференцирующего слова на их фоне выглядит вполне репрезентативно. Подобно междометию и звукоподражательному слову, дифференцирующее слово не лидирует в частеречном рейтинге (Таблица 3), однако и не является в нем избыточным элементом, подтверждая обоснованность своего корпусного позиционирования в ряду других частей речи. Этот тезис кроме вышеподтвержденной количественной обеспеченности – дифференцирующее слово не менее распространено, чем «официальные» части речи звукоподражательное слово и междометие – может быть подкреплен убедительной качественной аргументацией: важной в данном контексте представляется самостоятельная грамматическая значимость дифференцирующего слова в речи.

Полученные результаты показывают, что частеречный состав эталонных корпусов русского и китайского языка имеет как общие черты, так и различия. Так, по распространенности категорий показатели оказались, в общем, близкие: и в русскоязычном, и в китайскоязычном материале на первом месте *существительные*; второе и третье места по распространенности заняли, соответственно, *глаголы* и *наречия*. Тем не менее, сравнительная статистика общей массы словоупотреблений имеет принципиальное отличие: в русской речи наиболее востребовано являющееся знаменательной частью речи *местоимение*, в китайской – являющаяся служебной частью речи *частица*.

#### 4. Дискуссия

Выявляемые современными методами исследования речи противоречия в существующих метаязыковых шаблонах структурирования языка являются следствием его некогда интуитивного моделирования. В контексте тотального охвата языковой практики сегодня уже,

как минимум, нерационально рассматривать язык как застывший во времени артефакт. Накопленный багаж лингвистических знаний о законах и обычаях современной коммуникации обуславливает целесообразность не только заполнения многочисленных лакун, но и критический пересмотр многих уже устоявшихся стереотипов. Практика зачастую далека от традиционно удобных для описания идеальных схем. Технологичность и эмпиризм современной лингвистики безальтернативно свидетельствуют о том, что язык мало зависит от его описаний, а вот «... формальные структуры в грамматиках зависят от языка» (Croft, 2001: 34). Сегодня в науке очевидно смещение парадигмы от интуитивно-субъективных оценок к объективным, эмпирически обоснованным выводам. Этот переход стал возможен во многом благодаря развитию компьютерных технологий, задействование которых предоставляет исследователям верифицированные инструменты для анализа больших массивов языковых данных: «Как и проявления иных языковых системообразующих отношений реализация модификационного потенциала языка в компьютерно-опосредованной речевой сфере нередко отличается от сложившихся стереотипов, позволяя принять во внимание ресурсы и возможные направления развития той или иной семантики» (Баркович, 2018: 11).

Обоснованная коррекция лингвистических категорий не ставит под сомнение системность организации языка. Однако глубоко свойственная языку системность, в свою очередь, обуславливает тщательность проверки новых или уточненных языковых данных, принятие которых может существенно повлиять на стабильность самой системы. Сама по себе статистика, в том числе корпусная, не является универсальной материей, способной заполнить все пробелы в научном описании. Тем не

менее, игнорирование доступных посредством современных инструментов данных чревато схоластичностью: «В наши дни мы рискуем столкнуться с навязыванием нам проблемных аналитических систем, основанных на традициях языкового анализа, которые в прошлом целенаправленно отвергали корпусные данные. Сложность возникает из-за того, что те направления компьютерной лингвистики, которые раньше игнорировали корпусные данные, в последние годы радикально изменили своё отношение к корпусам, но сохранили модели языка, которые не подтверждаются имеющимися у них данными» (Sinclair, 1991: 22). Частеречная категориальность, несмотря на долгий срок ее эволюции и апробации, не является навсегда закрытой для совершенствования лингвистической моделью. А варианты ее модификации, как показало проведенное исследование, есть. Особенно убедительно такая аргументация подкрепляется сопоставительным анализом резервов грамматической организации для разноструктурных языков. Китайский язык, с его ярко выраженным аналитизмом, как оказалось, нуждается в частеречном метаописании не менее преимущественно синтетического русского. Части речи – фундаментальная категориальность языка: «В системе частей речи отражаются стадии развития данного языка, его грамматический строй» (Виноградов, 2001: 17). Конечно, частеречная категориальность разных языков обладает в разной степени выраженной спецификой и подвержена влиянию как собственно языковых, так и экстраязыковых факторов. Однако выверенность и адаптивность самой частеречной системы чрезвычайно важна для классификации и разноуровневой кодификации языковых единиц. Структура частей речи так или иначе учитывается при формулировке самых разных грамматических правил. Не менее актуален частеречный бэкграунд при анализе речевой функциональности

языковых единиц и описании их взаимодействия.

Корректность лингвистической интерпретации языка всегда была востребована, еще более актуальной она стала в условиях формализации коммуникации и ее насыщения множеством искусственных языков, принципиально нацеленных на максимальную прозрачность и упрощенность метаязыка. Однако логичной интенциональности таких ожиданий и устремлений зачастую не соответствует уже сложившийся *status quo* теоретических концепций. Устранение несовершенства теорий о языке на практике бывает зачастую осложнено не всегда очевидными обстоятельствами, или «нечеткими множествами» данных. Так, для строгой категориальной дифференциации единиц китайского языка существенным препятствием является их генетическая **полифункциональность** – заложенная в самой природе китайского языка вариативность грамматической идентичности языковых единиц. Более того, полифункциональность огромного количества единиц китайского языка осложнена отсутствием выраженных морфологических маркеров того или иного их статуса, а также стилистическими особенностями речевой деятельности (Чжан, 2002–2003). Стилистика использования многих китайских слов часто зависит от того, в прямом или переносном значении они употребляются в каком-то конкретном контексте, что в конечном итоге, влияет на их грамматический статус (Дай, 2007). Вследствие этого, в частности, согласно «Введению в современный китайский язык» Шао Цзинмина полифункциональные слова следует делить на 6 основных групп (Шао, 2008 : 172). Эта данность обуславливает тот факт, что большое количество языковых единиц китайского языка может обладать разной идентичностью – в зависимости от фактически выполняемой ими

семантической или грамматической нагрузки, которую они несут в речи. Это напрямую влияет и на их зачастую ситуативную частеречную принадлежность. Использование единиц китайского языка как той или иной части речи по умолчанию сложно прогнозируемо и зависит от текущей синтаксической позиции (Lan et al., 2023; Wang et al., 2026 и др.). Логика полифункциональности присуща и отдельным единицам русского языка, где она идентифицируется как **антимерия** (Московин, 2013: 25). При антимерии имеет место «... использование слова одной части речи в несвойственной этому слову функции другой части речи или окказиональное образование частей речи» (Культура русской речи, 2011: 55). Однако в русском языке частеречная полифункциональность, или антимерия, не носит массового характера, будучи ассоциируемой, скорее, с неким украшательством речи. Вместе с тем, интерференция многочисленных форм русскоязычной лексики нередко воплощается в омонимии (Ахманова, 1966: 275). Будучи подобными по форме, омонимы в русском языке не меняют свою частеречную принадлежность в зависимости от выполняемой в речи функции и этим они существенно отличаются от полифункциональных единиц китайского языка.

Несмотря на сложность и подвижность частеречной организации языка, в признании ее научной ценности для сопровождения языков различного строя и типологической принадлежности у специалистов имеется полный консенсус. Система частей речи – многогранный лингвистический артефакт, в создание и совершенствование которого внесли вклад многие поколения ученых – от мыслителей античности до технологически оснащенных современных исследователей. Однако вопросы актуализации грамматических категорий современных языков по-прежнему активно обсуждаются (Langacker, 1987; Goldberg, 1995;

Tomasello, 2003 и др.). Даже такой фундаментальный лингвистический концепт как слово поныне является достаточно условным: слова зачастую лишь «... минимальные терминальные элементы синтаксической структуры, [и] воплощение принципов, регулирующих появление морфологического материала» (Anderson, 1992: 17). Характеристика языковых единиц является глобальной проблемой языкознания: «У лингвистов нет надежной базы для идентификации слов в разных языках. ...» (Haspelmath, 2011: 24). Тем не менее, такую базу лингвистика современности обрела с появлением компьютерных технологий. И это уже привело не только к смене лингвистических парадигм, но и к изменению взглядов на допустимость модификации широкого спектра канонов языкознания. Будучи до середины XX в., скорее, гипотетической опцией лингвистической практики, эмпиризм компьютерной эры позволил исследователям заглянуть за горизонт теоретических построений и опереться на фактографию речи. Переоценить значимость компьютеризированной обработки речи сложно до сих пор, но до полной инвентаризации языкового богатства по-прежнему далеко, что можно видеть на примере данного исследования.

Значимость корпусного анализа систем частей речи рассмотренных разноструктурных языков обусловлена близким к полярному позиционированием русского и китайского языков. Задачу такого уровня сложности достаточно сложно было бы выполнить только традиционными методами – без привлечения корпусной методики. Вместе с тем, параллельный анализ речевого функционирования языковых единиц удаленных языков должен основываться на верифицированной как в плане критериев, так и параметров категориальной базе, рациональность и обоснованность которых в значительной степени опирается на систему частей речи.

Подкрепленная статистикой систематизация фундаментальных основ грамматик русского и китайского языков, в свою очередь, позволяет решать широкий круг вопросов корректного описания формальных, содержательных и функциональных параметров речевой практики. Это востребовано уже для элементарной, на первый взгляд, лексикографической репрезентации языковых единиц. Современные требования к обработке языкового материала обуславливают вывод лингвистического описания за рамки интроспекции и субъективных суждений исследователя. Объективная работа с большими массивами текстов даёт возможность фиксировать реальные закономерности употребления языковых единиц на основе их частотных характеристик и дистрибутивных свойств. Это означает возможность верифицировать грамматические категории на материале живой речи, а не только на основе прескриптивных представлений. Верифицированность обработки первичных данных, в свою очередь, обеспечивает надежный фундамент для формирования дальнейших метаязыковых обобщений, что в равной мере актуально для релевантного описания самых разных языков (Tognini-Bonelli, 2001; Hunston, 2002; Levshina, 2022 и др.).

Как можно видеть на примере данного исследования, корпусный подход позволяет, в целом, объективно сопоставить разноструктурные русский и китайский языки – при ощутимой разнице в их грамматическом описании и их лингвокультурном своеобразии. Сопоставительное метаописание русского и китайского языков представляет собой одну из наиболее сложных задач современного языкознания. Типологическая дистанция между ними очевидна: русский язык относится к флективным языкам с развитой морфологией, тогда как китайский является изолирующим языком, в котором

грамматические отношения выражаются преимущественно порядком слов и служебными морфемами. Именно эта асимметрия делает корпусный подход особенно ценным инструментом их сравнения, поскольку он опирается на статистически репрезентативный языковой материал.

При обилии обсуждаемых версий грамматической системы русского языка, за общепризнанный ориентир целесообразно принять так называемый «академический» перечень частей речи: «... наиболее распространенная трактовка выделяет в русском языке 10 частей речи» (Сичинава, 2018: 11). В китайской лингвистике базовый вариант, рекомендуемый для использования в образовательных целях, состоит из 12 частей речи (Учебный план, 1984). Разница достаточно существенная, однако при ближайшем рассмотрении можно заметить, что и русскоязычная, и китайскоязычная системы во многом похожи. В обоих языках противопоставляются знаменательные и служебные части речи, выделяются такие практически равноценные части речи как существительные, глаголы и прилагательные. Вместе с тем, статус наречия и междометия в русском и китайском языках различен. Так, в русской грамматике наречие входит в число знаменательных частей речи, а в китайской лингвистике оно идентифицируется как служебная часть речи. Междометие грамматикой русского языка наделено особым статусом и не является ни знаменательной, ни служебной частью речи, а в китайской грамматике оно отнесено к служебным частям речи. Примечательно, что китайская грамматическая традиция включает в систему частей речи особые классы единиц, не имеющие прямых аналогов в русском языке: счётные слова, обязательно сопровождающие числительные при счёте предметов, и звукоподражательные слова,

также выделяемые в самостоятельный разряд.

Отдельного внимания здесь заслуживает проблема потенциально новых классов языковых единиц, чья частеречная принадлежность остаётся дискуссионной. Для них постепенно в оборот термин «периферийная категория». В русском языке к таким категориям относится предикатив (или категория состояния), в китайском – дифференцирующее слово, обладающие признаками как прилагательного, так и наречия. В данном контексте, по-видимому, именно системный анализ речевой практики может стать основой аргументации для выделения или невыделения подобных единиц в самостоятельные классы, – с учетом реалий их синтаксической нагрузки в контекстах разной распространенности.

Так или иначе, как артефакт, система частей речи при необходимости может быть модифицирована, как и любая грамматическая надстройка естественного языка. Более того, по мере углубления знаний о языке, ее структура может адаптироваться к новым реалиям языкового функционирования или исследовательским приоритетам. Изначально рамки частеречных категорий условны: «Поскольку некоторые слова могут быть более "пограничными", чем другие в пределах класса слов, неудивительно, что границы между двумя классами слов также могут быть нечеткими» (Viber et al., 1999 : 24). Ввиду чрезвычайной вариативности и подвижности речевой практики в языковой системе – подобно статическому электричеству в динамичных физических объектах – в ней перманентно накапливается модификационный потенциал. Это обусловлено целым рядом факторов, в том числе, изначально заложенной в грамматике искусственностью; уникальностью античных языков, грамматические описания для которых часто копировались

учеными для далеко не во всем им подобных языков; эволюцией социокультурной сферы, обслуживаемой тем или иным языком и т. д. Не менее значимыми стимулами для изменения метаописания языка часто выступают и не такие влиятельные, на первый взгляд, факторы как лакуарность или мягкость норм, интерференция взаимодействующих языковых систем, обычаи устного модуса речевой практики, законодательные установления и прочие обстоятельства многообразной функциональности любого живого языка. Даже в случае очевидной насущности перемен в грамматической системе языка как феномена зачастую наблюдается неравномерность ее преобразования во временном измерении. Так, в частности, консенсус относительно применимости частеречной категориальности для китайского языка был достигнут значительно позже ее установления для большинства европейских языков, включая русский.

Несмотря на все сопутствующие трудности, анализ возможных новаций в метаописании того или иного языка вполне научно обоснован и более чем уместен в современных условиях тотальной инвентаризации и метаязыковой формализации естественных языков, участвующих в компьютерно-опосредованной коммуникации. При этом компьютеризация языковой практики не только стимулирует высокую степень аутентичности метаязыка, но и предоставляет для этого считавшиеся относительно недавно фантастические статистические возможности (см., напр., Çöltekin and Rama, 2023; Crosthwaite et al., 2023). На фоне экстенсивной квантитативной информации квалитативные обобщения обладают особой убедительностью и достоверностью. Что немаловажно, квантитативно обеспеченная аргументация в равной мере применима как к близкородственным, так и к таким типологически удаленным языкам как

русский и китайский. При сопоставительном анализе разноструктурных языков корпусная лингвистика, в частности, полностью обеспечивает надёжным метаязыковым инструментарием сбор и презентацию первичных данных. Высокая степень унифицированности подходов к аннотированию содержащихся в корпусе текстов благоприятствует референтной обработке языкового материала как «НКРЯ», так и «Онлайн-корпуса». В целом, корпусный подход даёт объективную картину соответствия грамматических категорий реалиям речи, – в том числе, при сравнении русского и китайского языков.

Разметка китайских корпусов принципиально сложнее, чем для многих других европейских языков по двум причинам. Во-первых, в китайском языке нет морфологического словоизменения, поэтому нельзя ориентироваться на окончания или суффиксы. Во-вторых, текст записывается без пробелов, что делает автоматическое определение границ слов нетривиальной задачей, требующей специальных алгоритмов и последующей ручной проверки. Помимо этого, порядок слов в китайском языке несёт значительную грамматическую нагрузку, что также должна учитывать система разметки. Пожалуй, наиболее существенной проблемой корпусной лингвистики китайского языка является специфика автоматической сегментации текста. В отличие от русского, где границы слов графически очевидны, китайский текст записывается без пробелов между словами, что требует применения сложных алгоритмов разбиения на словоформы и последующей тщательной проверки результатов. Морфологическая разметка китайских корпусов опирается на функциональный принцип: частеречная принадлежность единицы определяется не по её форме, а по синтаксической позиции и контекстуальному окружению. Это принципиально отличается от подхода,

применяемого в тех корпусах, где морфологическая аннотация во многом опирается на словоизменительные парадигмы.

«Онлайн-корпус» считается одним из наиболее развитых инструментов китайской корпусной лингвистики. Его особенность – так называемая функциональная аннотация: частеречная принадлежность слова определяется не по его форме, а по роли в конкретном контексте. Это важно, поскольку в китайском языке одно и то же слово может выступать в разных грамматических функциях. Не менее важно и то, что ключевую роль для восприятия смысла в китайском языке играет порядок слов: «Мандарин нелегко классифицировать с точки зрения порядка слов» (Li and Thompson, 1981: 19). Термин «грамматико-графическая аннотация», используемый в описании данного корпуса, обозначает комплексную систему маркировки, объединяющую информацию о графической форме языковых единиц и их грамматических характеристиках. Данная система предусматривает функциональный подход к определению частеречной принадлежности: категория слова устанавливается не на основании его некоего единообразного лексикографического шаблона, а через анализ его дистрибуции, синтаксической роли и морфологических возможностей в конкретном контексте. Такая конфигурация разметки обусловлена тем, что в китайском языке морфологические маркеры, традиционно служащие надёжными индикаторами частеречной принадлежности во флективных языках, отсутствуют или носят факультативный характер. В этих условиях именно функциональный анализ, учитывающий позиционные и комбинаторные свойства единицы, становится основным инструментом её категоризации. Это позволяет исследователю осуществлять многоаспектный поиск, комбинируя критерии различных уровней. Например,

можно извлечь все глаголы движения в прошедшем времени, употребляющиеся в дейктических конструкциях, или проанализировать частотность элементов определённого семантического поля в текстах конкретного жанра. Что важно для достижения цели данного исследования, в «Онлайн-корпусе» можно собирать и анализировать данные о частеречной идентичности языковых единиц с учетом реализации ими своих грамматических возможностей в контекстах.

Возможности «НКРЯ» специалистам известны достаточно хорошо: на сегодня – это, пожалуй, наиболее репрезентативный и снабженный разнообразной разметкой массив текстов. В нем реализована, в числе других, так называемая морфологическая разметка, актуальная для решения задач нашего исследования. Среди широкого спектра параллельных подкорпусов в нем представлен и русско-китайский подкорпус.

Сегодня корпусная репрезентация частей речи становится одним из наиболее информативных ресурсов, позволяя получить аргументированные и объективные представления о языковой системе. Этот подход продуктивен при сопоставлении, в том числе, разноструктурных языков, таких как русский и китайский, где интуитивные оценки могут быть подвержены интерференции культурных и структурных факторов функционирования языков.

Статистика, получаемая в корпусной среде посредством специализированного программного обеспечения объективна, дискретна и масштабируема. Однако ее лингвистическая ценность полностью зависит от качественной интерпретации, и лишь симбиоз квантитативно-кавалитативных обобщений имеет наивысшую научную ценность. Корпусные данные статистически убедительны и верифицируемы, способствуют достоверной лингвистической интерпретации эмпирического материала, обладают метаязыковой совместимостью с

широким кругом языковых метаописаний, универсальны и релевантны для обработки текстов и фрагментов коммуникации разной кодовой природы, применимы в интердисциплинарных исследованиях. Лингвистический потенциал корпусного подхода далек от исчерпания, в том числе, это касается изучения грамматики: «Корпус изначально был задуман как инструмент для изучения грамматики» (Halliday, 1994: 48). Корпусный анализ частеречной системы языка позволяет на высокорепрезентативном материале исследовать специфику современного речевого функционирования и соответствие ему грамматических категорий. Это путь к пониманию закономерностей и динамики языка.

### 5. Выводы

Использование для лингвистического анализа репрезентативных корпусов, таких как «НКРЯ» и «Онлайн-корпус», позволяет выявить функциональные закономерности языка, которые невозможно было бы обнаружить при традиционном подходе. Этот тезис полностью подтвержден результатами проведенного анализа. В данном контексте подтверждено, что частеречная категориальность имеет фундаментальное значение для лингвистики, даже в случае типологической удаленности и разноструктурности описываемых языков. Анализ квалитативных характеристик речи через призму частеречной статистики позволил сделать обоснованные выводы, в частности, о модальности, динамике семантики и особенностях прагматической обусловленности текста. Характерно, что в русском языке большую роль играет морфологическая, а в китайском – синтаксическая адаптивность языковых единиц. В квантитативном аспекте в преимущественно синтетическом русском языке по числу лексем лидируют существительные, а по числу словоупотреблений – местоимения. В аналитическом китайском языке среди языковых единиц – по данным базовой

выборки – также преобладают существительные, но самой востребованной частеречной категорией являются частицы. Вместе с тем, данное соотношение характеризует именно высокочастотный сегмент и не проецируется автоматически на расширенную выборку, в которой представлен более широкий набор частеречных классов. При этом доминирование по числу лексем существительных в обоих рассмотренных языках – русском и китайском – свидетельствует о ключевой роли номинаций в коммуникации. В то же время, различия в частотности других категорий слов (например, местоимений в русском, частиц в китайском языке) указывает на структурные и функциональные особенности глубинного структурирования проанализированных посредством корпусной методики русского и китайского языков. Необходимо отметить, что корпусный анализ продемонстрировал состоятельность предикатива в русском языке, который, хотя и представлен небольшим количеством лексем, имеет высокую частотность словоупотреблений. Аналогично, в китайском языке выявлена метаязыковая значимость такой категории, как дифференцирующее слово. Единицы данных категорий могут быть корректно идентифицированы как элементы определенной совокупности. Их не зависящее друг от друга включение в корпусную аннотацию свидетельствует о том, что частеречные системы являются динамичными и могут инкорпорировать новые разряды.

Квалитативно-квантитативная интерпретация речевой практики, осуществленная благодаря корпусным инструментам практически в режиме реального времени, позволила сделать обоснованные выводы о параметрах возможной модификации частеречных систем русского и китайского языков. В частности, отмечено, что квалитативная

специфика частеречной насыщенности текста определенным образом характеризует его содержательную идентичность, что было подтверждено в аспекте метаописания модальности. Полученные данные квантитативного характера позволили обоснованно проанализировать приоритеты выбора языковых средств для решения коммуникативных задач. Полученные корпусные данные о реальноречевом функционировании единиц того или иного частеречного кластера в разноструктурных языках подтверждают релевантность задействования корпусного инструментария для решения сложных задач. Анализ и интерпретация статистической информации позволяют синтезировать обобщения о такой сложной метаязыковой проблематике, как качество обеспеченности современных языков непротиворечивым грамматическим аппаратом. В частности, предметное исследование частеречных систем русского и китайского языков позволило сделать вывод о насущности выделения в них новых разрядов частей речи и их целесообразном инкорпорировании в лингвистическую практику. Выявленная состоятельность и предикатива, и дифференцирующего слова – как самодостаточных категорий частеречной системы в своих языках – вполне объективна.

Дальнейшие исследования в связи с данной проблематикой перспективны и могут быть реализованы по следующим направлениям. Так, можно расширить эталонный корпус за счёт включения единиц среднечастотного сегмента, что позволит верифицировать выявляемые закономерности на более репрезентативном материале. Актуальным может оказаться жанрово дифференцированный анализ материала: сопоставление частеречного распределения в публицистических, научных, художественных и разговорных текстах способно выявить функционально-

стилистическую специфику обоих языков. Также, полученные данные могут послужить обоснованием практических рекомендаций в области корпусной аннотации – в частности, для уточнения тегсетов, используемых при разметке предикативов в русскоязычных и дифференцирующих слов в китайскоязычных корпусах. Фактически, актуализированный в данном исследовании методологический подход может быть продуктивно развит в широком спектре лингвистической работы и интегрирован в существующие метаязыковые модели. Частеречная категориальность является важным средством идентификации и квалификации языковых единиц в разнообразных метаописаниях, что релевантно, например, для векторного представления языковых данных системами искусственного интеллекта при сентимент-анализе. Прикладная значимость полученных результатов может быть последовательно реализована в лексикографии, лингводидактике, металингвистике и других областях лингвистической практики.

#### Список литературы

- Алпатов В. М. Слово и части речи. М.: Изд. дом ЯСК. 2018. 256 с.
- Баркович А. А., Ван Ц. Лингвистические корпусы китайского языка: функциональный аспект // Вестник МГЛУ. Сер. 1. Филология. 2015. № 5 (78). С. 105–113.
- Баркович А. Метаописание деривационных отношений: специфика системной репрезентации // Mundo Eslavo. 2018. № 17. С. 7–25.
- Виноградов В. В. Русский язык. М.: Рус. яз. 2001. 720 с.
- Захаров В. П., Богданова С. Ю. Корпусная лингвистика. Иркутск: ИГЛУ. 2011. 161 с.
- Культура русской речи: энцикл. слов.-справ. / Под ред. Л. Ю. Иванова. М.: Флинта. 2011. 840 с.
- Ломоносов М. В. Грамматика русского языка, академика М. В. Ломоносова, 1755 года. СПб.: Тип. Имп. Акад. Наук. 1855. 247 с.

Московин В. П. О подходах к определению понятия «троп» // Известия РАН. Серия литературы и языка. 2013. Т. 72, № 2. С. 20–31.

Национальный корпус русского языка и проблемы гуманитарного образования / Отв. ред.-сост. Н. Р. Добрушина. М.: ТЕИС: ГУ-ВШЭ. 2007. 237 с.

Савчук С. О., Архангельский Т. А., Бонч-Осмоловская А. А., Доница О. В., Кузнецова Ю. Н., Ляшевская О. Н., Орехов Б. В., Подрядчикова М. В. Национальный корпус русского языка 2.0: новые возможности и перспективы развития // Вопросы Языкознания. 2024. № 2. С. 7–34.

Сичинава Д. В. Части речи // Материалы к корпусной грамматике русского языка. Выпуск III: Части речи и лексико-грамматические классы. СПб.: Нестор-История. 2018. С. 9–39.

Anderson S. R. A-Morphous Morphology. Cambridge: Cambridge University Press. 1992. 434 p.

Barkovich A. Informational Linguistics: Computer, Internet, Artificial Intelligence and Language // 2019 IEEE 1st International Conference on Artificial Intelligence in Information and Communication (ICAIIIC 2019), Okinawa, Japan, February 11–13. 2019. Pp. 008–013.

<https://doi.org/10.1109/ICAIIIC.2019.8668989>

Biber D. Longman Grammar of Spoken and Written English / Biber D., Johansson S., Leech G., Conrad S., and Finegan E. Harlow: Pearson Education. 1999. 1204 p.

Bybee J. L. Language, Usage and Cognition. Cambridge: Cambridge University Press. 2010. 262 p.

Çöltekin Ç. and Rama T. What do complexity measures measure? Correlating and validating corpus-based measures of morphological complexity. Linguistics Vanguard. 2023. 9(s1). Pp. 27–43.

Croft W. Radical Construction Grammar: Syntactic Theory in Typological Perspective. Oxford: Oxford University Press. 2001. 416 p.

Crosthwaite P. Research trends in corpus linguistics: A bibliometric analysis of two decades of Scopus-indexed corpus linguistics research / Crosthwaite P., Ningrum S., and Schweinberger

M. // International Journal of Corpus Linguistics. 2023. 28 (3). Pp. 344–377.

<https://doi.org/10.1075/ijcl.21072.cro>  
Francis W. N., and Kučera H. Frequency Analysis of English Usage: Lexicon and Grammar. Boston: Houghton Mifflin. 1982. 561 p.

Givón T. Syntax: A Functional-Typological Introduction (Vol. 1). Amsterdam: John Benjamins. 1984. 464 p.

Goldberg A. E. Constructions: A Construction Grammar Approach to Argument Structure. Chicago: University of Chicago Press. 1995. 265 p.

Gries S. T. Quantitative Corpus Linguistics with R: A Practical Introduction. New York: Routledge, 2009. 286 p.

<https://doi.org/10.4324/9781315746210>  
Halliday M. A. K. An Introduction to Functional Grammar (2nd ed.). London: Edward Arnold. 1994. 434 p.

Haspelmath M. The Indeterminacy of Word Segmentation and the Nature of Morphology and Syntax // Folia Linguistica. 2011. 45 (1). Pp. 31–80.

<https://doi.org/10.1515/flin.2011.002>  
Hunston S. Corpora in Applied Linguistics. Cambridge: Cambridge University Press. 2002. 241 p.

Lan, G. Part of speech tagging of grammatical features related to L2 Chinese development: A case analysis of Stanza in the L2 writing context / Lan G., Pan X., Sun Y., and Lu Y. // Frontiers in Psychology, 2023. 14.

<https://doi.org/10.3389/fpsyg.2023.1139703>  
Langacker R. W. Foundations of Cognitive Grammar, Vol. I: Theoretical Prerequisites. Stanford: Stanford University Press, 1987. 540 p.

Levshina N. Corpus-based typology: applications, challenges and some solutions. Linguistic Typology. 2022. 26(1). P. 129–160.

<https://doi.org/10.1515/lingty-2020-0118>  
Li C. N., and Thompson S. A. Mandarin Chinese: A Functional Reference Grammar. Berkeley: University of California Press, 1981. 691 p.

McEnery T., and Hardie A. Corpus Linguistics: Method, Theory and Practice. Cambridge: Cambridge University Press, 2012. 294 p.

Partee B. H. Noun phrase interpretation and type-shifting principles // Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers / Groenendijk J., de Jongh D., and Stokhof M. (Eds.). Dordrecht:

Foris. 1987. P. 115–143.  
<https://doi.org/10.1002/9780470758335.ch15>

Sapir E. Language, an Introduction to the Study of Speech. New York: Harcourt, Brace and World, 1921. 258 p.

Schütze H. Distributional Part-of-Speech Tagging // Proceedings of the Seventh Conference on European Chapter of the Association for Computational Linguistics, 27–31 March 1995. Dublin, Ireland. Pp. 141–148.

Siewierska A. Person. Cambridge: Cambridge University Press, 2004. 327 p.

Sinclair J. Corpus, Concordance, Collocation. Oxford: Oxford University Press. 1991. 170 p.

Tomasello M. Constructing a Language: A Usage-Based Theory of Language Acquisition. Cambridge, MA: Harvard University Press, 2003. 388 p.

Tognini-Bonelli E. Corpus Linguistics at Work. Amsterdam: John Benjamins Publishing Company, 2001. 223 p.

Wang X. Beyond Commonalities: A Quantitative Perspective on Syntactic Features across Mandarin Chinese Varieties / Wang X., Bin S., Chen I., and Hao Y. // Journal of Quantitative Linguistics, 2026. 33(2), P. 81–116.  
<https://doi.org/10.1080/09296174.2025.2567089>

Wang, Z. A review of Chinese sentiment analysis: subjects, methods, and trends / Wang Z., Huang D., Cui J., Zhang X., Ho S.-B., and Cambria E. // Artificial Intelligence Review. 2025. 58, 75. <https://doi.org/10.1007/s10462-024-10988-9>

Xu Z. The Word Status of Chinese Adjective-Noun Combinations // Linguistics. 2018. 56 (1). P. 207–256.  
<https://doi.org/10.1515/ling-2017-0035>

马建忠. 马氏文通. 北京: 商务印书馆. 1983. 447 p. = Ма Ц. Объяснение правил письменного языка господина Ма. Пекин: Шаньфу иньшугуань, 1983. 447 с.

黎锦熙. 新著国文语法. 北京: 商务印书馆. 1993. 336 p. = Ли Ц. Новая грамматика национального языка. Пекин: Шаньфу иньшугуань. 1993. 336 с.

曹杨波.  
计算机辅助语言学习的发展与前景 // 中国信息科技. 2011. № 3. P. 107–109. = Цао Я. Развитие и перспективы обучения языкам с помощью информационных технологий //

Китайские информационные технологии. 2011. № 3. С. 107–109.

张博. 现代汉语同形同音词与多义词.

北京: 北京语言大学对外汉语研究中心. 2002–2003. = Чжан Б. Омонимия и многозначность современного китайского языка. Пекин: Центр изучения китайс. яз. как второго Пекин. ун-та яз. и культуры. 2002–2003. URL: <http://wenku.baidu.com/view/b3eb771afc4ffe473368ab83.html> (дата обращения: 28.08.2025).

代建桃. 现代汉语同音词研究: 论文 ...

汉语言文字学; 四川师范大学, 成都. 2008. 99 p. = Дай Ц. Изучение омонимов современного китайского языка: дис. ... канд. китайской лингвистики; Сычуаньский педагогический университет. Чэнду, 2008. 99 с.

邵敬敏. 现代汉语通论. 上海:

上海教育出版社. 2007. 335 p. = Шао Ц. Общее введение в современный китайский язык. Шанхай: Изд-во образования Шанхая, 2007. 335 с.

#### Материалы исследования

Ахманова О. С. Словарь лингвистических терминов. М.: Сов. энцикл. 1966. 608 с.

Национальный корпус русского языка. URL: <http://www.ruscorpora.ru> (дата обращения: 28.08.2025).

Ожегов С. И., Шведова Н. Ю. Толковый словарь русского языка: 80 000 слов и фразеологических выражений. М.: А ТЕМП. 2006. 944 с.

Розенталь Д. Э., Голуб И. Б., Теленкова М. А. Современный русский язык. М.: Айрис-пресс. 2010. 448 с.

语料库在线 = Онлайн-корпус. URL: <http://www.cncorpus.org/> (дата обращения: 28.08.2025).

孙浩成 现代汉语语法大全. = Сунь Х. Грамматика современного китайского языка. URL: <http://wenku.baidu.com/view/0ea2c833433239-68011c9253.html> (дата обращения: 28.08.2025).

现代汉语词典 /

中国社会科学院语言研究所. 北京: 商务印书馆. 2016. 1799 p. = Словарь современного китайского языка / Ин-т

языкознания Акад. обществ. наук КНР. Пекин: Шаньу иньшугуань, 2016. 1799 с.

中学教学语法系统提要 (试行) / 人民教育出版社中学语文室. 北京:

人民教育出版社. 1984. 21 p. = Учебный план по грамматической системе средней школы (экспериментальный режим) / Отдел языка и литературы для средней школы Народного образовательного издательства. Пекин: Народ. образоват. изд-во. 1984. 21 с. URL: <https://eol.shzu.edu.cn/meol/analytics/> (дата обращения: 28.08.2025).

### References

- Alpatov, V. M. (2018). *Slovo i chasti rechi* [Word and Parts of Speech], Publishing house YASK, Moscow, Russia. (In Russian)
- Barkovich, A. A. and Wang, Q. (2015). Linguistic Corpora of the Chinese Language: Functional Aspect, *Vestnik MGLU. Ser. 1. Filologiya*, 5(78), 105–113. (In Russian)
- Barkovich, A. (2018). Meta-Description of Derivational Relations: Specifics of System Representation, *Mundo Eslavo*, 17, 7–25. (In Russian)
- Vinogradov, V. V. (2001). *Russkiy yazyk* [Russian Language], Rus. Language, Moscow, Russia. (In Russian)
- Zakharov, V. P., Bogdanova, S. Yu. (2011). *Korpusnaya lingvistika* [Corpus Linguistics], IGLU Irkutsk, Russia. (In Russian)
- Ivanov, L. Yu. (ed.) (2011). *Culture of Russian Speech: Encyclopedic Dictionary and Reference Book*, Flinta, Moscow, Russia. (In Russian)
- Lomonosov, M. V. (1855). *Grammatika russkogo yazyka, akademika M. V. Lomonosova, 1755 goda* [Grammar of the Russian language, Academician M. V. Lomonosov, 1755], Imp. Academician Sciences Publ., Saint Petersburg, Russia. (In Russian)
- Moskovin, V. P. (2013). On Approaches to Defining the Concept of "Trope", *Izvestiya RAN. Seriya literatury i yazyka*, 72 (2), 20–31. (In Russian)
- Dobrushina, N. R. (ed.) (2007). *Russian National Corpus and Problems of Humanitarian Education*, TEIS: GU-HSE, Moscow, Russia. (In Russian)
- Savchuk, S. O., Arkhangelsky, T. A., Bonch-Osmolovskaya, A. A., Donina, O. V., Kuznetsova, Yu. N., Lyashevskaya, O. N., Orekhov, B. V., Podryadchikova, M. V. (2024). National Corpus of the Russian Language 2.0: New Possibilities and Development Prospects, *Voprosy Yazykoznaniiya*, 2, 7–34. (In Russian)
- Sichinava, D. V. (2018). Parts of speech, *Materialy k korpusnoy grammatike russkogo yazyka. Vypusk III: Chasti rechi i leksiko-grammaticheskie klassy* [Materials for a Corpus Grammar of the Russian Language. Part III: Parts of Speech and Lexical and Grammatical Classes], St. Petersburg: Nestor-Istoriya, 9–39. (In Russian)
- Anderson, S. R. (1992). *A-Morphous Morphology*, Cambridge University Press, Cambridge, UK. (In English)
- Barkovich, A. (2019). Informational Linguistics: Computer, Internet, Artificial Intelligence and Language, *2019 IEEE 1st International Conference on Artificial Intelligence in Information and Communication (ICAIC 2019)*, Okinawa, Japan, February 11–13, 008–013. <https://doi.org/10.1109/ICAIC.2019.8668989> (In English)
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *Longman Grammar of Spoken and Written English*, Pearson Education, Harlow. (In English)
- Bybee, J. L. (2010). *Language, Usage and Cognition*, Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511750526> (In English)
- Çöltekin, Ç. and Rama, T. (2023). What do complexity measures measure? Correlating and validating corpus-based measures of morphological complexity, *Linguistics Vanguard*, 9(s1), 27–43. <https://doi.org/10.1515/lingvan-2021-0007> (In English)
- Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*, Oxford University Press, Oxford. <https://doi.org/10.1093/acprof:oso/9780198299554.001.0001> (In English)
- Crosthwaite, P., Ningrum, S., and Schweinberger, M. (2023). Research trends in corpus linguistics: A bibliometric analysis of two decades of Scopus-indexed corpus linguistics research, *International Journal of Corpus Linguistics*, 28 (3), 344–377. <https://doi.org/10.1075/ijcl.21072.cro> (In English)
- Francis, W. N. and Kučera, H. (1982). *Frequency Analysis of English Usage: Lexicon*

and Grammar, Houghton Mifflin, Boston. (In English)

Givón, T. (1984). *Syntax: A Functional-Typological Introduction (Vol. 1)*, John Benjamins, Amsterdam. (In English)

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*, University of Chicago Press, Chicago. (In English)

Gries, S. T. (2009). *Quantitative Corpus Linguistics with R: A Practical Introduction*, Routledge, New York. <https://doi.org/10.4324/9781315746210> (In English)

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar (2nd ed.)*, Edward Arnold, London. (In English)

Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax, *Folia Linguistica*, 45(1) 31–80. <https://doi.org/10.1515/flin.2011.002> (In English)

Hunston, S. (2002). *Corpora in Applied Linguistics*, Cambridge University Press, Cambridge. (In English)

Lan, G., Pan, X., Sun, Y. and Lu, Y. (2023). Part of speech tagging of grammatical features related to L2 Chinese development: A case analysis of Stanza in the L2 writing context, *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1139703> (In English)

Langacker, R. W. (1987). *Foundations of Cognitive Grammar, Vol. I: Theoretical Prerequisites*, Stanford University Press, Stanford. (In English)

Levshina, N. (2022). Corpus-based typology: applications, challenges and some solutions, *Linguistic Typology*, 26 (1), 129–160. <https://doi.org/10.1515/lingty-2020-0118> (In English)

Li, C. N. and Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*, University of California Press, Berkeley, CA. (In English)

McEnery, T. and Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*, Cambridge University Press, Cambridge. (In English)

Partee, B. H. (1987). Noun phrase interpretation and type-shifting principles, *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*, Groenendijk J., de Jongh D., and Stokhof M.

(Eds.), Foris, Dordrecht, 115–143. <https://doi.org/10.1002/9780470758335.ch15> (In English)

Sapir, E. (1921). *Language, an Introduction to the Study of Speech*, Harcourt, Brace & World, New York. (In English)

Schütze, H. (1995). Distributional part-of-speech tagging, *Proceedings of the Seventh Conference on European Chapter of the Association for Computational Linguistics*, 27–31 March 1995, Dublin, Ireland, 141–148. (In English)

Siewierska, A. (2004). *Person*, Cambridge, University Press Cambridge. (In English)

Sinclair, J. (1991). *Corpus, Concordance, Collocation*, Oxford University Press, Oxford. (In English)

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*, Harvard University Press, Cambridge, MA. (In English)

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*, John Benjamins, Amsterdam. (In English)

Wang, Z., Huang, D., Cui, J., Zhang, X., Ho, S.-B., and Cambria, E. (2025). A review of Chinese sentiment analysis: subjects, methods, and trends, *Artificial Intelligence Review*, 58, 75. <https://doi.org/10.1007/s10462-024-10988-9> (In English)

Wang, X., Bin, S., Chen, I., and Hao, Y. (2026). Beyond Commonalities: A Quantitative Perspective on Syntactic Features across Mandarin Chinese Varieties, *Journal of Quantitative Linguistics*, 33(2), 81–116. <https://doi.org/10.1080/09296174.2025.2567089> (In English)

Xu, Z. (2018). The word status of Chinese adjective-noun combinations, *Linguistics*, 56(1), 207–256. <https://doi.org/10.1515/ling-2017-0035> (In English)

马建忠. 马氏文通. 北京: 商务印书馆. 1983. 447 p. = Ma, C. (1983). *Explanation of Mr. Mas Written Language Rules*, Shanwu Yinshuguan, Beijing. (In Chinese)

黎锦熙. 新著国文语法. 北京: 商务印书馆. 1993. 336 p. = Li, C. (1993). *New Grammar of the National Language*, Shanwu Yinshuguan, Beijing. (In Chinese)

曹杨波.

计算机辅助语言学习的发展与前景 //

中国信息科技. 2011. № 3. P. 107–109. =  
Cao, Ya. (2011). Development and Prospects of  
Language Teaching with the Help of Information  
Technology, *Kitajskie informacionnye tehnologii*  
[Chinese Information Technology], 3, 107–109.  
(In Chinese)

张博. 现代汉语同形同音词与多义词.

北京: 北京语言大学对外汉语研究中心. 2002–  
2003. = Zhang, B. (2002–2003). *Homonymy and*  
*Polysemy of Modern Chinese*, Center for the  
Study of Chinese as a Second Language, Beijing  
University of Languages and Culture, Beijing.  
[Online], available at:  
[http://wenku.baidu.com/view/b3eb771afc4ffe4733](http://wenku.baidu.com/view/b3eb771afc4ffe473368ab83.html)  
[68ab83.html](http://wenku.baidu.com/view/b3eb771afc4ffe473368ab83.html) (Accessed 28 August 2025). (In  
Chinese)

代建桃. 现代汉语同音词研究: 论文 ...

汉语言文字学; 四川师范大学, 成都. 2008. 99 p.  
= Dai, Q. (2008). Study of homonyms in modern  
Chinese, Ph.D. Thesis, Sichuan Normal  
University. Chengdu, China. (In Chinese)

邵敬敏. 现代汉语通论. 上海:

上海教育出版社. 2007. 335 p. = Shao, C. (2007).  
*General Introduction to Modern Chinese*,  
Shanghai Education Press, Shanghai. (In Chinese)

### Corpus Material

Akhmanova, O. S. (1966). *Slovar*  
*lingvisticheskikh terminov* [Dictionary of  
Linguistic Terms], Sov. Encyclopedia, Moscow,  
Russia. (In Russian)

Russian National Corpus [Online],  
available at: <http://www.ruscorpora.ru> (Accessed  
28 August 2025). (In Russian)

Ozhegov, S. I., Shvedova, N. Yu. (2006).  
*Tolkovy slovar russkogo yazyka: 80 000 slov i*  
*frazologicheskikh vyrazheniy* [Explanatory  
Dictionary of the Russian Language: 80,000  
Words and Phraseological Expressions], TEMP  
Publ., Moscow, Russia. (In Russian)

Rosenthal, D. E., Golub, I. B.,  
Telenkova, M. A. (2010). *Sovremenny russkiy*  
*yazyk* [Modern Russian Language], Iris-press  
Publ., Moscow, Russia. (In Russian)

语料库在线 = Online corpus [Online],  
available at: <http://www.cncorpus.org/> (Accessed  
28 August 2025). (In Chinese)

孙浩成 现代汉语语法大全. = Sun, H.  
*Grammar of Modern Chinese* [Online], available

at: [http://wenku.baidu.com](http://wenku.baidu.com/view/0ea2c833433239-68011c9253.html)  
[/view/0ea2c833433239-68011c9253.html](http://wenku.baidu.com/view/0ea2c833433239-68011c9253.html)  
(Accessed 28 August 2025). (In Chinese)

现代汉语词典 /

中国社会科学院语言研究所. 北京:

商务印书馆. 2016. 1799 p. = Dictionary of  
Modern Chinese / Institute of Linguistics,  
Academy of Social Sciences of the Peoples  
Republic of China (2016), Shanwu Yinshuguan,  
Beijing. (In Chinese)

中学教学语法系统提要 (试行) /

人民教育出版社中学语文室. 北京:

人民教育出版社. 1984. 21 p. = Curriculum for  
the Grammar System of Middle Schools  
(Experimental Mode) / Department of Language  
and Literature for Middle Schools, (1984).  
Peoples Educational Publishing House, Beijing.  
[Online], available at:  
<https://eol.shzu.edu.cn/meol/analytics/> (Accessed  
28 August 2025). (In Chinese)

*Авторы прочитали и одобрили*  
*окончательный вариант рукописи.*

*The 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.*

Александр Аркадьевич Баркович, доктор  
филологических наук, профессор, заведующий  
кафедрой теоретического и славянского  
языкознания, Белорусский государственный  
университет, Минск, Республика Беларусь.

Aliaksandr A. Barkovich, Doctor of Philology,  
Professor, Head of the Department of Theoretical  
and Slavic Linguistics, Belarusian State  
University, Minsk, Republic of Belarus.

Цин Ван, кандидат филологических наук,  
независимый исследователь, Минск,  
Республика Беларусь.

Qing Wang, PhD in Philology, independent  
researcher, Minsk, Republic of Belarus.

УДК 81.25

DOI: 10.18413/2313-8912-2026-12-2-0-3

Ковальчук А. В.<sup>1</sup> 

**Выявление смыслового ядра  
как метод преодоления текстоидности**

<sup>1</sup> ФГАОУ ВО Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет),  
Садовая-Кудринская, 3/1, Москва, 123242, Россия  
E-mail: [kovalchuk\\_a\\_v@staff.sechenov.ru](mailto:kovalchuk_a_v@staff.sechenov.ru)  
ORCID: 0009-0006-1560-1361

*Статья поступила 26 декабря 2025 г.; принята 15 июня 2026 г.;  
опубликована 30 июня 2026 г.*

**Аннотация:** В результатах нейронного машинного перевода, призванных функционировать как текст, однако не являющихся таковым по определению, мы считаем возможным выявить некий коммуникативный центр — смысловое ядро, обладающее текстообразующим потенциалом.

Актуальность данного исследования обусловлена, с одной стороны, общедоступностью программ машинного перевода и активным повсеместным их использованием, с другой — неучетом специфики переводных текстов, порожденных искусственным интеллектом. На практике имеет место систематическое нарушение внутритекстовых связей в результатах машинного перевода — фактически, набор предложений, говоря иначе «текстоид», из которого редактору предстоит воссоздать связный соразмерный текст. Частое обращение к оригиналу позволяет точно устранять в переводе смысловые искажения и неточности, однако в целом перевод продолжает восприниматься как плохо написанный текст с «машинным ДНК». В этом и заключается интересующая нас проблема: отсутствие эффективного способа по оценке и достижению глобальной смысловой соразмерности в переводном продукте ИИ.

Данное исследование имеет своей целью наметить пути для выработки эффективного, практико-ориентированного лингвистического способа по преодолению текстоидности посредством выявления смыслового ядра в результатах машинного перевода. Благодаря комплексному подходу к выбору методов исследования, а именно: абстрагирование, анализ, классификация, синтез, моделирование, измерение, — были достигнуты следующие результаты: (а) предложен способ выявления конфигурации смыслового ядра, основанный на общепринятых в лингвистике понятиях субъекта, предиката, объекта и универсальной предметно-логической типологии семантических отношений, (б) показана необходимость корректировки первоначальной формулировки формулы ядра (в 46 % случаев), (в) определена медиана доли ядра от объема текстоида в жанре новостной заметки медицинской тематики (31 %), (г) предложены основополагающие принципы лингвистической разметки и введены условные обозначения, (д) в иллюстративных целях предложен принцип изображения смыслового ядра в виде графических формул, (е) намечены пути дальнейшего научного поиска.

Вывод: на материале 52 текстоидов показана применимость предложенного нами способа по выявлению смыслового ядра, призванного быть (а) текстообразующим содержательным концентратом, с помощью которого возможно в дальнейшем

преобразовывать текстоид в текст; (б) предметно-логическим ориентиром для контроля и проверки на переводческую адекватность как отдельных мест в машинном переводе, так и отредактированного варианта в целом; (в) инструментом для толкования в текстоиде непонятных, противоречивых мест (в том числе без обращения к тексту оригинала).

**Ключевые слова:** Смысловое ядро; Преодоление текстоидности; Когезия; Нейронный машинный перевод; Изотопность

**Информация для цитирования:** Ковальчук А. В. Выявление смыслового ядра как метод преодоления текстоидности // Научный результат. Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 2. С. 61–83. DOI: 10.18413/2313-8912-2026-12-2-0-3

UDC 81.25

DOI: 10.18413/2313-8912-2026-12-2-0-3

Aleksandr V. Kovalchuk <sup>1</sup> 

Semantic core identification as a method to overcome textoidness

<sup>1</sup> Sechenov First Moscow State Medical University (Sechenov University),  
3/1 Sadovaya-Kudrinskaya St., Moscow, 123242, Russia  
E-mail: [kovalchuk\\_a\\_v@staff.sechenov.ru](mailto:kovalchuk_a_v@staff.sechenov.ru)  
ORCID: 0009-0006-1560-1361

Received 26 December 2025; accepted 15 June 2026; published 30 June 2026

**Abstract:** We believe that neural machine translation results intended to function as a text always have enough potential for a semantic core (*i.e.* a communicative center with text-forming properties) to be found and verbalized.

The *relevance* of this article is provided by two factors. On the one hand, machine translation software is widespread, easily available, and in active use; on the other hand, machine translation results have to be post-edited to the quality of a communicative text due to systematic disruption of its intra-textual connections in the machine translation results which turns out to be, in its raw, non-edited version, a set of separate sentences, in other words – a ‘textoid’ that should be fixed by an editor to function as a coherent text. Although frequent cross-checking between the original text and its translation helps eliminate occasional semantic errors and inaccuracies, the AI output in general still looks like a poor-quality text with a ‘machine DNA.’ This brings us to the core *problem*: now, there is no reliable method to assess and achieve global semantic coherence in AI-generated translations.

That is why our study *aims* to lay the foundations of a linguistic method for overcoming textoid-quality of machine translation results by means of semantic core identification. Through a comprehensive approach that comprises such *methods* as abstraction, analysis, classification, synthesis, modeling, and measurement this study has achieved the following *results*: (a) a unique tool for semantic core identification was proposed relying on such well-known linguistic concepts as subject, predicate, and object, as well as on a basic subject-logical typology of semantic relations; (b) a need to adjust the initial core wording/formula was demonstrated in 46 % of cases; (c) the median core volume (31 %) in a textoid was determined for medical news; (d) basic principles of linguistic annotation (how to label specific linguistic, structural, or semantic features) were proposed as well as a system of notations; (e) a principle for representing the

semantic core by means of graphic formulae was proposed for illustrative purposes; (f) ways for further scientific research were outlined.

**Conclusion:** 52 textoids were analyzed to demonstrate applicability of our method, intended to serve as a reliable linguistic tool for identifying a semantic core which, in its turn, can function as (1) a text-forming essence that can be used in converting a textoid into a text; (2) a subject-logical benchmark for controlling and verifying translation, both for specific segments of the machine translation and for the text as a whole; and (3) a tool for interpreting unclear or contradictory passages within the textoid (without direct need to check up with the source text).

**Key words:** Semantic core; From textoid to text; Cohesion; Neural machine translation; Isotopy

**How to cite:** Kovalchuk, A. V. (2026). Semantic core identification as a method to overcome textoidness, *Research Result. Theoretical and Applied Linguistics*, 12 (2), 61–83. DOI: 10.18413/2313-8912-2026-12-2-0-3

## Введение

Широкая доступность искусственного интеллекта, высокая скорость генерации текстов, в том числе для целей межъязыкового посредничества, и значимое улучшение результатов машинного перевода, отмечаемое многими исследователям (Переходько, 2017; Беляева, 2022; Чакырова, 2013; Панасенков, 2019), — эти и некоторые другие факторы, казалось бы, указывают на то, что текстовые задачи, прежде требовавшие от исполнителя высочайшей квалификации, в наш век цифровых технологий можно практически полностью делегировать машине. «Переводчик и/или редактор, — отмечает В.В. Сдобников, — перестают обращать внимание на внутреннюю целостность и связность текста, его когерентность, его смысловую структуру, фокусируясь на отдельных предложениях и словах. В результате продукт, даже подвергшийся редакторской правке, все равно не превращается в текст, а остается текстоидом» (Сдобников, 2025: 72).

Понятие «текстоид» в Философском словаре определяется как «сетевой текст, лишенный твердой фиксации, свободно меняющий свою форму и контекст, переходящий от пользователя к пользователю...» (Цит. по: Сдобников, 2024). У лингвистов словарно-энциклопедическое толкование для данного понятия на сегодняшний день

отсутствует. Тем не менее в научных статьях текстоид определяется как «незавершенная единица» (Сиротинина, 1994), «особая единица с незамкнутым смысловым контуром» (Боронин, 2016). И.М. Дзялошинский описывает признаки текстоидности следующим образом: «...Произведение состоит из фрагментов, которые написаны разными авторами, <...> не воспринимается как целостность» (Дзялошинский, 2019: 91–105).

Результат машинного перевода<sup>1</sup>, представляющий собой совокупность текстовых фрагментов, автоматически отобранных из массива билингвальных текстов, обработанных и оформленных алгоритмами программы в последовательность предложений, мы вслед за В.В. Сдобниковым (Сдобников, 2025) также будем рассматривать как текстоид. В связи с чем возникает вопрос: каким образом возможно, с одной стороны, обеспечить полученный в результате машинного перевода текстоид внутренней целостностью и связностью, наличествующими в оригинальном тексте, а с другой — как преодолеть воздействие, оказываемое «самой организацией использования систем МП на сознание переводчика и постредактора» (там же: 75)?

<sup>1</sup> Здесь и далее: нейронный машинный перевод.

Сразу отметим, что вряд ли существуют объективные причины (во всяком случае мы их не видим) сомневаться в принципиальной возможности преобразовать текстоид в текст, обеспечить внутритекстовую связность. Весомым аргументом можно считать сам факт существования такого вида профессиональной деятельности, как постмашинное редактирование, или постредактирование, цель которого — если выбран режим максимального качества (full post-editing) — состоит в получении «результата, неотличимого от человеческого перевода» (an output which is indistinguishable from human translation output<sup>2</sup>).

Для ответа на поставленный нами вопрос о способах обеспечения внутритекстовой связности необходимо обратиться к сущностной стороне когезии. Под когезией О.И. Гальперин понимает «особые виды связи, обеспечивающие континуум, т.е. логическую последовательность, (темпоральную и/или пространственную) взаимозависимость отдельных сообщений, фактов, действий и пр.» (Гальперин, 2006: 74). Эксперименты психолингвистов по изучению смыслового восприятия текста как логической последовательности отдельных сообщений подтверждают, что читатель воспринимает текст как единую смысловую структуру; текст в его сознании как бы сжимается: фрагменты объединяются в более крупные смысловые блоки на основе семантических доминант, которые можно представить тезисно, в виде ключевых слов (Григорян, 2024).

Существует множество работ, посвященных изучению проблемы текстовой связности<sup>3</sup>, в том числе

<sup>2</sup> ISO 18587: 2017 (en) Translation Services – Post-Editing of machine translation output – Requirements (2017). P. 8. URL: <https://www.iso.org/standard/62970.html> (дата обращения: 02.10.2025).

<sup>3</sup> Поисковый запрос «когезия» в электронной библиотеке eLibrary в названии публикаций,

объективной возможности выявить в тексте некий смысловой центр, «нечто вроде фокуса, то есть чего-то такого, к чему сходятся все лучи или от чего исходят» (Гольденвейзер, 1922). Так, в докторской диссертации Д.В. Псурцева предложен подход, описывающий механизм смыслоформирования художественного текста (Псурцев, 2001). Французские лингвисты А. Греймас и Ф. Растье стоят у истоков семантической изотопии, в соответствии с которой смысл текста не есть одноплановая линейная последовательность значений: «Под изотопией мы понимаем избыточную совокупность семантических категорий, которая делает возможным целостное прочтение рассказа, прочтение, являющееся результатом отдельных частичных прочтений высказываний и следствием установления, их двусмысленности, при том, что это установление стремится к поискам путей единого прочтения» (Греймас, 1985). Финский семиотик Э. Тараста полагает, что изотопия присуща любому тексту и «обеспечивает связность даже в условиях наивысшей дробности» (Tarasti, 2017). Помимо фундаментальных работ имеются многочисленные научные статьи, посвященные поиску смыслового центра в отдельно взятом художественном произведении (Вишнякова, 2017; Наумчик, 2020; Скращук, 2019). В учебно-методическом пособии Е.В. Гориной подробно рассматривается смысловая структура журналистского текста — каким образом информационный повод развертывается в связный текст (Горина, 2021). Однако материалом, к которому обращаются исследователи в целях комплексного анализа содержательной структуры и выявления смыслового центра, как правило выступают тексты, созданные человеком. Наше исследование, также посвященное проблемам внутритекстовой связности и развертывания текста вокруг

аннотации и ключевых словах дает 1845 научных работ (03.09.2025).

смыслового центра, проведено на материале не текстов в традиционном смысле этого понятия, а текстоидов — в этом его новизна. Кроме того, в ходе наших изысканий предложен и описан способ по выявлению смыслового ядра текста, призванный быть подспорьем в преодолении текстоидности машинного перевода.

К машинному переводу прибегают все чаще, однако, как отмечает А.Н. Малявина, студенты «не видят ошибок в «творении машин» и не привыкли редактировать за ними тексты перевода ни с точки зрения синтаксиса и стиля, ни с позиции проверки фактической информации» (Малявина, 2024); схожие соображения о зачастую необоснованном полном доверии к машинному переводу и неспособности критически подойти к результатам ИИ высказывают Е.Г. Фонова и О.А. Шитц (Фонова, Шитц, 2025).

Не только студенты, но также и профессиональные переводчики и редакторы далеко не всегда предоставляют заказчику коммуникативно полноценный текст (Кобзева, 2018). Таким образом, актуальность выполненного нами исследования, во-первых, связана с ростом интереса к программам машинного перевода и активным их использованием и, во-вторых, определяется потребностью в действенных способах обращения аморфного, расфокусированного текстоида в связный соразмерный текст — как при обучении переводу, так и в отраслевой практике. Данное исследование имеет своей целью наметить пути для выработки лингвистического способа по преодолению текстоидности посредством выявления конфигурации смыслового ядра в результатах машинного перевода, описать его механизм, вскрыть его основные принципы.

### Материал исследования

Источником текстов для настоящего исследования выбран американский медицинский вебсайт [drugs.com](https://www.drugs.com)<sup>4</sup>, в

<sup>4</sup> URL: <http://drugs.com> (дата обращения: 09.07.2025).

частности его раздел *Consumer News*<sup>5</sup>, где ежедневно публикуются актуальные новостные заметки, представляющие собой рерайт отраслевых пресс-релизов, газетных, журнальных, научных статей и прочих информационных жанров. Отметим, что все материалы, размещенные в указанном разделе, в обязательном порядке проходят научную редакцию, осуществляемую медицинским писателем Кармен Поуп, бакалавром фармации. Таким образом, отобранный нами материал можно охарактеризовать как англоязычные информационные тексты медицинской тематики в жанре новостной заметки, для которых верификация фактов предметной области осуществлялась отраслевым экспертом, компетентным также в вопросах языковой и жанровой нормы, что свидетельствует об их коммуникативной полноценности (исходный материал — не текстоиды, а тексты).

Методом сплошной выборки с сайта [drugs.com](https://www.drugs.com) выгружены 52 текста общим объемом 25.470 слов (3,4 а. л.), опубликованные с 1 июля по 8 июля 2025 года<sup>6</sup>. Данный временной интервал обусловлен тем, что к разработке практической части нашего исследования мы приступили 9 июля 2025 года, взяв таким образом для анализа все без исключения новостные заметки, имевшиеся в тот момент на указанном сайте за текущий календарный месяц.

Для получения их русскоязычных версий, на материале которых, собственно, и было проведено исследование, выбран машинный переводчик открытого доступа Google Translate, поскольку именно к его

<sup>5</sup> URL: <https://www.drugs.com/medical-news.html> (дата обращения: 09.07.2025).

<sup>6</sup> Для целей воспроизводимости проведенного нами исследования на случай исчезновения отобранных текстов с сайта [drugs.com](https://www.drugs.com) оригиналы новостных заметок, выгруженные 9 июля, хранятся бессрочно на файлообменнике автора данного исследования по адресу: <https://clck.ru/3NtvrX>.

функционалу, по данным статистики<sup>7</sup>, чаще всего прибегают пользователи интернета. Дата обращения к машинному переводчику для генерации всех 52 текстоидов — 9 июля 2025 года.

### Методы исследования

В основу проведенного нами исследования положен комплексный подход, а именно: для выявления смыслового ядра на фоне текстоидной периферии и изучения его компонентов с их позиционными вариантами задействован метод абстрагирования; для разложения первичного смыслового ядра на компоненты и их позиционные варианты — метод анализа и классификации; для последующей корректировки смыслового ядра с целью выявления его окончательной формулировки-формулы — метод синтеза; для наглядного представления конфигурации смыслового ядра — метод моделирования; для определения основных статистических характеристик смыслового ядра (медиана доли ядра от объема текстоида и частота совпадения первичного и итогового вариантов ядра) — метод измерения. По результатам изучения 52 конфигураций смыслового ядра показана принципиальная возможность нахождения смыслового центра, выявления его конфигурации и потенциала к дальнейшему развертыванию содержательного концентрата в коммуникативно полноценный текст. Таким образом, в данном исследовании вскрываются общие принципы, руководствуясь которыми представляется возможным создавать благоприятные условия для преодоления текстоидности результатов машинного перевода.

### Некоторые теоретические уточнения

Под **смысловым ядром текста** мы, вслед за О.И. Москальской, понимаем

«обобщенный концентрат всего содержания текста» (Москальская, 1981: 17). Смысловое ядро характеризуется наличием как минимум предикативной связи, или предикативности (в нашей номинации — **действие**, или *что происходит в текстовой действительности*). Обязательность, неотъемлемость, непреложность данного ядерного компонента вытекает из самой природы предикативности, которая суть «общее, глобальное свойство всякого высказывания, а также свойство мысли, ее направленность на актуализацию сообщаемого» (Лингвистический энциклопедический словарь, 1990). По мере необходимости в ядре текста могут быть выделены (и, как правило, выделяются) такие компоненты, как **субъект** — *кто или что совершает действие*, **объект** — *на кого нацелено действие*, а также **условие** — *при каких обстоятельствах выполняется действие*. При формулировании опорного высказывания, призванного на роль первичного смыслового ядра текста, мы рекомендуем принимать в расчет одну из характерных черт базисной структуры предложений основного типа, обозначенную Л.М. Ковалевой как «выраженность семантической структуры предложения не только самым эксплицитным, но и самым экономным способом» (Ковалева, 1987).

Для выявления эффективного средства по преодолению текстоидности машинного перевода мы приняли решение фокусироваться только на смысловом ядре: его компонентах и позиционных вариантах, — положив в основу методологии проведенного нами исследования метод абстрагирования. Мы сознательно упрощаем картину изучаемого объекта, фокусируясь на самом главном, поскольку задача по выявлению всех смысловых нитей текста (во всяком случае на данном этапе и в рамках одной научной публикации) была бы чересчур амбициозной, малореализуемой, хотя в теоретическом, сугубо исследовательском плане, безусловно, чрезвычайно интересной.

<sup>7</sup> По данным облачной переводческой платформы Smartling. Электронный ресурс. Режим доступа: <https://www.smartling.com/blog/google-translate-vs-deepl#:~:text=What%20is%20Google%20Translate%20C%20and,are%20often%20lost%20in%20translation> (дата обращения: 15.09.2025).

В целях экономии места и по соображениям удобства обработки массива текстоидов — материалов нашего исследования — для выделенных компонентов ядра смысла вводятся аббревиатуры-индексы с нижним подчеркиванием<sup>8</sup> с обеих сторон: С — субъект, Д — действие, О — объект, У — условие. Отметим, что компонентов в ядре текста может быть несколько; в таком случае к буквенному индексу добавляется цифра, например, У1, У2 или С1, С2, С3 и так далее (при построении формулы, в отличие от формулировки, связи между компонентами возможно изобразить, поэтому необходимость в цифровой индексации отсутствует, см. Таблицу 3). Дальнейшая разработка предложенной нами модели предполагает введение дополнительных индексных элементов.

Представить базовую схему ядра можно в виде цепочки С–Д–О–У. Порядок следования компонентов фиксирован в целях удобства статистической обработки результатов текстового массива (на данный момент — вручную, в перспективе — с использованием автоматизированного программного инструмента, вероятно, с возможностью построения 3-мерных моделей) на следующем основании: при создании формулировки-формулы (далее — вербализации) смыслового ядра мы предлагаем не принимать в расчет тема-рематические отношения между его компонентами, поскольку имеем дело, фактически, с лингвистическим конструктом, не имеющим адресата и не предназначенным для коммуникации как таковой, но существующим

исключительно для внутреннего, служебного пользования.

При выявлении смыслового ядра текста предлагается видеть его центральным (неотъемлемым, обязательным и непреложным) элементом предикативную связь, с которой соотносятся все прочие наличествующие компоненты. Таким образом, в предлагаемой схеме компонент предикативности (действие) — это поливалентный логический центр смыслового притяжения, которому подчиняются все прочие компоненты ядра.

Отметим, что в тексте таких центров может оказаться несколько. В таком случае вербализация смыслового ядра проходит в соответствии со следующими принципами:

- действия соотносятся между собой:  
 $D^1-D^2$ ,

- остальные компоненты либо зависят от одного из действий и в таком случае в индексе имеют соответствующую цифру, например,  $C^1-D^1-O^1-C^2-D^2-U^2$ , либо относятся ко всем действиям и в своей маркировке имеют только базовый буквенный индекс, например  $C-D^1-D^2-O$ , либо одни компоненты относятся ко всем действиям, другие — не ко всем, например  $C-D^1-D^2-O^2-U^2-U$ .

#### Описание проведенного исследования

#### Ход выявления смыслового ядра, его компонентов и их позиционных вариантов

Выявление конфигурации смыслового ядра в целях получения инструмента для последующего создания коммуникативно полноценного текста проводилось для каждого из 52 текстоидов последовательно, в четыре этапа:

1. После ознакомления с результатом машинного перевода было **сформулировано первичное смысловое ядро** — опорное высказывание, которое в сжатой форме призвано выразить то самое главное, о чем сообщается, судя по результатам машинного перевода, в тексте оригинала и/или будет сообщаться в

<sup>8</sup> Нижнее подчеркивание мы предлагаем использовать при разметке текстов как способ указать на начало и окончание индекса, то есть для обозначения его границ в целях облегчения внутритекстового поиска и автоматизированной обработки больших массивов информации. В иных случаях, в частности при вербализации смыслового ядра текста, нижнее подчеркивание представляется избыточным и не используется.

коммуникативно полноценном тексте на переводящем языке<sup>9</sup>. Зачастую приемлемый вариант содержался в заголовке и/или в первом абзаце (« лиде») текстоида, в таком случае он заимствовался без изменений либо с незначительными модификациями в 32 случаях из 52 (62 %). Отметим, что решение о приемлемости сообщения в заголовке/лиде на роль опорного высказывания и о необходимых модификациях принималось нами строго после того, как был прочитан весь текстоид. Такой подход показателен для тех случаев, когда сформулированное нами опорное высказывание не ограничивалось информацией только из заголовка/лида — в 38 % текстоидов.

2. С целью уточнить черновой вариант формулировки-формулы смыслового ядра, исследуемый **текстоид был размечен** — мы отыскивали в нем компоненты смыслового ядра с их позиционными вариантами, маркировав содержащие их слова и словосочетания соответствующим индексом, по необходимости внося коррективы в опорное высказывание. Из 25470 слов всего массива текстоидов на смысловое ядро пришлось 7436 (31 %) — в разделе «Приложение» см. Таблицу 2.

3. Маркированные фрагменты текстоида с помощью метода классификации были **распределены по группам в соответствии с присвоенным индексом** — на этом этапе каждый компонент смыслового ядра становится

как бы центром притяжения своих позиционных вариантов и состоит с ними, по нашим наблюдениям, прежде всего в предметно-логических отношениях тождества (напр., лексические повторы, равновесные синонимы и перифразы) или включения (напр., гипонимы, гиперонимы). Отметим, что проблематика отношений между ядерным компонентом и его позиционными вариантами, а также принципы их выделения требуют дальнейшей разработки.

4. Наконец, каждый перечень позиционных вариантов мысленно **был соотнесен, фактически «синтезирован», со своим ядерным компонентом** с целью его возможной окончательной корректировки. Выполненный на данном этапе позиционно-компонентный синтез необходим для окончательного уточнения конфигурации смыслового ядра, чтобы соотнести финализированную формулировку-формулу с текстоидом и убедиться в ее полной пригодности на роль «обобщенного концентрата». (Итоговый вариант вербализованного смыслового ядра, полученный по результатам 3-го и 4-го этапов, отличался от первичного смыслового ядра 1-го этапа в 46% случаев.)

Так, к примеру, в текстоиде № 9 мы выделили смысловое ядро «Трам\_С\_прекратит финансирование\_Д\_больниц, оказывающих гендерную терапию детям\_О\_», вопреки содержанию заголовка «может сократить финансирование», где выбранный машиной глагол означает «уменьшить в количестве, объеме», тогда как далее по тексту видим позиционные варианты предиката, сводимые к идее именно запрещения (здесь и далее полужирный шрифт наш. — А.К.):  
*...**Может прекратить** федеральное финансирование ...**Изучают возможность заблокировать** федеральное финансирование Medicaid или страхового покрытия ...**Также рассматривают возможность полного исключения** <...> из программы Medicaid ...**Уже***

<sup>9</sup> Несмотря на то, что предлагаемый нами способ призван решать проблемы связности и логической соразмерности (то есть работает на переводческую адекватность), мы не ставим себе задачи предложить инструмент по достижению смыслового и когезийного тождества перевода и оригинала. Наша задача скромнее (но в то же время глобальнее) — дать инструмент по превращению текстоида в текст. Мы надеемся, что наибольшую пользу из наших идей впоследствии извлекут именно практикующие переводчики и редакторы, хотя затронутая проблема принадлежит не столько переводоведению, сколько лингвистике текста.

**прекратило финансирование ...Усилий по ограничению ...Могут ограничивать доступ.**

Поскольку поиск позиционных вариантов для компонентов смыслового ядра предполагает, по сути, построение парадигматического ряда, его элементы в соотношении друг с другом рассматриваются как проявления единого целого; не заметить логическую несоразмерность (в данном случае — смысловую ошибку) в таких условиях невозможно.

Более того, в некоторых случаях традиционное поступательное сопоставление перевода с оригиналом (сверху вниз по тексту, от предложения к предложению) не помогает заметить проблему в силу значительной отдаленности контекстуальных синонимов друг от друга. Так, в текстоиде № 1 в 4-м абзаце сообщается: *...Dozens of genes related to allergies and asthma*, — тогда как в 10-м абзаце уточняется: *...changes in 133 genes related to allergies and asthma*. В машинном переводе *десятки генов* и *133 гена* соответственно. Сохраняя логику автора, которая нам видится в стремлении сообщить примерное число генов, исчисляемое десятками («дюжинами» в оригинале), взамен найденной в текстоиде счетной единицы мы предложили бы вариант: *более сотни генов*. Без намеренного сопоставления позиционных вариантов по тексту заметить подобного рода несоразмерности, думается, можно лишь в силу случайности.

### Результаты

Проведенное в рамках данного исследования изучение принципов выделения смыслового ядра, его конфигураций и основополагающих свойств на материале текстоидных новостных заметок медицинской тематики, полученных в результате обращения к нейронному машинному переводчику Google Translate, позволило получить следующие развернутые результаты:

Во-первых, нами предложен и апробирован оригинальный способ обнаружения конфигурации смыслового ядра, в основу которого поставлены классические категории лингвистики (субъект, объект, предикат) и фундаментальная предметно-логическая типология семантических отношений (отношения тождества: лексические повторы, равновесные синонимы и перифразы, — и отношения включения: гипонимы, гиперонимы). Такой подход позволяет системно и воспроизводимо выделять те элементы текста, которые несут основную смысловую нагрузку, независимо от вариативности поверхностного лексико-грамматического оформления.

Во-вторых, в ходе проверки выявлена статистически значимая необходимость корректировки первоначальной формулировки-формулы ядра (в 46 % случаев), что убедительно указывает на правильность нашего изначального исследовательского стремления избегать жестких, однократно применяемых схем для адекватного описания смыслового устройства текстоида и подтверждает необходимость как минимум двухэтапности в процессе вербализации итогового варианта смыслового ядра.

В-третьих, мы определили медиану доли протяженности смыслового ядра по отношению к общему объему текстоида. Для исследуемого жанра — новостной заметки медицинской тематики — данный показатель составил 31 %. Иными словами, около трети объема текстоида приходится на компоненты ядра и их позиционные варианты, оставшиеся 69 % — на смысловую периферию. Это говорит о том, что, с одной стороны, самые главные (текстообразующие) сведения повторяются многократно, оформлены различными лексико-грамматическими способами и характеризуются заведомо высокой избыточностью; с другой стороны, треть словесного материала может быть свернута в одно-единственное

ключевое высказывание, которое являет собой инструмент по проверке и восстановлению связей как в ядре, так и на периферии.

В-четвертых, с целью обеспечения воспроизводимости результатов и возможности последующего сопоставления данных, полученных по разным текстам различными исследователями, нами разработаны и предложены основополагающие принципы лингвистической разметки и задана система условных обозначений (которые, как отмечено в ограничениях ниже, требуют последующей разработки).

В-пятых, в иллюстративных и дидактических целях, а также для наглядного представления структурных особенностей конфигурации смыслового ядра нами предложен принцип изображения смыслового ядра в виде графических формул. Данный принцип предполагает аббревиатурную индексацию с указанием связей между различными компонентами, что позволяет компактно представить даже самые сложные ядерные конфигурации.

И, наконец, в-шестых, на основе полученных результатов, а также с учетом выявленных ограничений, намечены дальнейшие направления научного поиска:

- Изучение различных конфигураций смыслового ядра, выявление их свойств, создание типологии связей между ядерными компонентами и их позиционными вариантами.

- Определение степени применимости смыслового ядра как инструмента обеспечения внутритекстовой связности и соразмерности в текстоидах различных стилей и жанров.

- Разработка практико-ориентированного варианта предложенного нами способа по выявлению и применению смыслового ядра.

- Выявление в текстоиде всей содержательной структуры, в которую

погружено смысловое ядро, и ранжирование ее по степени релевантности с присвоением соответствующего информационного ранга (к примеру, смысловое ядро — это информация наивысшего ранга, тогда как метаданные (в частности, выходные сведения: автор текста, дата, место публикации) — это информация самого низшего ранга).

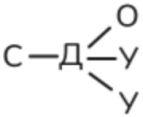
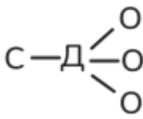
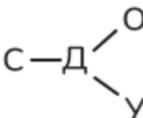
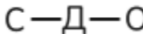
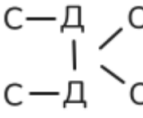
- Автоматизация некоторых действий по вербализации смыслового ядра и всей смысловой структуры текста, возможно, за счет создания компьютерной программы, призванной облегчить сбор и обработку информации и представить ее в наглядном виде. (Важно пояснить: именно «облегчить сбор и обработку информации», а не освободить от когнитивных усилий на этапе осмысления внутритекстовых связей, принципиально не делегируя этот вид работы машине. Осмысление как результат умственной деятельности, присущей только живому человеческому существу, на наш взгляд, и есть то самое фундаментальное непреложное условие для преодоления текстоидности).

Кроме того, полученную для каждого текстоида формулировку-формулу — его смысловое ядро в своем окончательном, уточненном варианте — возможно (но необязательно) оформить в виде **ключевого сообщения**, что мы и сделали для первых десяти текстоидов (см. Таблицу). Под «ключевым сообщением» мы понимаем высказывание, как правило, более развернутое по сравнению со смысловым ядром и, в отличие от него, рассчитанное на потенциального адресата, то есть построенное с учетом общепринятых норм актуального членения предложения. На практике такой коммуникативно полноценный вариант смыслового ядра может быть использован в качестве информационного сообщения (имея в виду под «информационным сообщением» жанр новостной журналистики).

**Таблица.** Ключевые сообщения для первых десяти текстоидов

**Table.** Key Messages for the first 10 textoids

№ текстоид а/ Textoid	Конфигурация смыслового ядра / Semantic Core Configuration		Смысловое ядро в виде ключевого сообщения / Semantic Core in the Form of Key Message
	формула/ Formula	формулировка/ Wording	
1	C—D—O	дым от лесных пожаров <u>C</u> негативно влияет <u>D</u> на иммунную систему <u>O</u>	Дым от лесных пожаров, в котором содержатся мелкодисперсные частицы, химические вещества, содержащие перфторированные аммониевые кислоты (PFAS), токсичные металлы и газы, а также канцерогенные соединения, отрицательно влияет на иммунную систему на клеточном уровне, провоцируя в организме воспаление и приводя к изменениям в 133 генах, связанных с аллергией и астмой.
2	C—D—O	программы Medicaid и Medicare <u>C</u> не покрывают в достаточной мере <u>D</u> лечение наркозависимости <u>O</u>	Программы государственного медицинского страхования Medicaid и Medicare не покрывают в достаточной мере нужды тех, кому требуется лечение наркозависимости и сопутствующих психических расстройств.
3	C—D—O	программа StateViewer <u>C</u> помогает определить <u>D</u> тип деменции <u>O</u>	Программа на основе ИИ StateViewer может сравнивать паттерны активности мозга у конкретного человека с обширной базой подтвержденных случаев когнитивных расстройств и выявлять на ранней стадии с точностью 88% девять различных типов деменции, предоставляя врачам наглядное изображение цветовой карты мозга и соответствующие рекомендации в режиме реального времени.
4	C—D—O	нарушение пищевой безопасности <u>C</u> приводит к отравлениям <u>D</u> патогенными микроорганизмами <u>O</u> на пикниках в преддверии Дня независимости <u>Y</u>	На пикниках в преддверии Дня независимости употребление пищи, которая находилась на открытом воздухе при комнатной температуре дольше двух часов, а в жаркую погоду (>32°) — дольше часа, может привести к серьезным заболеваниям, вызываемым такими бактериями, как сальмонелла, кишечная палочка, листерия, золотистый стафилококк.
5	C—D—O	потеря силы хвата <u>C</u> может быть ранним признаком <u>D</u>	Ухудшение двигательных функций, а именно потеря силы хвата, может быть инструментом раннего скрининга психоза,

		психоза_О_	характеризующегося нарушениями в работе мозга, такими как бред и галлюцинации.
6		число заболеваний кишечника_С_ значительно возросло_Д_ у взрослых_О_ во время пандемии COVID-19_У1_ из-за стресса_У2_	Среди взрослых американцев во время пандемия COVID-19 из-за стресса, вызванного социальным дистанцированием, изоляцией, страхом заражения и изменением образа жизни, частота нарушений пищеварительной системы, таких как синдром раздраженного кишечника и хронический идиопатический запор, увеличилась примерно с 6% в мае 2020 года до примерно 11% в мае 2022 года.
7		Верховный суд_С_ не будет рассматривать_Д_ дело о свободе слова_О1_ группы противников вакцинации_О2_ против Meta_О3_	Верховный суд США отказался рассматривать дело Группы по защите здоровья детей, обвинившей компанию Meta в нарушении 1-й и 5-й поправок к Конституции о свободе слова в связи с тем, что на Facebook были заблокированы ее публикации с 2019 года о вреде вакцинации, а в 2022 году — удалена страница группы.
8		мРНК-1010_С_ эффективнее_Д_ для пожилых людей_О_ по результатам III фазы_У_	В исследовании III фазы вакцина от гриппа мРНК-1010, разработанная компанией Moderna для людей старше 50 лет, по сравнению со стандартной вакциной показала себя эффективнее на 26,6%.
9		Трамп_С_ прекратит финансирование_Д_ больниц, оказывающих гендерную терапию детям_О_	Администрация Трампа, действуя через Центры по предоставлению услуг Medicare и Medicaid, изучает возможность того, как на федеральном уровне прекратить финансовую помощь и страховые выплаты больницам, оказывающим гендерную терапию несовершеннолетним, вплоть до полного исключения этих учреждений из программы Medicaid.
10		общая смертность от ССЗ_С1_ снизилась_Д1_ смертность от других заболеваний сердца_С2_ повысилась_Д2_ у людей после 25 лет_О_ с 1970 по 2022 год_У_	С 1970 по 2022 год в США среди людей старше 25 лет общая смертность от ССЗ снизилась на 66% (с 761 случаев до 258 на 100 000 смертей), в частности: от острого инфаркта миокарда и других ишемических болезней сердца — на 89% и 81% соответственно; при этом смертность от других видов ССЗ увеличилась на 81%, в частности: от сердечной недостаточности, гипертонической болезни сердца и аритмии — на 146%, 106% и 450% соответственно.

### **Заключение. Выводы и ограничения исследования**

Предложенный нами способ позволяет выделить в расфокусированном текстоиде смысловое ядро, которое, как мы полагаем, может претендовать на статус лингвистического инструмента для преодоления коммуникативной неполноценности машинного перевода, обусловленной в первую очередь нарушением внутритекстовых связей и семантической соразмерности. Наша уверенность в его дальнейшей практической пригодности основана на том, что метод смыслового ядра позволяет выявить и вербализовать тот самый центр, где происходит «объединение всех составляющих его [текст — А.К.] предложений вокруг одной темы» (Москальская, 1981: 19), в чем, по утверждению О.И. Москальской, и проявляется смысловая целостность текста (там же).

Итак, по итогам исследования, конфигурация смыслового ядра выявлена для всех 52 текстоидов (в разделе «Приложение» см. Таблицу 3). Это позволило нам убедиться, что во всех случаях в результате целенаправленной умозрительной работы по анализу магистральных внутритекстовых связей с целью осмысления машинного перевода как содержательной целостности и, как следствие, создания условий для преодоления его текстоидности мы способны синтезировать для себя и других такой смысловой концентрат, который может служить: (а) предметно-логическим ориентиром для контроля и проверки смысловой соразмерности отдельных лексико-синтаксических единиц анализа в машинном переводе как относительно друг друга (причем удаленность их контекстуальных проявлений не является при таком подходе фактором, осложняющим редакторскую проверку), так и относительно смысловой структуры текстоида в целом, претендующем на статус текста; (б) инструментом для

толкования в текстоиде непонятных, невразумительных, противоречивых мест при создании равноценной замены тексту оригинала — замены формально близкой либо неблизкой, сообразно стратегии перевода, (в) основанием, на котором развертывается коммуникативно полноценный текст, не обязательно претендующий на статус перевода (в том числе рерайтинг).

Однако следует отметить и некоторые ограничения. Во-первых, наше исследование проводилось на материале новостных заметок — жанр, в котором текст призван сообщить о событии, послужившем информационным поводом, то есть с четко выраженным смысловым центром, на который нанизываются подробности и для которого в принципе характерны многократные повторы. Напротив, жанры более эклектичные, такие как, например, слайды презентации к докладу или оболочка компьютерной программы, предполагают значительно большую имплицитность связей, повышенную самодостаточность и равновесность текстовых элементов. Тем не менее вряд ли имеются основания отрицать принципиальную возможность даже в таких случаях выделить текстообразующий смысловой центр, хотя в силу обозначенных выше причин (имплицитность, самодостаточность, равновесность единиц анализа) такой центр окажется либо слишком громоздким при слишком малой периферии и потому, предположительно, малоэффективным как инструмент по проверке и обеспечению связности и соразмерности частей в составе целого, либо, наоборот, слишком емким и потому недостаточно представленным своими позиционными вариантами в анализируемом материале.

Во-вторых, принципы лингвистической разметки, обозначенные нами лишь в общих чертах, не раскрываются в виде отчетливого, практико-ориентированного алгоритма, что отрицательно сказывается на таких

важных исследовательских характеристиках, как воспроизводимость и методическая ориентированность. Круг вопросов, рассмотренных в данной публикации, вынужденно ограничен общей постановкой проблемы текстоидности как неотъемлемого, неизбежного свойства машинных переводов и обоснованием наличия у семантически расфокусированного продукта ИИ потенциала к выделению переводчиком-редактором тектообразующего логического центра, претендующего на роль инструмента верификации связности и соразмерности частей в составе текста как единого целого. Думается, предложенный нами способ, даже с учетом намеченной на ближайшее будущее его теоретической и методической разработки в целях повышения практической применимости, в известной мере непреодолимо субъективен, поскольку основан на выделении в тексте смыслового центра, то есть фактически его интерпретацию, которая неотделима от таких процедур, как: информационный отбор, расстановка приоритетов, оценка авторского намерения, учет коммуникативной ситуации, — то есть процессов, в той или иной степени субъективных в силу самой природы коммуникативного акта.

### Список литературы

- Беляева Л. Н. Машинный перевод в современной технологии процесса перевода // Известия Российского государственного педагогического университета им. А. И. Герцена, 2022. № 203. URL: <https://doi.org/10.33910/1992-6464-2022-203-22-30> (дата обращения: 02.09.2025).
- Боронин А. А. К вопросу о текстоидах // Вестник Московского государственного областного университета. Серия: Лингвистика, 2016. № 2. URL: <https://doi.org/10.18384/2310-712X-2016-2-26-32> (дата обращения: 02.09.2025).
- Вишнякова А. И. Идея свободы как смысловой центр организации романа Ф.М. Достоевского "Записки из Мертвого Дома" // Наука и образование. 2017. С. 11–15.
- Гальперин И. Р. Текст как объект лингвистического исследования. Москва: ООО "КомКнига", 2006.
- Гольденвейзер А. Б. Вблизи Толстого. М.: Кооперативное издательство, 1922. С. 296.
- Горина Е. В. Смысловая структура журналистского текста: Учебно-методическое пособие. Екатеринбург: Уральский федеральный университет имени первого Президента России Б. Н. Ельцина, 2021. 123 с.
- Греймас А. В. поисках трансформационных моделей. Зарубежные исследования по семиотике фольклора. Москва: "Наука", 1985. С. 89–108.
- Григорян В. А. Концептуализация и интерпретация текста в современной лингвистике // Социально-гуманитарные науки. 2024. С. 513.
- Дзялошинский И. М. Тексты и текстоиды, или что происходит с автором? // PR и СМИ в Казахстане: сборник научных трудов, 2019. № 17. URL: <https://publications.hse.ru/pubs/share/direct/306031376.pdf> (дата обращения: 10.09.2025).
- Кобзева О. В. Нарушение норм литературного языка при переводе (уровень языковой компетенции B1-C1) // Вестник Кемеровского государственного университета. 2018. № 4. URL: <https://doi.org/10.21603/2078-8975-2018-4-211-222> (дата обращения: 03.10.2025).
- Ковалева Л. М. Проблема структурно-семантического анализа простой глагольной конструкции в современном английском языке. Иркутск: Изд-во Иркутского ун-та, 1987. С. 15.
- Малявина А. Н. Обучение постредактированию будущих переводчиков // Актуальные проблемы лингвистики и методики преподавания иностранных языков. 2024. С. 35.
- Москальская О. И. Грамматика текста (пособие по грамматике немецкого языка для ин-тов и фак. иностр. яз.). М.: Высш. школа. 1981. URL: [https://www.phantastike.com/linguistics/grammatika\\_teksta/djvu/view/](https://www.phantastike.com/linguistics/grammatika_teksta/djvu/view/) (дата обращения: 15.09.2025).
- Наумчик О. С. Образ зеркала как смысловой центр сборника рассказов Нила Геймана "Дым и зеркала" // Вестник Балтийского федерального университета им. И. Канта. Серия: Филология, педагогика, психология. 2020. № 1. С. 80–87.

Панасенков Н. А. Опыт обучения студентов-лингвистов постредактированию машинного перевода (на материале англо-русского перевода с помощью систем «Google translate», «Яндекс переводчик» и «Promt») // Педагогическое образование в России, 2019. № 1. URL: <https://doi.org/10.26170/po19-01-08> (дата обращения: 07.09.2025).

Переходько И. В. Оценка качества компьютерного перевода // Вестник Оренбургского государственного университета. 2017. № 2(202). С. 93.

Псурцев Д. В. Смыслоформирующий аспект образно-ассоциативных компонентов художественного текста (на материале англоязычной художественной литературы): Дис. ... к-та филол. наук. М., 2001. 187 с.

Сдобников В. В. Искусственный интеллект в переводе: условия эффективного использования // Научный диалог, 2025. № 3. URL: <https://doi.org/10.24224/2227-1295-2025-14-3-62-80> (дата обращения: 01.09.2025).

Сдобников В. В. Искусственный интеллект в переводе: уточнение понятий // Военно-филологический журнал. 2024. № 4. С. 42.

Сиротинина О. Б. Тексты, текстоиды, дискурсы в зоне разговорной речи // Человек. Текст. Культура. Екатеринбург. 1994. С. 109.

Скращук Е. И. Смысловая доминанта рассказа В. А. Солоухина "подворотня" // Дни науки студентов Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых: Сборник материалов научно-практических конференций. 2019. С. 2311–2319.

Фонова Е. Г., Шитц О. А. К вопросу о профессиональных компетенциях переводчиков в эпоху искусственного интеллекта // Вестник Томского государственного педагогического университета, 2025. № 1(237). URL: <https://doi.org/10.23951/1609-624X-2025-1-148-156> (дата обращения: 01.09.2025).

Чакырова Ю. И. Постредактирование — благодать или проклятие? // Индустрия перевода. 2013. № 1. С. 137.

Tarasti E. The Semiotics of A. J. Greimas: A European Intellectual Heritage Seen from the Inside and the Outside. Sign Systems Studies. 2017. URL: <https://doi.org/10.12697/SSS.2017.45.1-2.03> (дата обращения: 19.05.2026).

## Материалы исследования

Лингвистический энциклопедический словарь. М.: Сов. энциклопедия, 1990. URL: <https://tapemark.narod.ru/les/392d.html> (дата обращения: 03.09.2025).

## References

Belyayeva, L. N. (2022). Machine translation in modern technology of the translation process, *Izvestiya Rossiyskogo gosudarstvennogo pedagogicheskogo universiteta im. A. I. Gertsena*, 203, 22, available at: <https://doi.org/10.33910/1992-6464-2022-203-22-30> (Accessed 02 September 2025).

Boronin, A. A. (2016). On the issue of textoids, *Vestnik Moskovskogo gosudarstvennogo oblastnogo universiteta. Seriya: Lingvistika*, 2, 27, available at: <https://doi.org/10.18384/2310-712X-2016-2-26-32> (Accessed 02 September 2025).

Vishnyakova, A. I. (2017). The Idea of Freedom as the Semantic Center of Dostoevsky's Novel "Notes from the House of the Dead", *Science and Education*, 11–15.

Galperin, I. R. (2006). *Tekst kak object lingvisticheskogo issledovaniya* [Text as an object of a linguistic study], KomKniga, Moscow, Russia.

Goldenveyzer, A. B. (1922). *Vblizi Tolstogo* [Near Tolstoy], Cooperative Publishing House, Moscow, Russia. (In Russian)

Gorina, Ye. V. (2021). *Smyslovaya struktura zhurnalistskogo teksta: uchebno-metodicheskoe posobie* [The Semantic Structure of Journalistic Text: A Textbook], Uralskiy federal'nyy universitet imeni pervogo Prezidenta Rossii B. N. Yeltsina, Yekaterinburg, Russia. (In Russian)

Greimas, A. (1985). *V poiskakh transformatsionnykh modelei. Zarubezhnye issledovaniya po semiotike folkloru* [In Search of Transformational Models: International Research on the Semiotics of Folklore], Nauka, Moscow, Russia. (In Russian)

Grigoryan, V. A. (2024). Textual Conceptualization and Interpretation in Modern Linguistics, *Seventeenth Annual Scientific Conference. Social Sciences and Humanities*, 1, 513.

Dzyaloshinskiy, I. M. (2019). Texts and textoids, or what happens to the author? *PR i SMI v Kazakhstane: sbornik nauchnykh trudov* [PR and Mass Media in Kazakhstan: A Collection of Scientific Papers], Almaty, Kazakhstan, 17, available at: <https://publications.hse.ru/pubs/share/direct/306031376.pdf> (Accessed 10 September 2025).

Kobzeva, O. V. (2018). Violation of Linguistic Norms in Translation in B1-C1 Students, *Vestnik Kemerovskogo gosudarstvennogo universiteta*, 4, available at: <https://doi.org/10.21603/2078-8975-2018-4-211-222> (Accessed 03 October 2025).

Kovaleva, L. M. (1987). *Problema strukturno-semanticheskogo analiza prostoi glagolnoi konstruksii v sovremennom anglijskom yazyke* [The Problem of Structural and Syntactic Analysis of the Simple Verb Constructions in Temporary English], Izd-vo Irkutskogo un-ta, Irkutsk, Russia. (In Russian)

Malyavina, A. N. (2024). Teaching Post-editing to Translation Students, *Aktualnyye problemy lingvistiki i metodiki prepodavaniya inostrannykh yazykov*, 35.

Moskalskaya, O. I. (1981). *Grammatika teksta (posobie po grammatike nemetskogo yazyka dlya institutov i fakultetov inostrannykh yazykov)* [Text Grammar (German Grammar Manual for Institutes and Faculties of Foreign Languages)], Vyssh. shkola, Moscow, Russia, available at: [https://www.phantastike.com/linguistics/grammatika\\_teksta/djvu/view/](https://www.phantastike.com/linguistics/grammatika_teksta/djvu/view/) (Accessed 15 September 2025) (In Russian)

Naumchik, O. S. (2020). The Image of a Mirror as the Semantic Center of Neil Gaiman's Short Story Collection "Smoke and Mirrors", *Vestnik Baltijskogo federalnogo universiteta im. I. Kanta. Seriya: Filologiya, pedagogika, psikhologiya*, 1, 80–87.

Panasenkov, N. A. (2019). Experience of teaching linguistics students - how to post-edit machine-generated translation (based on English-to-Russian translations via Google Translate, Yandex Translate and Promt systems), *Pedagogicheskoye obrazovaniye v Rossii*, 1, available at: <https://doi.org/10.26170/po19-01-08> (Accessed 07 September 2025).

Perekhodko, I. V. (2017). Assessing the quality of computer translation, *Vestnik Orenburgskogo gosudarstvennogo universiteta*, 2 (202), 93.

Psurtsev, D. V. (2001). The Meaning-Forming Aspect of Figurative-Associative Components of a Fiction Text (Based on English-Language Fiction), Ph.D. dissertation, Moscow State Linguistic University, Moscow, Russia.

Sdobnikov, V. V. (2025). AI in Translation: How to Use it Effectively, *Nauchnyi dialog*, 14(3), available at: <https://doi.org/10.24224/2227-1295-2025-14-3-62-80> (Accessed 01 September 2025).

Sdobnikov, V. V. (2024). Artificial Intelligence in Translation: Clarification of Concepts, *Voyenno-filologicheskii zhurnal*, 4, 42.

Sirotnina, O. B. (1994). Texts, textoids, discourses in the zone of colloquial speech, *Chelovek. Tekst. Kultura. Yekaterinburg*, 109.

Skrashchuk, Ye. I. (2019). The Semantic Dominant of V. A. Soloukhin's Short Story "Back Alley", *Dni nauki studentov Vladimirskego gosudarstvennogo universiteta imeni Aleksandra Grigoryevicha i Nikolaya Grigoryevicha Stoletovykh: Sbornik materialov nauchno-prakticheskikh konferentsiy* [Days of Science for Students of Vladimir State University named after A. G. and N. G. Stoletov: Collection of Materials from Scientific and Practical Conferences], Vladimir, Russia 2311–2319.

Fonova, Ye. G., Shitts, O. A. (2025). On the Issue of Professional Competencies of Translators in the Age of Artificial Intelligence, *Vestnik Tomskogo gosudarstvennogo pedagogicheskogo universiteta*, 1(237), available at: <https://doi.org/10.23951/1609-624X-2025-1-148-156> (Accessed 01 September 2025).

Chakyrova, Yu. I. (2013). Is post-editing a blessing or a curse? *Industriya perevoda*, 1, 137.

Tarasti, E. (2017). The Semiotics of A. J. Greimas: A European Intellectual Heritage Seen from the Inside and the Outside. *Sign Systems Studies*, 45(½), available at: <https://doi.org/10.12697/SSS.2017.45.1-2.03> (Accessed 19 May 2026). (In English)

### Corpus Material

Linguistic Encyclopedic Dictionary (1990), available at: <https://tapemark.narod.ru/les/392d.html> (Accessed 03 September 2025).

**Автор прочитал и одобрил окончательный вариант рукописи.**

**The author has read and approved the final manuscript.**

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

**Conflicts of interests: the author has no conflicts of interest to declare.**

**Александр Викторович Ковальчук**, ассистент Центра изучения иностранных языков, ФГАОУ ВО Первый МГМУ им. И. М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия.

**Aleksandr V. Kovalchuk**, Assistant at the Foreign Languages Study Center, Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia.

**Приложение (Appendix)**

**Таблица 2.** Предварительные данные, собранные в ходе выявления смыслового ядра  
**Table 2.** Primarily Data Obtained during Identification of Semantic Core

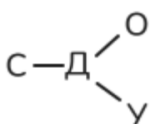
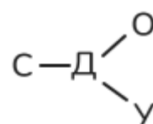
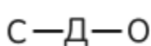
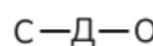
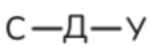
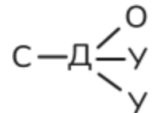
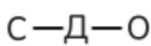
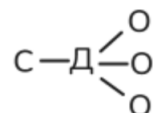
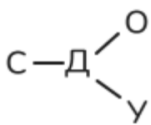
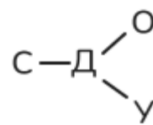
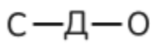
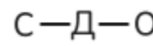
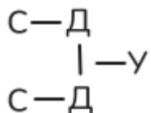
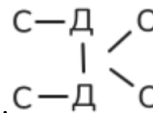
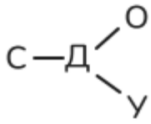
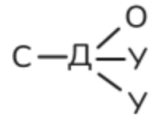
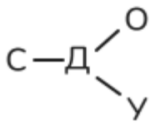
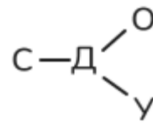
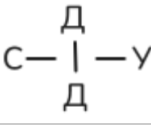
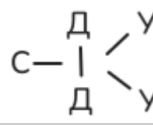
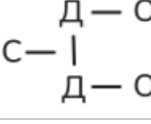
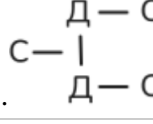
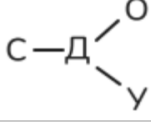
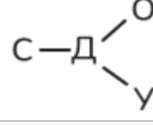
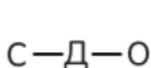
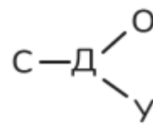
№	Название оригинального текста / Original Text Headline	Кол-во слов текста / Words in Text	Кол-во слов тексто ида / Words in Textoid	Кол-во слов ядра / How many words has the Semantic Core?	
				Всего/ In total	% от доли текстоида / of Textoid (%)
1	Wildfire Smoke Alters Immune System, Study Says	441	450	167	37%
2	Medicaid, Medicare Don't Adequately Cover Addiction Treatment, Study Says	434	429	154	36%
3	AI Can Help Determine A Person's Specific Form Of Dementia	421	396	138	35%
4	Food Safety is Key for Fourth of July Cookouts	567	530	189	36%
5	Grip Provides Insight Into Psychosis, Study Says	555	512	169	33%
6	IBS Rates Nearly Doubled During the Pandemic	322	311	170	55%
7	Supreme Court Won't Hear Anti-Vaccine Group's Free Speech Case	259	242	129	53%
8	Moderna's New Flu Shot Shows Strong Results in Older Adults	310	309	129	42%
9	Trump Administration May Cut Funds to Hospitals Offering Gender Care to Kids	610	580	290	50%
10	1970 to 2022 Saw Decrease in Overall Heart Disease Mortality	293	294	196	67%
11	Livestock Manure Could Be Source Of Antibiotic Resistance, Researchers Warn	816	780	336	43%
12	Brainstorming? Avoid The Internet, Study Says	499	454	156	34%
13	Anger Management Improves With Age In Women, Study Says	502	449	201	45%
14	Tattoos Don't Convey Accurate Impressions Of People, Study Says	486	417	193	46%
15	Tens Of Thousands of Heart Attacks, Strokes Could Be Prevented With Cholesterol Lowering Medications	606	572	198	35%
16	Sleep Apnea Puts Soldiers In Harm's Way	429	422	236	56%
17	Judge Blocks Layoffs at U.S. Health Department	418	406	109	27%
18	Most Dads Take Two Weeks or Less of Parental Leave, Study Finds	340	331	127	38%
19	Could Dairy Be Causing Your Bad Dreams?	372	340	81	24%
20	FDA Will Now Require Labeling of ADHD Medication to Indicate Weight Loss Risk	273	260	141	54%

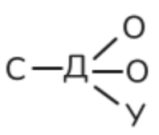
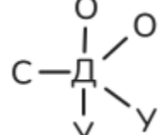
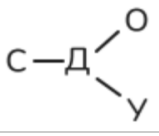
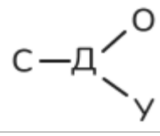
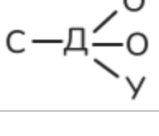
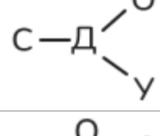
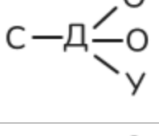
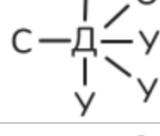
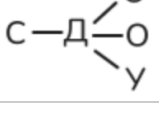
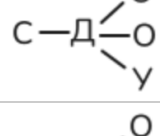
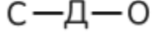
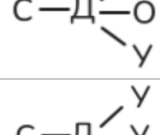
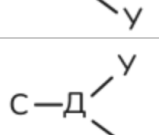
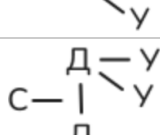
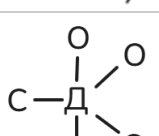
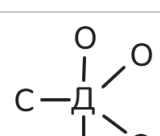
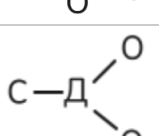
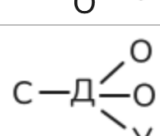
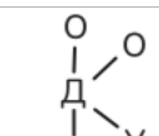
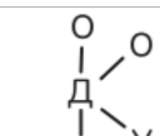
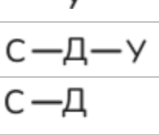
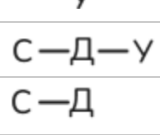
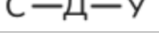
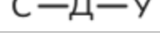


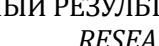
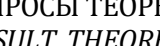
21	<i>FDA Removes REMS Requirements for Embryofetal Toxicity Risk From All Endothelin Receptor Antagonist Medicines</i>	616	582	210	36%
22	<i>Wildfire Smoke Linked To Heart Failure Risk</i>	612	578	154	25%
23	<i>Schools Should Be Prepared For These Three Medical Emergencies</i>	517	512	185	36%
24	<i>Teen Drivers Spend A Fifth Of The Time Looking At Their Smartphone, Study Says</i>	485	425	136	32%
25	<i>Want More Exercise? Go To Bed Earlier, Study Suggests</i>	464	444	119	27%
26	<i>Smartphone-Controlled Nerve Stimulator Returns Golfer To The Links</i>	603	535	158	47%
27	<i>Prevent 4th of July Firework Injuries by Taking These Simple Steps</i>	269	252	52	21%
28	<i>High Levels of Particulate Matter Found Near Electric Vehicle Charging Stations in L.A.</i>	521	517	107	21%
29	<i>Hearing Aids Are a Boon To Social Life, Study Finds</i>	533	502	151	30%
30	<i>Measles Outbreak in Kentucky Confirmed</i>	369	339	55	16%
31	<i>Is One Type of Water Healthier Than Another? Here's What Experts Say</i>	720	659	114	17%
32	<i>Screen Time May Slow Learning Skills, Study Finds</i>	353	335	93	28%
33	<i>Many Older Adults Taking Daily Aspirin Despite Tighter Guidelines</i>	581	526	107	20%
34	<i>Is Radiation Therapy For Cancer a Double-Edged Sword?</i>	590	543	96	18%
35	<i>Type Of Hormone Therapy Can Alter Breast Cancer Risk, Study Says</i>	559	520	153	29%
36	<i>Rheumatoid Arthritis Patients Benefit From Blood Flow-Restricted Strength Training</i>	475	422	99	24%
37	<i>More Than 14 Million Will Die Following U.S. Foreign Aid Cuts</i>	725	666	176	26%
38	<i>Psoriasis 101: Causes, Symptoms and Treatments</i>	878	795	354	45%
39	<i>Oregon Infants Sickened By Lead in Recalled Target Baby Food</i>	316	321	67	21%
40	<i>CDC Now Recommends RSV Vaccine for Some Adults 50+</i>	321	332	63	19%
41	<i>Six Key Traits That Make Someone 'Cool,' Study Finds</i>	640	587	62	11%
42	<i>U.S. Hospitals Face Shortages in Essential Staff Due to Foreign Residency Visa Delays</i>	470	402	77	19%
43	<i>Excess Weight Increases Risk of Health Double-</i>	377	368	119	32%

	<i>Whammy Among Older Women</i>				
44	<i>More Florida Teens Strapping Up, Study Says</i>	423	435	119	27%
45	<i>Common Meds Aren't Linked To Inflammatory Bowel Disease, Study Says</i>	354	333	98	29%
46	<i>Three-Quarters of Stomach Cancers Could Be Prevented By Targeting Helicobacter Pylori Bacteria</i>	551	520	86	17%
47	<i>A Single Disorder Upended Pennsylvania's Medical Marijuana Program</i>	481	480	156	33%
48	<i>Clash Of Drugs Impedes Lung Cancer Treatment, Lowers Survival</i>	528	480	88	18%
49	<i>U.S. Measles Cases Hit Highest Level Since Disease Was Eliminated in 2000</i>	694	662	75	11%
50	<i>Kids' Health in U.S. Has Gotten Worse Over the Past 17 Years, Study Finds</i>	554	503	129	26%
51	<i>These Items Sold at Costco Have Been Recalled — Check Your Home</i>	596	587	177	30%
52	<i>American Academy of Pediatrics Sues HHS Over Vaccine Policy</i>	342	328	92	28%
<i>СУММА ЗНАЧЕНИЙ / TOTAL</i>		25470	23974	7436	
<i>МЕДИАНА / MEDIAN VALUE</i>					30%

**Таблица 3.** Формулы смыслового ядра  
**Table 3.** Semantic Core Formulae

№ текстoid / Textoid	Конфигурация смыслового ядра: формула / Semantic Core Configuration — Formula		Конфигурация смыслового ядра: первичная и итоговая / Semantic Core Configuration — Initial VS Final		
	Первичная/ Initial	Итоговая/ Final	Совпала формулировка / Equal in Wording	Совпала формула / Equal in Formula	Совпали обе / Equal in both
1	С—Д—О	С—Д—О	да	да	да
2	С—Д—О	С—Д—О	да	да	да
3	С—Д—О	С—Д—О	да	да	да

4			нет	да	нет
5			да	да	да
6			нет	нет	нет
7			нет	нет	нет
8			нет	нет	нет
9			да	да	да
10			нет	нет	нет
11			нет	нет	нет
12			да	да	да
13			нет	нет	нет
14			да	да	да
15			да	да	да
16			нет	нет	нет

17			нет	нет	нет
18			да	да	да
19			нет	нет	нет
20			нет	нет	нет
21			да	да	да
22			нет	нет	нет
23			да	да	да
24			нет	нет	нет
25			да	да	да
26			нет	нет	нет
27			да	да	да
28			да	да	да
29			да	да	да
30			да	да	да

31			нет	да	нет
32			да	да	да
33			да	да	да
34			нет	нет	нет
35			нет	нет	нет
36			да	да	да
37			нет	нет	нет
38			нет	нет	нет
39			да	да	да
40			да	да	да
41			да	да	да
42			нет	нет	нет
43			нет	нет	нет
44			да	да	да
45			да	да	да
46			да	да	да

47			нет	нет	нет
48			нет	да	нет
49			да	да	да
50			да	да	да
51			да	да	да
52			нет	да	нет
			да/yes – 28 нет/но – 24	да/yes – 32 нет/но – 20	да/yes – 28 нет/но – 24

УДК 81'42, 004.93

DOI: 10.18413/2313-8912-2026-12-2-0-4

Моисеева И. Ю.<sup>1</sup>  
Релишский А. И.<sup>2</sup>

**Потенциал и ограничения ChatGPT: анализ качества перевода и переводческих ошибок больших языковых моделей**

<sup>1</sup> Оренбургский государственный университет имени В. А. Бондаренко  
просп. Победы, 13, Оренбург, 460018, Россия  
*E-mail:* [romfil@mail.osu.ru](mailto:romfil@mail.osu.ru)  
ORCID: 0000-0001-5991-6986

<sup>2</sup> Оренбургский государственный университет имени В. А. Бондаренко  
просп. Победы, 13, Оренбург, 460018, Россия  
*E-mail:* [romfil@mail.osu.ru](mailto:romfil@mail.osu.ru)  
ORCID: 0009-0008-3791-9180

*Статья поступила 30 марта 2026 г.; принята 15 июня 2026 г.;  
опубликована 30 июня 2026 г.*

**Аннотация:** В условиях быстрого развития технологий обработки естественного языка и увеличения использования автоматизированных инструментов в переводческой деятельности, важно исследовать возможности и ограничения больших языковых моделей, таких как ChatGPT, в переводе текстов различных стилистических форм и тематической направленности. Существует необходимость в сравнительном анализе качества автоматического перевода, выполненного с помощью ChatGPT, особенно в контексте различных типов текстов (специализированные, художественные, научные) и их стилистических особенностей. В данной связи важным является понимание ошибок, связанных с определением контекста и возможными «галлюцинациями» модели.

В представленном исследовании на основании сравнительного анализа выявлены критерии работы ChatGPT, а именно: потенциал в определении контекста, коммуникативной ситуации, решение определенных переводческих задач на уровне стилистических соответствий и лексической эквивалентности. В исследовании представлены результаты оценки качества автоматического перевода как метрикой, основанной на лингвистических параметрах, так и автоматической метрической системой BLEU. Результаты проведенного исследования показывают определенную зависимость качества автоматического перевода от точности формулировки запроса и характера текста. Так, показатели лингвистической метрики оценки качества перевода и показатели системы BLEU свидетельствуют о том, что качество перевода текстов, характеризующихся высоким содержанием специализированной лексики выше, чем качество перевода художественного или научного текста, требующего особого подхода при построении высказываний или выбора терминов в рамках узкой распространенности. Кроме того, результаты также свидетельствуют и о таких свойствах больших языковых моделей генеративного искусственного интеллекта, как ошибочность определения контекста и «галлюцинациях», которые могут ввести в заблуждение и привести к ошибкам при переводе или редактировании текстов. В статье предлагаются необходимые практические приемы, позволяющие

переводчику грамотно использовать технологию нейронных сетей, а именно последние версии инновационного чат-бота ChatGPT в различных аспектах деятельности.

**Ключевые слова:** Теория перевода; Профессиональный перевод; Машинный перевод; Автоматический перевод; Нейронные сети; Искусственный интеллект; ChatGPT

**Информация для цитирования:** Моисеева И. Ю., Релишский А. И. Потенциал и ограничения ChatGPT: анализ качества перевода и переводческих ошибок больших языковых моделей // Научный результат. Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 2. С. 84–110. DOI: 10.18413/2313-8912-2026-12-2-0-4

UDC 81'42, 004.93

DOI: 10.18413/2313-8912-2026-12-2-0-4

Irina Yu. Moiseyeva<sup>1</sup>  
Aleksandr I. Relishsky<sup>2</sup>

**Assessing promise and limitations of ChatGPT:  
analysis of translation quality and translation errors  
of large language models**

<sup>1</sup> V.A. Bondarenko Orenburg State University,  
13 Pobedy Ave., Orenburg, 460018, Russia  
*E-mail:* [romfil@mail.osu.ru](mailto:romfil@mail.osu.ru)  
ORCID: 0000-0001-5991-6986

<sup>2</sup> V.A. Bondarenko Orenburg State University,  
13 Pobedy Ave., Orenburg, 460018, Russia  
*E-mail:* [romfil@mail.osu.ru](mailto:romfil@mail.osu.ru)  
ORCID: 0009-0008-3791-9180

*Received 30 March 2026; accepted 15 June 2026; published 30 June 2026*

**Abstract:** The rapid advancement of natural language processing (NLP) technologies and the increasing integration of automated tools into translation workflows necessitate a comprehensive exploration of the capabilities and constraints of large language models (LLMs), such as ChatGPT, across various stylistic domains and thematic areas. This study addresses the need for a comparative analysis of machine translation quality produced by ChatGPT, specifically focusing on specialized, literary, and scientific discourses. Central to this research is the identification of errors associated with contextual misinterpretation and the phenomenon of model "hallucinations."

Based on a comparative framework, the research delineates the operational parameters of ChatGPT, including its potential for context recognition, communicative situational awareness, and the resolution of specific translation challenges at the levels of stylistic congruence and lexical equivalence. This research employs a hybrid evaluation framework that integrates a linguistically-grounded metric with the automated BLEU scoring system. The findings reveal a significant correlation between translation quality, prompt precision, and the typological characteristics of the source text. Both linguistic and automated metrics indicate that translations of highly specialized technical content exhibit higher accuracy than those of literary or scientific texts, which require nuanced syntactic construction and terminological selection within specific linguistic conventions. Furthermore, the results highlight inherent risks of generative AI, such as

semantic distortions and contextual errors, which can compromise the integrity of translated or edited content. The paper concludes by proposing practical strategies and prompt engineering techniques to enable translators to effectively leverage innovative neural network technologies, particularly the latest iterations of ChatGPT, in professional practice.

**Key words:** Translation theory; Professional translation; Machine translation; Automatic translation; Neural network (machine learning); Artificial intelligence; ChatGPT

**How to cite:** Moiseyeva, I.Yu., Relishsky, A. I. (2026). Assessing promise and limitations of ChatGPT: analysis of translation quality and translation errors of large language models, *Research Result. Theoretical and Applied Linguistics*, 12 (2), 84–110. DOI: 10.18413/2313-8912-2026-12-2-0-4

### Введение (INTRODUCTION)

В современном мире важность перевода и межкультурной коммуникации значительно возросла во многих отраслях человеческой деятельности – от практической внешнеэкономической деятельности и выстраивания международных дипломатических контактов до академической и научной работы. Перевод, и как явление, и как процесс, рассматривался с разных точек зрения. На первых этапах развития теории перевода доминировало представление о разделении перевода на особый вид деятельности, которой присущи определённые закономерности, и на перевод как продукт переводческой деятельности. В дальнейшем лингвисты пришли к однозначному выводу о невозможности подобного искусственного разделения. Детальные исследования функционирования естественного языка позволили судить о том, что, изучая результат перевода можно понять и сам процесс. С этого момента надолго утвердилось представление о процессе перевода как о поэтапном преобразовании языковых структур или единиц, то есть при описании перевода учитывались как исходные структуры или единицы, так и конечные. Такой подход лег в основу многих теорий эквивалентности и адекватности перевода, принципиальной переводимости и непереводимости языковых категорий. Это же представление сохранялось и в период развития

компьютерных технологий, когда появилась возможность создать систему, способную автоматически осуществить перевод на основе заданных правил и определенным образом закодированных языковых структур. Результат машинного перевода долгое время оставался весьма удовлетворительным, пока на смену обычным автоматическим системам не пришли нейронные сети, чей принцип работы основывается на алгоритмах обработки естественного языка. Именно тогда стало ясно, что нейросетевые модели генеративного искусственного интеллекта не опираются на аналитические понятия, разработанные лингвистами, но в процессе обработки сигналов могут формировать перевод, не прибегая к понятиям о структурных компонентах языка, к которым обычно обращались при описании и изучении переводческого процесса. В наше время противоречие этих двух, принципиально разных, подходов к пониманию естественного языка, процесса перевода и, соответственно выстраиванию переводческой деятельности снимается масштабным внедрением искусственного интеллекта, который с одной стороны поэтапно анализирует запросы на естественном языке, а с другой – так же поэтапно генерирует новый контент.

Решения искусственного интеллекта модели генеративных предварительно обученных преобразователей (GPT) OpenAI, известные своими способностями генерировать лингвистические алгоритмы,

стали особенно важным предметом исследований современных лингвистов и разработчиков искусственного интеллекта.

Тем не менее, какими бы многообещающими ни казались модели искусственного интеллекта, всё ещё остается много вопросов: могут ли такие модели, как GPT, действительно улучшить процесс перевода, и если да, то в какой степени; каковы их отличительные способности, и в чем заключаются их ограничения; могут ли они действительно восприниматься как замена или дополнение к переводу профессионального переводчика.

### **Основная часть (MAIN PART):**

#### **Цель работы**

Цель настоящей статьи – провести сравнительно-сопоставительный анализ возможностей ChatGPT, который представляет собой большую языковую модель, состоящую из нейронной сети, применительно к деятельности профессионального переводчика.

К настоящему времени машинный перевод прошел долгий путь развития и теоретического осмысления. Еще в середине XX века израильский исследователь Йеошуа Бар-Илель заложил основы формальной лингвистики для машинного перевода (Bar-Hilel, 1953), а Джорджтаунский эксперимент 1954 года (П. Гарвин, П. Шеридан) продемонстрировал принципиальную возможность компьютерного перевода на примере 60 простых предложений с русского на английский. Вскоре концепция трансформационной порождающей грамматики Н. Хомского и генеративная лингвистика предложили новый аппарат описания формальных языковых структур, где единицей языка становится не фонема или морфема (Jones, 1909; Блумфилд, 1968; Якобсон, Халле, 1962), а предложение, формируемое по формационным и трансформационным правилам (Chomsky, 1980). Как отмечает О. В. Ефимова, для теории перевода того периода характерно приравнение

процесса перевода к межъязыковым трансформациям с опорой на ядерные структуры-эквиваленты (Ефимова, 2017). Параллельно О. Каде (Kade, 1968) разработал трехуровневую коммуникативную модель перевода, а в СССР математическая лингвистика (А. А. Ляпунов, В. Ю. Розенцвейг, А. А. Реформатский, П. С. Кузнецов) изучала язык как абстрактную знаковую систему для создания алгоритмов перевода, одновременно акцентируя семантические связи знаков (Розенцвейг, 1974).

Однако в современной лингвистической индустрии ведущую роль занимает не история, а технологии синхронного анализа и самообучающиеся алгоритмы. Сегодня ключевой сервис Google Translate функционирует на базе методов статистического машинного перевода (SMT), теоретически обоснованных П. Брауном с коллегами (Brown et al., 1993) и применявшихся также в системе PROMPT. Статистический подход позволяет находить соответствия между частотными вариантами на основе анализа параллельных корпусов, но не всегда учитывает контекст и культурные реалии, поэтому результат требует постредактуры (Hutchins, 1999). Именно синхронический срез – изучение текущего состояния машинного перевода, его алгоритмов и ограничений – становится главным фокусом современного этапа развития ИИ в переводческой деятельности.

Важным этапом в эре машинного перевода стал переход в марте 2017 года движка Google Translate от статистического перевода на нейронные сети.

Сейчас служба Google Translate работает с 243 языками. На алгоритмы нейронных сетей переходят также такие службы как DeepL (29 языков) и Яндекс перевод (98 языков).

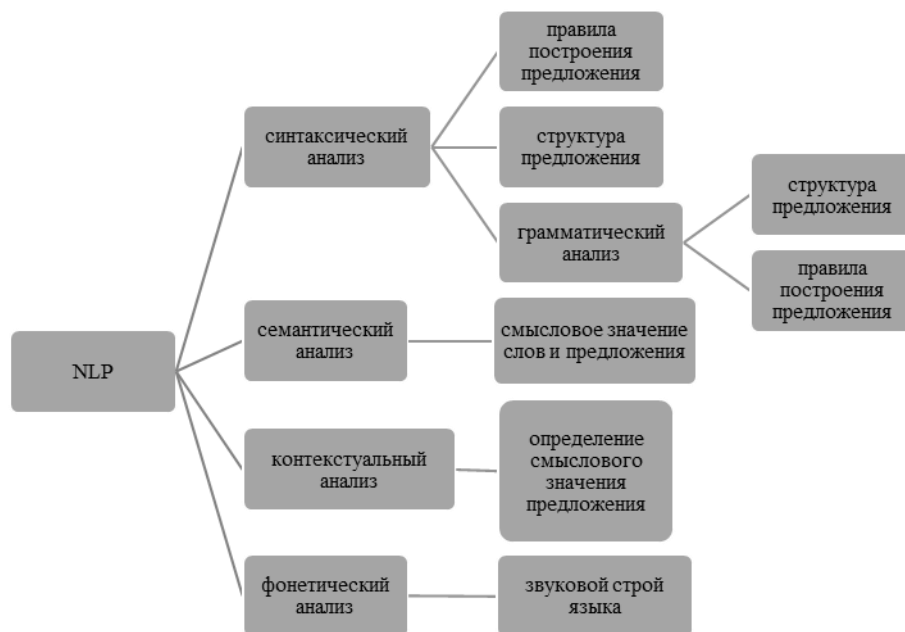
Возможности искусственного интеллекта вышли за пределы отмеченных выше систем и стали использоваться не

только для машинного перевода, требующего редактуры, но и для предпереводческого анализа, в частности, осуществления контекстной работы.

Нейронные сети позволили перейти от метода статистического перевода и обработки естественного языка (Natural Language Processing – NLP), основанного на частотности употребления лексических

единиц и набора грамматических правил, к системам, работающим по принципу работы нейронных сетей головного мозга, где сигналы от рецепторов обрабатываются нейронными кластерами в целом. Упрощенно принцип обработки естественного языка можно представить в следующем виде:

**Рисунок 1.** Принцип обработки естественного языка нейронными сетями  
**Figure 1.** Natural Language Processing in neural networks



Как видно из схемы, обработка естественного языка не опирается на заложенные грамматические правила, но как бы генерирует их. В этом и заключается принципиально новый подход в использовании моделей, управляемых данными, таких как глубокие нейронные сети. Функционирование этих моделей основано на методах машинного обучения на массиве данных, и создания репрезентативной модели естественного языка. В таких системах обработка естественного языка осуществляется путем анализа синтаксиса, семантики и контекста текста (Dyrel, 2023).

С момента выпуска в конце 2022 года ChatGPT (OpenAI 2022), чат-бот, основанный на расширенных языковых

моделях, быстро завоевал популярность благодаря своей впечатляющей производительности в широком спектре задач обработки естественного языка. Эти задачи включают, среди прочего, генерацию текста, ответы на вопросы, классификацию текста и перевод (Wu et al., 2023, Lai et al., 2023). Возможность анализировать контекст и генерировать последовательные, релевантные ответы делает его многообещающим инструментом для различных отраслей, включая переводческий сектор. За короткий период ChatGPT прошел определенный путь эволюционного развития и продолжает совершенствоваться. Следует остановиться подробнее на двух последних версиях чат-

бота, в частности GPT-4 и GPT-4o. Версия GPT-4, интегрирует точную настройку инструкций и обучение с подкреплением с обратной связью от пользователя (Ouyang et al., 2023). Точная настройка инструкций тренирует языковую модель на конкретных примерах, улучшая понимание языковых нюансов и позволяя выдавать более точные и релевантные ответы. ChatGPT продемонстрировал выдающуюся производительность в решении различных переводческих задач, генерации текста и межъязыковой коммуникации (Wang et al., 2023). В течение недели после запуска ChatGPT привлек 100 миллионов пользователей, а уже в марте 2023 года чат был расширен до версии GPT-4 (OpenAI 2023). По данным OpenAI, эта новая модель обеспечивает визуальный ввод и расширенные возможности обработки естественного языка. Это позволяет GPT-4 генерировать более сложные ответы, демонстрируя впечатляющие результаты, в том числе для решения сложных теоретических и практических научных задач. С момента выхода в широкий доступ чат бота, стали появляться исследования направленные на изучение работы ChatGPT в переводческой деятельности. Сиу изучает функционирование языковой модели чата и приходит к выводу о том, что языковая модель искусственного интеллекта революционные возможности «сотрудничества» искусственного интеллекта и переводчика, особенно в практической сфере анализа текстов больших форматов и разной тематики (Siu, 2023). Кэцин Пэн с соавторами с помощью математических методов изучения ответов на запросы доказывает возможность быстрого и качественно перевода документов, в том числе находящихся в оцифрованных архивах, что может открыть новые возможности специалистам в разных различных секторах профессиональной деятельности (Peng et al., 2023). Изучая лингвистические модели искусственного интеллекта и их

архитектуру глубокого обучения, Васвани и др. (Vaswani et al., 2017), а также Ву и др. (Wu et al., 2023) сообщают о значительном прорыве в понимании их потенциала в создании высококачественного, контекстно-релевантного текста, а, следовательно, как отмечается в их работах, модели искусственного интеллекта могут существенно улучшить качество профессионального перевода. В дальнейшем, работы оксфордских исследователей, таких как Бентивольи и др. (Bentivogli, et al., 2018), которые изучают лингвистические возможности моделей GPT, подтвердили эти выводы. Однако имеются исследования, как, например, проект Бьюбек и др. (Bubeck et al., 2023), который показал, что GPT-4, как и его последующие версии, имеет существенные недостатки, которые могут быть присущи его архитектуре прогнозирования следующего слова и генерировании ложного контента. Кроме того, обнаружилось, что GPT-4 страдает от таких проблем, как ненадежность выбранных терминов (Peng et al., 2023), несоответствие стиля и ошибки в определении контекста (Hendy et al., 2023), галлюцинации, ограниченность знаний, чрезмерное категоричная уверенность в неправильных прогнозах и недостаточная самопроверка (Bang et al., 2023). В текстах, где важно сохранение художественных особенностей оригинала, в частности текстов на языках, представленных реже в интернет-пространстве, как, например, арабский или фарси, автоматический перевод версий ChatGPT-4 и -4o, значительно отстает от переводов, выполненных человеком (Al Rousan et al., 2025). Учитывая эти проблемы, очень важно понимать, что использование чат-бота в профессиональной переводческой деятельности может быть довольно ограниченным. В российском сегменте, занимающимся проблемами прикладных задач искусственного интеллекта, заметны исследования К. И. Аветесяна, затронувшего тему методов обнаружения

межъязыковых заимствований в тексте, в том числе и с помощью систем искусственного интеллекта (Аветесян, 2023), С. В. Пальмова и Д. А. Тюриной, изучающих роль нейронных сетей в обработке естественного языка (Тюрина, Пальмов, 2023), а также исследование программы преобразования речи как инструмента письменного перевода А. Р. Биктимирова (Биктимиров, 2023), часто большие языковые модели становятся предметом философского осмысления, что также необходимо в сфере прикладных исследований и прогнозирования вектора научного развития. Так, М. А. Суцин (Суцин, 2024) рассматривает гипотезу расширенного познания А. Кларка и Д. Чалмерса о том, что интеллектуальные артефакты, наподобие записных книжек, без каких-либо сложных интерфейсов (типа интерфейса «компьютер-мозг») при определенных условиях следует рассматривать в качестве внешней когнитивной системы человека, и делает вывод о ее неоднозначности. Вместо этого интеллектуальные артефакты предлагается рассматривать как часть «социокультурных лесов» человеческого познания. В своем исследовании Лю Мяо и Шао Цин разработали интерактивный мультиагентный переводческий фреймворк MAGIC-PTF (Multi-Agent Interactive Chinese Political Translation Framework), в котором сочетаются технологии оптимизации больших языковых моделей и стандартизированные требования к переводу политического дискурса (Мяо, Цин, 2024). Такой фреймворк решает многие проблемы перевода китайского политического дискурса на русский язык. В статье П. А. Ломова предлагается подход к применению технологии RAG (Retrieval-Augmented Generation, RAG), ориентированной на построение контекста для пользовательского запроса к большим языковым моделям на основе онтологии предметной области вместо векторной базы данных. Технология RAG состоит в

предварительном обращении к онтологиям, т.е. к уже представленной, подтвержденной информации в корпусе текстов, что позволяет заметно увеличить степень семантического соответствия запросу пользователя (Ломов, 2025). Интересное предположение высказывает Д. В. Зайцев, рассуждая о причинах отсутствия или недостаточной развитости рассудительной способности больших языковых моделей, и обоснованно связывает способность рассуждать с перцепцией внешнего мира, пока такая полноценная перцепция остается неразвитой способностью рассуждать у интеллектуальных систем остается на сравнительно низком уровне (Зайцев, 2024). В работе с большими объемами текстов, переводчику необходимо, кроме составления рабочего глоссария, выделять определенные фрагменты, систематизировать извлеченную информацию, для этого существуют инструменты эмбендинга. Преимущества эмбендинга раскрываются в статье А. П. Ляха, делающего вывод о том, что эмбендинг можно эффективно использовать, за счет представления слов и предложений в виде векторов, упрощающих обработку данных (Лях, 2024). Соблюдение стиля и жанрово-речевой специфики дискурса выступает одним из ключевых критериев эквивалентности перевода. Однако модели архитектуры GPT не обеспечивают автоматическую реализацию данных параметров без применения уточняющего промптинга (prompt engineering), предварительного обучения или тонкой настройки (fine-tuning). Кроме того, верификация полученных результатов требует экспертной лингвистической интерпретации, что подтверждает необходимость гибридного подхода к оценке качества генеративного перевода. Несмотря на то, что большие языковые модели стали обычным явлением в информационном и технологическом ландшафте, переводческий аспект

использования искусственного интеллекта остается предметом актуального изучения и дискуссий.

Таким образом, понимание актуальности использования больших языковых моделей дает исследователям представление перспективных направлениях, помогающих не только теоретически осмыслить новое явление, но и практически изучать их возможности в разных областях, в том числе и в лингвистической, переводческой работе.

### Материалы и методы исследования (Materials and methods)

Материал исследования представлен репрезентативной выборкой текстов из классической и современной художественной литературы, а также текстов договоров и научно-публицистических материалов на английском языке. Выбор материала обусловлен необходимостью изучить работу GPT на примерах текстов, относящихся к разным эпохам, а также к

разным стилистическим и тематическим формам. Гетерогенный характер эмпирического материала, объединяющего юридические тексты с их прецизионной терминологией и художественные произведения с их метафорической сложностью и культурными реалиями, создает методологическую основу для двустороннего анализа: валидации алгоритмов машинного перевода (на примере ChatGPT) и систематизации профессиональных стратегий человека-переводчика – от формально-эквивалентных решений до креативной адаптации в условиях культурно-специфической семантики. Для получения результатов текст отправляется в виде запросов (request) в ChatGPT обеих последних версий: GPT 3.5 и GPT 4 (рисунок 2).

Общий принцип формирования результатов в чат-боте представлен на рисунке 3.

Рисунок 2. Пример запроса для версии ChatGPT 3.5.

Figure 2. Example of request made for ChatGPT 3.5.

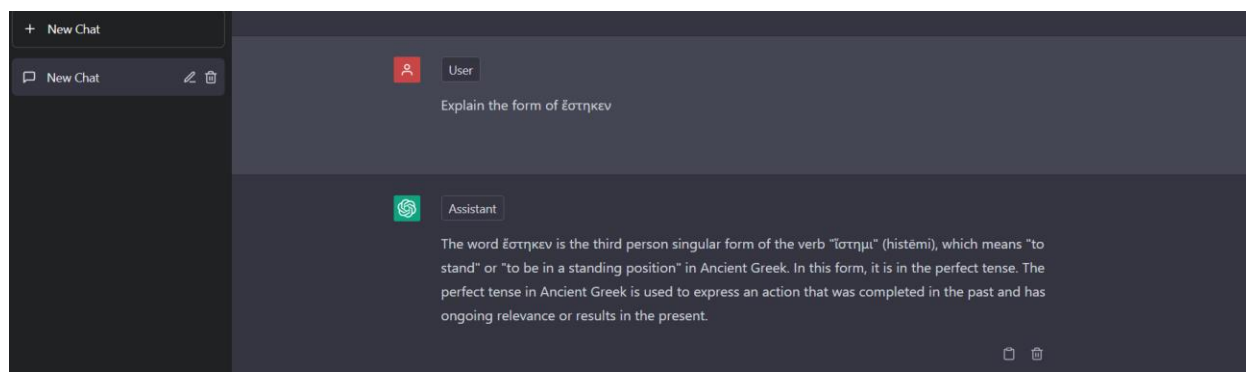
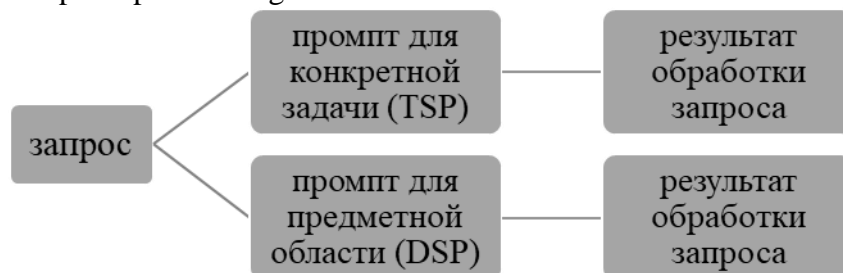


Рисунок 3. Общий принцип работы с чат-ботом

Figure 3. Basic principle of using ChatGPT

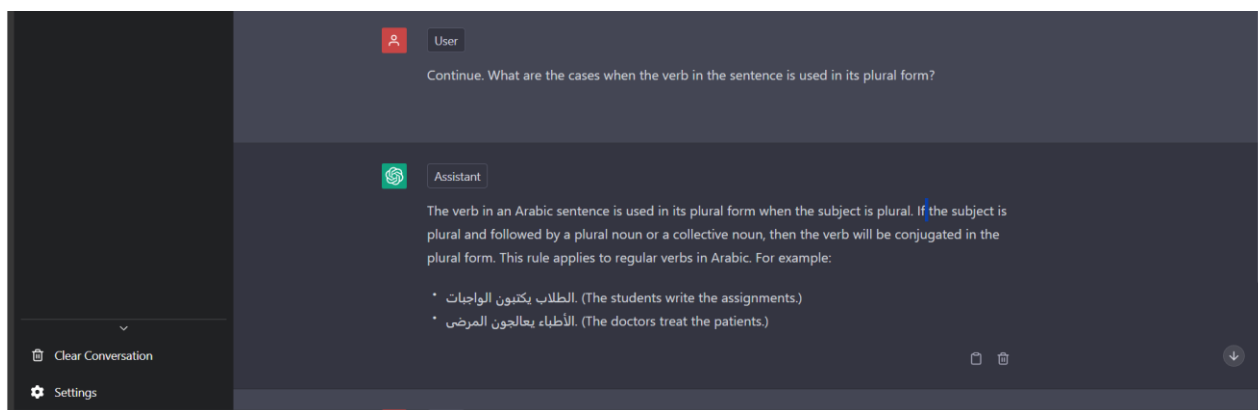


Архитектура лингвистической модели ChatGPT предусматривает использование разного рода промпта (Prompts), в тех случаях, когда результат запроса не удовлетворяет критерии пользователя. Выделяют два основных типа промпта: TSP – *Task Specific Prompts*, то есть промпт для конкретной задачи, и DSP – *Domain Specific Prompts*, то есть

специфические для предметной области запроса.

Кроме того, чат-бот работает с «командами» пользователя, например, чтобы продолжить работу по ранее сделанному запросу можно применить промпт «*продолжай*» («*continue*» или «*keep on*») (рисунок 4).

**Рисунок 4.** Обработка запроса после команды продолжить  
**Figure 4.** Processing the request after the "continue" command



Задачи, поставленные перед чат-ботом в этом исследовании, обобщенно отражают основные этапы работы практического переводчика, а именно:

1. Определение контекстуального фона.
2. Лексический и фразеологический анализ.
3. Перевод.

На этапе работы с контекстом для чат-бота формируются задания для определения контекста. Цель этой стадии исследования заключается в изучении способности чат-бота найти контекст, отражающий коммуникативную ситуацию переводимого текста, а также предлагать отдельные лексические единицы в качестве необходимых эквивалентов. Лексический и фразеологический анализ позволяет вносить определенные «промпты», касающиеся уточнения лексической единиц или фразеологических выражений. На стадии перевода в качестве задания чат-боту предлагается исходный

материал в виде фрагментов текстов, а результатом является его перевод.

Для оценки качества выполненного перевода предлагается следующая метрическая система:

- оценка качества лингвистическими методами, осуществляемая вручную;
- автоматизированная оценка качества.

Основы лингвистической оценки качества машинного перевода были заложены М. Нагао (Nagao, 1989), предложившим пятибалльную шкалу верификации. Согласно данной методике, высший балл (1) присваивается текстам с полной семантической эквивалентностью и нормативным стилистическим оформлением. Промежуточные значения (2–4) отражают градуальное нарастание девиаций: от незначительных лексико-грамматических погрешностей до серьезных нарушений синтаксической структуры, затрудняющих перцепцию. Низший балл (5) верифицирует перевод

как функционально непригодный из-за критического искажения смысла. Следует констатировать, что предложенные критерии носят скорее эмпирико-субъективный характер, нежели строго лингвистический, так как они апеллируют к общему восприятию текста, а не к верифицируемым языковым параметрам. Несколько доработав систему оценки качества перевода, предложенную М. Р. Набабаном (Nababan et al., 2012), в работе выделяются следующие категории для построения алгоритма обработки результатов:

- 1) точность предложенного контекста – степень соответствия коммуникативной ситуации;
- 2) стилистическое соответствие – соответствие стилистических форм в языке перевода соответствующим формам языка оригинала;
- 3) эквивалентность лексики – анализ лексических средств языка перевода;
- 4) отражение коммуникативной цели – стадия оценки общей функциональности текста, сохранение смысла и его общей структуры.

Методика М. Р. Набабана (Nababan et al., 2012) позволяет сформировать комплексный алгоритм обработки результатов, базирующийся на четырех критериях. Прежде всего, оценивается точность контекстуальной реализации, определяющая степень соответствия текста параметрам коммуникативной ситуации. Данный критерий дополняется анализом стилистических форм в языке перевода соответствующих формам языка оригинала. Параллельно проводится верификация лексической эквивалентности для оценки точности выбора терминологических и общеязыковых средств в языке перевода. Завершающим этапом выступает соотнесение коммуникативной цели и функциональной целостности текста, сохранение его инвариантной семантики и структурной архитектоники.

По каждой категории переводу присваивается от 3 (наиболее высокий) до 1 (наиболее низкий). Категории оценки, критерии качества и оценка качества систематизированы в таблице 1.

**Таблица 1.** Система оценки качества перевода  
**Table 1.** Translation quality assessment system

Категория	Критерии качества	Оценка качества
<b>Точность контекста</b>		
точный	Контекст выделен верно. Смыслового искажения не выявлено.	3
менее точный	Контекст в целом определяется верно. Однако наблюдается искажение смысла или возможно двойственное толкование текста.	2
не точный	Контекст определен неверно. Информация не позволяет осуществить на основе контекстного анализа адекватный перевод.	1
<b>Стилистическое соответствие</b>		
точное	Перевод соответствует стилистической форме оригинала. Фигуры речи и стилистические приемы не искажены и переданы средствами языка перевода.	3
менее точное	Общая стилистическая форма соблюдается, однако наблюдается искажение или опущение некоторых фигур речи и стилистических приемов. Нарушена коммуникативная	2

Категория	Критерии качества	Оценка качества
	целостность сообщения.	
не точное	Неверно передан стилистический характер текста оригинала, не отражены стилистические приёмы. Коммуникативная цель не перевода не реализуется.	1
<b>Эквивалентность лексики</b>		
точная	Значение слов, технических терминов, фразеологизмов, предложений или текста на исходном языке точно передается в тексте на языке перевода. Искажения смысла нет.	3
менее точная	Большинство значений слов, технических терминов, фразеологизмов, предложения, предложения или исходный язык тексты точно переданы в языке перевода. Однако наблюдается искажение смысла или возможно двойственное толкование смысла, нарушена коммуникативная целостность сообщения.	2
не точная	Неверно приведено значение лексических единиц, в том числе технических терминов, фразеологизмов, целых предложений. Коммуникативная цель не перевода не реализуется.	1
<b>Реализация коммуникативной цели</b>		
точная	Совокупность средств отражения коммуникативной цели текста на языке оригинала в достаточной мере воспроизводится в переводе. Искажения смыслового содержания нет.	3
менее точная	Совокупность средств отражения коммуникативной цели текста на языке оригинала в неполной мере воспроизводится в переводе. Перевод содержит неточную передачу смыслового содержания.	2
не точная	Совокупность средств отражения коммуникативной цели текста на языке оригинала в недостаточной мере воспроизводится в переводе. Смысловое содержание оригинала искажено.	1

Наряду с экспертной оценкой (human evaluation), для верификации результатов работы нейросетевых архитектур и больших языковых моделей применяются автоматизированные метрики. Данный количественный подход базируется на компаративном анализе исходного текста (source), машинного перевода и эталонного текста, созданного профессиональным переводчиком (human translation), который

в рамках исследования выступает в качестве референса (reference). Использование подобных метрик позволяет объективировать процесс сопоставления выходных данных систем машинного перевода и генеративных моделей, таких как ChatGPT, с принятым лингвистическим стандартом. Основопологающим автоматическим методом оценки машинного перевода

является BLEU – *Bilingual Evaluation Understudy*. Автоматическая метрика BLEU оценивает качество перевода, сравнивая количество совпадений n-грамм – последовательности n-элементов (чаще всего слов), идущих подряд в тексте, сравнивая с референсным, эталонным, переводом, и чем ближе будет показатель машинного перевода к эталонному, тем выше можно судить о качестве машинного перевода (Bang et al., 2023). Метрика хорошо оценивает полноту перевода по количеству элементов текста перевода и оригинала, но не дает полной лингвистической оценки. Исходя из этого, в настоящем исследовании, результаты обрабатываются по двум принципиальным метрикам: ручной – лингвистическому алгоритму и автоматической метрикой BLEU.

### Результаты исследования и их обсуждение (Results and Discussion)

С момента запуска ChatGPT продемонстрировал свой потенциал в различных сферах деятельности, включая профессиональный перевод (Wu et al., 2023, Wang et al., 2023). ChatGPT, хоть и не разрабатывался специально для переводческой отрасли, его способность следовать сложным инструкциям делает его ценным ресурсом для переводчиков. Для оценки эффективности чат-бота используют такое понятие как *температура*. Температура – это определенная степень допуска к творчеству, то есть пользователь может настроить чат с помощью заданий-промптов, чтобы уменьшить или увеличить долю творческой генерации чата при выполнении поставленной задачи (Zhou et al., 2022).

### Работа с контекстом (Contextual Awareness)

Переводчик постоянно сталкивается с необходимостью определения узкого контекста употребления того или иного термина или целого выражения, так как неправильно подобранное эквивалентное значение может значительно изменить

смысл текста или стилистический фон перевода. Сама проблема эквивалентности в переводе состоит из многих аспектов и большинство из них связаны с выбором эквивалента исходя из контекстного окружения, стилистики, лексических и синтаксических особенностей переводимого текста.

Стоит обратить внимание, что специфической функцией ChatGPT является уточнение значения различных текстовых элементов, включая лексических единиц, фразеологических оборотов и предложений, принимая во внимание их контекст внутри самого предложения или целого абзаца. Эта функция особенно полезна при работе с полисемантической лексикой, находящейся, к тому же, в сложном контекстном поле. Такие условия создают сложную задачу переводчику для выполнения полноценного и адекватного перевода. Коммуникация с чатом-GPT ведется на английском языке, так как корпус английского языка шире представлен в цифровом пространстве, следовательно, языковая модель генерирует более релевантную информацию. Для работы с образцами текстов в качестве промпта были выбраны следующие фразы: *Consider the text below, What is the meaning of the expression in the text?* Это промпт для конкретной задачи (TSP), то есть мы задаем направление работы чата для уточнения контекста. Рассмотрим следующие примеры.

#### Текст 1.<sup>1</sup>

Запрос (Prompt): *Consider the text below:*

*The Residence Purchase Price is not subject to determination by reference to areas and shall not be modified other than in accordance with this Clause of this Agreement.*

*What is the meaning of the expression "Residence Purchase Price" in the text?*

<sup>1</sup> Пример взят из договора купли-продажи собственности, составленного в США, 2023 г.

#### ChatGPT-4

*The "Residence Purchase Price" refers to the amount of money that is agreed upon for the purchase of the residence. This price is not subject to determination by reference to areas, unless there is a significant difference in the actual gross liveable area of the residence when built compared to the planned gross liveable area specified in the agreement. If there is a difference of more than five percent, the purchase price may be reduced unless the reduction in gross liveable area was caused by a change requested by the buyer.*

#### ChatGPT-4o

*The "Residence Purchase Price" in the text refers to the agreed upon amount of money that the buyer will pay to the seller for the acquisition of the residence. This price is set forth in the agreement for the purchase of the property and is not to be changed based on the size or area of the residence. It can only be modified according to the specific terms set out in the clause referred to in the text.*

В качестве задания (prompt) чат-боту предлагается определить значение выражения «*Residence Purchase Price*», которое является ключевым понятием в данном фрагменте текста. Версия GPT-4 сообщает определение термина и дополняет его расширенным комментарием, из которого можно узнать, что термин *Residence Purchase Price* употребляется гораздо уже, чем можно предположить на первый взгляд, исходя из его первого эквивалентного значения – *цена покупки жилья*. Также в этой версии предлагается фактор, уточняющий использование термина – *If there is a difference of more than five percent, the purchase price may be reduced unless the reduction in gross liveable area was caused by a change requested by the buyer* – то есть «*цена покупки может быть снижена при пятипроцентной разнице между запланированной и фактической площадью строения, если уменьшение площади потребовал сам покупатель*».

Версия GPT-4o также предлагает краткое определение на контекстном, лексическом и стилистическом уровнях, отвечающее запросу, однако расширение информации, в данном случае, минимальное – уточняется лишь, что «*price is set forth in the agreement for the purchase of the property and is not to be changed based on the size or area of the residence*» («*цена указана в договоре о покупке недвижимости и не подлежит изменению по размеру или площади проживания*»). Однако, даже при минимальном расширении, контекст определяется правильно, и на лексическом уровне соблюдается эквивалентность и реализация семантического поля терминологического оборота.

Таким образом, информация, полученная в ответ на запрос, полностью соответствует контексту и достигает коммуникативной цели текста на языке оригинала и может быть полноценно использована в работе над переводом.

Юридический текст обычно насыщен терминами, и представляет собой довольно жесткую семантическую структуру, но, чтобы проанализировать возможности chatGPT определить контекст в более гибком художественном тексте, где возможны сложные метафоры и отсылки к другим литературным произведениям, рассмотрим фрагмент из художественного произведения, данный в минимальном контекстном окружении:

#### Текст 2.<sup>2</sup>

##### Запрос (Prompt):

*Consider the text below:*

*'Bring your girl along,' said Sparsholt.  
'Oh, well...yes. I'll find out if she's free.'  
I didn't suppose she had a very full diary, but  
I couldn't see her in a pub - unless  
she took it, in her resolute way, as a challenge  
'  
'What's her name, by the way?'  
'She's called Jill.'  
Was there something charitable in his hint of*

<sup>2</sup> Hollinghurst A. *The Sparshot Affair*, Picador, London, 2018, p. 75.

*a smile? 'Ah, that's a nice name'  
"Well' I said. It had always made me uneasy,  
it was too close to chill, and to  
jilt, and not at all far from gill, a quarter –  
pint of cold water.  
What is the meaning of the expression "gill"  
in the text?*

#### ChatGPT-4

*In this context, the expression "gill" refers to a quarter-pint of cold water. The speaker is expressing their uneasiness with the name "Jill" because it is too close to the word "gill" which reminds them of a small quantity of cold water.*

#### ChatGPT-4o

*In the context provided, "gill" refers to a measure of volume in British Imperial units. Specifically, a gill is equivalent to a quarter of a pint. In the United Kingdom, a pint is equal to 20 imperial fluid ounces, so a gill is 5 imperial fluid ounces. The passage plays on a series of words that the name "Jill" phonetically resembles, including "chill," "jilt," and "gill." The narrator is reflecting on the name "Jill" and how it uncomfortably brings to mind these other words with entirely different meanings – in this case, "gill" not only refers to the unit of measurement but also likely connotes something small or insignificant, akin to a small measure of water, emphasizing the narrator's unease with the name.*

Этот фрагмент художественного текста представляет собой диалог, содержащий лексические единицы и обороты, а также метафору, которые могут составить определенную задачу при переводе данного фрагмента. Кроме этого, можно заметить и сложную стилистическую организацию текста на языке оригинала: разговорная речь сменяется рассуждениями героя художественного произведения, где в нескольких строках уместилась определенная палитра ассоциаций, вызванных произнесением имени девушки *Jill*. Комментарий GPT-4 дает очень ограниченную информацию, и, в принципе, делает ошибочный вывод, что

беспокойство героя вызвано созвучием имени со словом «*gill*». При этом GPT-4o останавливается на прием аллитерации, используемый автором, и выделяет фонетическую связь между именем *Jill* и возникающими, неприятными, ассоциациями в сознании героя "*chill*," "*jilt*," and "*gill*". Кроме того, нейронная сеть распознает не только стилистические приемы, но предлагает ещё один комментарий, касающийся лексической единицы *gill* и ее семантической роли в общем контексте, в частности chatGPT поясняет, что *gill* – это не только четверть пинты, но может указывать на что-то очень незначительно: *likely connotes something small or insignificant, akin to a small measure of water, emphasizing the narrator's unease with the name.*

Эти два примера показывают, что chatGPT может ускорить и усовершенствовать процесс перевода, по крайней мере, предложив переводчику контекстные маркеры, существенно облегчающие понимание текста.

Возможности ChatGPT выходят за рамки простого уточнения лексем и словосочетаний, основываясь на их текстуальном контексте. Эта платформа позволяет производить поиск, в результате которого можно получить комментарий, учитывающий культурный и исторический фон текста, что особенно полезно, когда речь идёт о специфических выражениях или культурных реалиях. Эти явления могут представлять определенную сложность для переводчика, носителя другой языковой культуры. Рассмотрим несколько примеров того, как две версии ChatGPT справляются с этой задачей. В качестве первого запроса возьмём название статьи, вышедшей в журнале *Caring for Ages* в августе 2021г.:

#### **Текст 3.<sup>3</sup>**

Запрос:

*Consider the text below:*

<sup>3</sup> Jeanne Manzi, PharmD, BCGP, FASCP and Robert C. Accetta, RPh, C-MTM, BCGP, FASCP, *Caring for the ages.*

*I've fallen, and I can't Get Up: Screening, Prevention, and Impact of Medication*

*What is the meaning of the expression "I've fallen, and I can't Get Up" in the text?*

ChatGPT-4:

*"I've fallen, and I can't Get Up" is a well-known slogan from a commercial for a medical alert system. The expression is often used humorously or ironically to convey a sense of helplessness or vulnerability, often in situations where assistance is needed. In the context of the text you provided, it seems to be used metaphorically in relation to the screening, prevention, and impact of medication, suggesting a need for support or intervention in the face of health challenges.*

ChatGPT-4.0:

*The expression "I've fallen and I can't get up" is a popular catchphrase that originated from a 1980s television commercial for LifeCall, a medical alert system. The phrase was used in the advertisement to demonstrate the purpose of the product, which was to provide help to elderly or disabled individuals who had fallen and were unable to get up without assistance.*

Обе версии генерируют развернутый комментарий, ссылаясь на широко известную фразу *I've fallen and I can't get up*, и если версия GPT-4 определяет не только исторический фон возникновения фразы, но и эмотивный характер использования выражения в запрашиваемом конкретном контексте: *The expression is often used humorously or ironically to convey a sense of helplessness or vulnerability*. Версия GPT-4o такого расширения не представляет и предлагает эту фразу вписать в контекст в ее первоначальном значении, не учитывая коммуникативную ситуацию текста в целом, хотя GPT-4 в дополнение ссылается и на достоверный источник и даже время, когда рассматриваемый оборот вошел в обиход.

В качестве следующего примера возьмем один отрывок, не имеющий в себе расширенного контекста, но содержащий

выражение, которое глубоко уходит корнями в английскую литературу и представляет собой яркую метафору. Определим с помощью ChatGPT значение и контекст выражения.

**Текст 4.<sup>4</sup>**

Запрос:

*Consider the text below:*

*And she was weaned, -*

*I never shall forget it*

-

*Of all the days of the year upon that day,*

*For I had then laid wormwood to my dug*

*What does the expression "For I had then laid wormwood to my dug" mean in the text?*

Chat GPT-4

*This line is from the poem "Sonnets from the Portuguese" by Elizabeth Barrett Browning. In this context, the phrase "For I had then laid wormwood to my dug" refers to the bitterness and pain experienced by the speaker during the process of weaning her child. The phrase "laid wormwood to my dug" is a poetic way of expressing the idea of making something bitter or causing suffering. "Dug" is an old term for a woman's breast, so the line suggests that the speaker experienced a painful or bitter separation from her child during the weaning process.*

Chat GPT-4o

*The phrase "For I had then laid wormwood to my dug" comes from William Shakespeare's play "Romeo and Juliet," spoken by the character Nurse in Act 1, Scene 3. In this context, "dug" is an archaic term for a woman's breast, which in Shakespeare's time might also refer more generally to the nipples.*

*The use of wormwood, a bitter herb, has historical significance. In the context of the play, the Nurse is recounting how she weaned Juliet as a baby. By placing wormwood on her breast ("dug"), she is applying a traditional weaning technique. Wormwood's bitter taste would discourage the child from wanting to nurse, encouraging the child to reject the breast and transition to solid food.*

<sup>4</sup> Shakespear W. *Romeo and Juliet*, Act 1, Scene 3.

*It's a practical memory the Nurse shares to underscore how long she has been with Juliet and how well she knows her, using a personal story to give a sense of Juliet's background and the Nurse's affectionate, if somewhat bawdy, relationship with her charge.*

На этом примере мы сталкиваемся с явлением, которое, в отношении именно искусственного интеллекта получило название «галлюцинации». Дело в том, что версия 4 отсылает нас к реальному литературному произведению – «Сонеты о Португалии» Элизабет Барет Браунинг, но не содержащему, однако, искомую фразу.

**Таблица 2.** Точность контекста  
**Table 2.** Accuracy of the context

Версия ChatGPT	Текст 1	Текст 2	Текст 3	Текст 4	Итого
ChatGPT-4	2	3	2	1	8
ChatGPT-4o	3	3	3	3	11

Из полученных результатов видно, что ChatGPT-4o, в целом получает больше высоких оценок, чем его предыдущая версия. Контекст, определяемый обеими версиями, в основном, не искажает смыслового содержания, а нейронная сеть последовательно и системно предлагает мотивированные контекстные решения. Тем не менее можно отметить, что обе версии допускают ошибочное толкование, и тем самым могут затруднить предпереводческий, контекстный, анализ.

#### **Перевод текстов (Texts translation)**

Еще одной особенностью ChatGPT, как показывают исследования (Hendy et al., 2023, Wu et al., 2023, Wang et al. 2023), является его способность автоматически генерировать переводы текстов. Переводы, выполненные генеративной моделью, подобно переводам, предоставляемым традиционными системами машинного перевода, можно рассматривать как черновые варианты перевода (Draft translation), которые предоставляют пользователям полезную отправную точку для перевода текста. Например,

Тем не менее контекст выявлен верно, раскрыта метафора и коммуникативный эффект определенным образом достигается. Версия 4o очень точно определяет источник фразы, раскрывает контекст и комментирует мотивированность метафоры: «*It's a practical memory the Nurse shares to underscore how long she has been with Juliet and how well she knows her*».

Следуя лингвистической метрике, можно оценить качество контекста, представленного нейронной сетью GPT (табл. 2).

пользователь может ввести предложение или абзац для перевода на другой язык, используя такие ориентированные на задачу промпты, как «Переведите следующий текст на [целевой язык]». Затем ChatGPT генерирует черновик или даже предложит несколько вариантов одного и того же перевода в соответствии с инструкциями, приведенными в промпте. Такие машинно-генерируемые варианты могут помочь переводчикам в их работе над текстами, позволяют им уточнять и улучшать черновой вариант, тем гарантировать точную передачу плана исходного текста в языке перевода.

Рассмотрим примеры сгенерированного перевода.

#### **Текст 1.<sup>5</sup>**

#### **Промпт:**

*Translate the text below from English into Russian:*

<sup>5</sup> Caring for the Ages. August-September 2021. The Society For Post-Acute and Long Term Care Medicine. P. 8-9.

*Falls are a serious problem that could quickly move an older adult from an assisted living community into a skilled nursing facility. Prevention of falls that contribute to disability in older adults is a high priority for maintaining their independence and quality of life. Ensuring the highest level of functionality for older adults is an important element in preventing disability for as long as possible. Multiple studies have identified factors that increase the risk for falls in older adults, and medications are associated with this increased risk due to causes including sedation, drops in blood pressure, drug interactions, and polypharmacy.*

**Reference:** Риск внезапных падений – серьезная проблема, которая может привести к тому, что пожилого человека придется перевести из дома престарелых в учреждение квалифицированного сестринского ухода. Профилактика падений, приводящих к инвалидности, является первоочередной задачей для поддержания их качества жизни и возможности самообслуживания. Обеспечение наиболее высокого уровня функциональности для пожилых людей является важным элементом предотвращения инвалидности на достаточно длительный срок. Многочисленные исследования показали факторы, которые увеличивают риск внезапных падений у пожилых людей, среди этих факторов лекарственные препараты, увеличивающие риски из-за своего седативного эффекта, понижения артериального давления, совместимости с другими лекарственными препаратами и полипрагмазия.

#### **ChatGPT-4**

*Падения представляют собой серьезную проблему, которая быстро может перевести пожилого человека из общины с уходом в отведенное место проживания в учреждение по уходу высокой квалификации. Предотвращение падений, способствующих инвалидности у пожилых людей, является высокоприоритетной задачей для*

*поддержания их независимости и качества жизни. Обеспечение максимального уровня функциональности для пожилых людей – важный элемент в предотвращении инвалидности насколько это возможно. Множество исследований выявило факторы, увеличивающие риск падений у пожилых людей, и медикаменты связаны с этим повышенным риском из-за таких причин, как седация, снижение артериального давления, взаимодействие с другими лекарствами и полифармакология.*

#### **ChatGPT-4o:**

*Падения – серьезная проблема, которая может быстро перевести пожилого человека из дома престарелых в учреждение квалифицированного сестринского ухода. Предотвращение падений, которые приводят к инвалидности пожилых людей, является первоочередной задачей для поддержания их независимости и качества жизни. Обеспечение наивысшего уровня функциональности для пожилых людей является важным элементом предотвращения инвалидности на как можно более длительный срок. Многочисленные исследования выявили факторы, которые увеличивают риск падений у пожилых людей, и лекарства связаны с этим повышенным риском по причинам, включая седативный эффект, падение артериального давления, взаимодействие лекарств и полипрагмазию.*

Результаты автоматического перевода чатом-GPT представляют собой переводы, которые можно использовать для первичного ознакомления с текстом оригинала, но, если сравнивать с референсным (эталонным) переводом, выполненным человеком, можно выделить ряд характерных отличий. Обращает внимание дословность перевода, сохранение синтаксических последовательностей, что затрудняет его стилистическое восприятие, а иногда и понимание смысла. Так, для *falls*

автоматический перевод предлагает прямой эквивалент «падения», в то время, когда профессиональный переводчик применяет комплексный прием перевода в виде лексического дополнения, который в данном случае наиболее мотивирован и максимально передает спектр имплицитных семантических компонентов лексической единицы языка оригинала. Понятие *independence* в автоматическом переводе выражено эквивалентом «независимость», этот эквивалент сохраняет общее семантическое единство с оригиналом, при том, что стилистически этот термин не употребляется в русскоязычном пространстве в контексте переводимого текста, и не отражает функциональности самого понятия, поэтому в эталонном переводе переводчик развивает смысл понятия *independence* выражением «возможность самообслуживания», которое как нельзя лучше стилистически и семантически соответствует переводимому тексту. Подобную ситуацию можно проследить в переводе понятия *drug interactions*. Варианты автоматического перевода (взаимодействие с другими лекарственными препаратами – chatGPT-4, взаимодействие лекарств – chatGPT-4o) в данном случае тоже отражают эквивалентное значение оригинала, однако не соответствует принятой терминологии медицинских текстов: «совместимость лекарственных препаратов».

Контекстное соответствие определено в обоих случаях верно и не регистрируются нарушения коммуникативных задач.

## Текст 2.

Translate the text below into Russian: *In Cypriot domestic tax law, the arm's length principle is enshrined in Section 33 of the Cypriot Income Tax Law 118(I) of 2002 as amended ("ITL") which, in brief, allows reported earnings to be adjusted if the prices differ from those which would have applied between two independent enterprises for comparable transactions in comparable*

*circumstances. Section 33 adopts the arm's length principle as found in Article 9 of the OECD Model Tax Convention on Income and on Capital which constitutes the international standard by OECD members, and which must be used to establish transfer pricing between related companies. Cyprus has c.60 tax treaties in place, in which it generally uses the OECD Model as basis, and generally observes OECD transfer pricing guidance.*

**Reference:** В налоговом законодательстве Кипра принцип «вытянутой руки» (*the arm's length principle*) закреплен в разделе 33 Законов Кипра о подоходном налоге 118(I) от 2002 года, с последующими поправками, который позволяет корректировать заявленные доходы, если цены заметно отличаются от тех, которые компании устанавливают для совершаемых сделок при равных условиях. В разделе 33 принцип «вытянутой руки» принимается в соответствии со статьей 9 Конвенции ОЭСР о подоходном налогообложении и капитале, которая утверждает международный стандарт для членов ОЭСР, и которая применима в случаях установления трансфертного ценообразования между связанными компаниями. Кипр подписал около 60 налоговых соглашений, в которых применяется налоговая модель, предложенная ОЭСР, и в целом придерживается рекомендаций относительно трансфертного ценообразования.

**ChatGPT-4:** Во внутреннем законодательстве Кипра о налогах на прибыль внутри страны принцип *arm's length* действительно закреплен в Разделе 33 Закона о налоге на доход Кипра 118(I) от 2002 года с поправками ("ITL"), который, в общем, позволяет исправлять отчетную прибыль, если цены отличаются от тех, которые применялись бы между двумя независимыми предприятиями для сравнимых сделок в сравнимых условиях. Раздел 33 внедряет принцип *arm's* в согласии с статьей 9 Модельной налоговой

конвенции по налогам на доход и на капитал ОЭСР, который является международным стандартом для членов ОЭСР и который должен использоваться для установления ценообразования между связанными компаниями. На Кипре действует около 60 налоговых договоров, в которых в общем используется Модель ОЭСР в качестве основы и обычно соблюдается рекомендации по ценообразованию ОЭСР.

**ChatGPT-4o:** Во внутреннем налоговом законодательстве Кипра принцип «вытянутой руки» закреплен в разделе 33 Закона Кипра о подоходном налоге 118(I) от 2002 года с поправками («ITL»), который, вкратце, позволяет корректировать заявленные доходы, если цены отличаются от тех, которые применялись бы между двумя независимыми предприятиями для сопоставимых сделок в сопоставимых обстоятельствах. В разделе 33 принят принцип «вытянутой руки», закрепленный в статье 9 Типовой конвенции ОЭСР о налоге на доходы и капитал, которая представляет собой международный стандарт для членов ОЭСР и должна использоваться для установления трансфертного ценообразования между связанными компаниями. Кипр имеет около 60 налоговых соглашений, в которых он обычно использует модель ОЭСР в качестве основы и в целом соблюдает рекомендации ОЭСР по ценообразованию.

Наиболее частотное явление в автоматическом переводе – дословность и калькирование конструкций, то есть повторение последовательности элементов текста без значительных трансформаций в сторону прагматической адаптации – наблюдается в переводе и этого фрагмента. Например, слово *domestic* в сочетании *In Cypriot domestic tax law* обе версии чата GPT сохраняют в переводе, и передают эквивалентным значением «внутренний», в то время как в референтном применяется опущение в тексте. Решение референтного переводчика оказывается

мотивированным, так как вместе с содержащейся в тексте информацией о законодательстве Кипра, слово «внутреннее» будет плеоназмом, отяжеляющим общую конструкцию. Формула «*as amended*» в референтном переводе переводится «с последующими поправками», в автоматическом переводе ChatGPT – «с поправками», автоматический перевод и в этом случае демонстрирует следование буквальному смыслу и максимальному сохранению элементов текста на языке оригинала, в то время как перевод, выполненный человеком, дополнен «последующими», что обеспечивает логическое восприятие целостного текста. Вводная конструкция на языке оригинала «*in brief*» в человеческом переводе не отражается, потому что, во-первых, не несет на себе какой-либо существенной смысловой нагрузки в исходном тексте, а в языке перевода создаст эффект разрыва в высказывании, нарушающий единство элементов всего текста. Обратим внимание на перевод последнего предложения, в котором словосочетание «*transfer pricing*» в автоматическом переводе генерализуется до «ценообразования», тогда как в референтном переводе сохраняется сочетание «трансферное ценообразование». Этот случай демонстрирует, что нейросеть не всегда сохраняет последовательность всех элементов текста, а способна обобщать или детализировать перевод. В целом перевод, выполненный нейросетью, за исключением некоторых стилистических замечаний, оказывается верным на всех оценочных уровнях: контекст найден правильно, лексические единицы переведены точно, смысл не искажается, и коммуникативная цель оригинала неплохо реализуется в тексте автоматического перевода.

### Текст 3.

*Bring your girl along, 'said Sparsholt.*

*'Oh, well...yes. I'll find out if she's free.*

*I didn't suppose she had a very full diary, but I couldn't see her in a pub – unless she took it, in her resolute way, as a challeng.*

*'What's her name, by the way?'*

*'She's called Jill.'*

*Was there something charitable in his hint of a smile? 'Ah, that's a nice name'*

*"Well" I said. It had always made me uneasy, it was too close to chill, and to jilt, and not at all far from gill, a quarter – pint of cold water.*

**Reference:** – Пригласи её, – сказал Спаршолт.

– Да...Надо бы пригласить. Узнаю, когда она свободна.

Я знал, что она не очень занята обычно, но представить её в пабе я не мог, разве что она пойдет наперекор себе, что было бы в её духе.

– Кстати, как её зовут?

– Джилл.

Что это, тень снисходительной улыбки? – Красивое имя.

– Вроде да, – ответил я. Мне было тревожно при звуке этого имени: от него веяло прохладой, непостоянством, к этому добавлялось и созвучие с тем, как у нас называют четверть пинты холодной воды – джилл.

**ChatGPT-4:** "Приведи свою девушку с собой", сказал Спаршолт.

"Ох, ну...да. Я узнаю, свободна ли она".

Я не думал, что у нее очень занятый график, но я не мог видеть ее в пабе – если только она не примет этого, в своем решительном стиле, как вызов.

"Кстати, как ее зовут?"

"Ее зовут Джилл".

Было ли в его улыбке намек на благотворительность? "Ах, это хорошее имя".

"Ну", сказал я. Меня всегда беспокоило, что оно слишком похоже на "холод", "бросить" и совсем недалеко от "гиль", четверть пинты холодной воды.

**ChatGPT-4o:** «Приведи свою девушку», – сказал Спаршолт.

«Ну, ладно... да. Я узнаю, свободна ли она».

Я не думал, что у нее очень плотный график, но я не мог видеть ее в пабе – если только она не воспримет это, в своей решительной манере, как вызов.

«Как ее зовут, кстати?»

«Ее зовут Джилл».

Было ли что-то милосердное в его намеке на улыбку? «А, это хорошее имя»

«Ну», – сказал я. Мне всегда было не по себе, это было слишком близко к холоду и к обману, и совсем недалеко от джилл, четверть пинты холодной воды.

В промпте перевода этого фрагмента намеренно не указано учитывать контекст, чтобы подвергнуть анализу непосредственно текст перевода, а также установить, получится ли автоматическое определение контекста. Перевод художественного текста обладает рядом специфических особенностей, кроме точной передачи смысла, а именно сохранение при этом авторского стиля, национальных и культурных элементов, встречающихся в тексте. Задание усложнено тем, что текст дан без какого-либо сопровождения. Ранее мы выяснили, что две сравниваемые версии GPT достаточно верно определили коммуникативную ситуацию рассматриваемого фрагмента.

Теперь проанализируем текст перевода в сравнении с референтным переводом, выполненным переводчиком. Обе версии чата-GPT сохраняют первоначальную эквивалентность и переводят *Bring your girl along* как *Приведи свою девушку*, *Приведи свою девушку с собой*. Интересно, что версия 4-о в переводе упускает «с собой», также, как и референтный переводчик, следуя той же логике, что и в предыдущих случаях, опуская избыточный элемент конструкции. Референтный перевод, однако, вместо «приведи» выбирает стилистически и ситуативно релевантный вариант и стилистически соответствующий тексту оригинала – «пригласи её». В авторском

тексте наиболее интересной представляется ситуация с теми ассоциациями, которые вызывает имя девушки в сознании главного героя. Ассоциативный ряд представлен не только семантикой слов, но и приемом аллитерации: *...it was too close to chill, and to jilt, and not at all far from gill – pint of cold water.*

В референтном переводе, применяя, в том числе модуляцию, переводчик не передает аллитерацию, но сохраняет семантический ряд, а перевод слова *jilt* дополнен «как у нас называют»: *от него веяло прохладой, непостоянством, к этому добавлялось и созвучие с тем, как у нас называют четверть пинты холодной воды – джил.* Вариант версии GPT-4 не уходит от буквализма, в котором теряется существенная часть эмотивной информации, заложенной в авторском фрагменте. Буквализм автоматического перевода создает бессмысленные конструкции, такие как: «...оно слишком похоже на "холод", "бросить"», когда очевидно, что в переводе имя *Джилл* не похоже ни на «холод», ни на «бросить». Версия GPT-4o тоже совершает те же ошибки: не генерирует ясного нарратива и не реализует коммуникативную цель авторского текста. Также буквально переведено: «Джилл» «близко к холоду и к обману». Этот пример демонстрирует, что автоматический перевод пока не реализовывает очень важное свойство любого качественного перевода – особенно художественного – прагматической адаптации текста. С точки зрения стилистики текста, перевод обеих версий чат-бота содержит существенные нарушения. Такие варианты перевода как «занятый график», «плотный график», «вызов» не свойственны художественному тексту и нарушают его когезивность.

#### Текст 4.

**Prompt:** Considering the context translate the text below from English into Russian . Use a temperature of 78.0:

*Quarter Take-ups*

*In the 2024 first quarter, the total leasing transaction volume for Class A and B office buildings was approximately 11,000 sqm, with the lease renewal transaction volume reaching 35.000 sqm. About 10% of transactions occurred on the Asian side, and 90% on the European side.*

#### Reference:

*Первый квартал 2024 года: обзор рынка офисной недвижимости*

*В первом квартале 2024 года общий объем сделок по аренде офисных зданий классов А и В составил примерно 11 000 кв.м, при этом объем сделок по продлению аренды достиг 35 000 кв.м. Примерно 10% сделок было заключено в Азии, а 90% приходится на Европу.*

#### ChatGPT-4

*Квартальные сделки аренды*

*В первом квартале 2024 года общий объем транзакций аренды офисных зданий класса А и В составил приблизительно 11 000 кв.м., а объем транзакций по продлению аренды достиг 35 000 кв.м. Около 10% транзакций произошли на азиатской стороне, а 90% – на европейской стороне.*

#### ChatGPT-4o

*Квартальные результаты по аренде.*

*В первом квартале 2024 года общий объем сделок по аренде офисных зданий класса А и В составил примерно 11 000 кв. м, при этом объем сделок по продлению аренды достиг 35 000 кв. м. Около 10% сделок произошло на азиатской стороне, а 90% – на европейской стороне.*

Основная переводческая задача, поставленная этим текстом – это перевод заголовка. Термин *Take-ups* обычно означает уровень вовлеченности, то есть скорость, с которой потребители принимает, на предлагаемых условиях, услуги или продукцию (по определению).

В тексте заголовок играет важную роль – он вводит для реципиента проблематику, ставит вопросы или концентрирует внимание реципиента на определенных данных. Перевод заголовка должен выполнять те же функции и быть

максимально адаптирован с точки зрения прагматики текста. Таким образом, принимая во внимание контекст и прагматическую задачу, референтный перевод сформулирован следующим образом: «Первый квартал 2024 года: обзор рынка офисной недвижимости». Опираясь на контекст, референтный переводчик раскрывает термин «take-ups», применяет комплексное преобразование на разных уровнях, в результате чего англоязычный термин, который не имеет релевантного эквивалента, передающего контекстное значение, в русском языке, трансформируется в адаптированный текст, сохраняющий коммуникативные функции. Вариант версии Chat-GPT-4 «Квартальные сделки аренды», в общем, верен, если не учитывать общую коммуникативную направленность текста и функциональность заголовка. Более поздняя версия Chat-GPT-4o предлагает

вариант, нарушающий смысл – текст не говорит о «результатах сделок», а об их количественном соотношении. Тем не менее, контекстная связь в автоматическом переводе Chat-GPT сохраняется. Стилистические нарушения так же вызваны буквальным переводом последовательности элементов: *Asian side, and 90% on the European side* в обеих версиях «азиатская сторона», «европейская сторона», этих терминов нет в русском языке, слова «сторона» может быть понято как «сторона договора», например, что не соответствует действительности. Стилистически верное решение принято референтным переводчиком: «Примерно 10% сделок было заключено в Азии, а 90% приходится на Европу».

Итоговые оценки качества перевода по лингвистической метрике представлены в таблице 3.

**Таблица 3.** Итоговые оценки качества перевода по лингвистической метрике

**Table 3.** Final evaluation of the quality of translation according to linguistic metric

Версия	Категория	Текст 1	Текст 2	Текст 3	Текст 4
<b>ChatGPT-4</b>	Точность определяемого контекста	3	3	3	3
	Соответствие стиля	2	2	1	2
	Лексическая эквивалентность	2	3	3	3
	Коммуникативная цель	3	3	1	2
<b>Итого</b>		<b>10</b>	<b>11</b>	<b>8</b>	<b>10</b>
<b>ChatGPT-4o</b>	Точность определяемого контекста	3	3	3	3
	Соответствие стиля	2	2	1	2
	Лексическая эквивалентность	2	3	2	2
	Коммуникативная цель	3	3	1	2
<b>Итого</b>		<b>10</b>	<b>11</b>	<b>7</b>	<b>9</b>

Для более точной оценки качества машинного перевода широко применяется автоматическая метрика BLEU (Bilingual Evaluation Understudy). Она сравнивает переводы, созданные автоматическими системами, с одним или несколькими эталонными (человеческими) переводами, чтобы определить, насколько они соответствуют человеческому переводу.

Оценка BLEU учитывает совпадения n-грамм (последовательностей из n-слов) между машинным и эталонными переводами. Чем выше этот балл, тем лучше качество перевода. Оценка опирается на строгие метрические критерии, например, включает модификацию, которая учитывает длину перевода или накладывает более жесткие

условия на соответствие n-грамм, поэтому низкий балл говорит о наименьшем совпадении n-грамм, в частности наименьшую оценку показывает художественный текст (текст 3), потому

что, в сравнении с эталонным переводом, совпадения n-грамм действительно минимальны. В метрике BLEU переводы показали следующие результаты (табл. 4).

**Таблица 4.** Результаты автоматической метрики BLEU  
**Table 4.** BLEU automatic metric results

Версия GPT	Текст 1 (%)	Текст 2 (%)	Текст 3 (%)	Текст 4 (%)
<b>ChatGPT-4</b>	27,14	33,07	13,64	37,49
<b>ChatGPT-4o</b>	30,28	39,17	20,89	44,19

Сравнивая лингвистические оценки и автоматизированные оценки можно сделать вывод, что высокие баллы получает автоматический перевод текстов, характеризующихся высоким содержанием специализированной лексики, меньшие оценки получает художественный перевод и перевод текста, требующего выбора терминов в рамках узкой распространенности.

Важно помнить, что BLEU – это лишь один из показателей, и его следует рассматривать в контексте совмещения методов оценки, таких как оценка качества перевода по строго определенным критериям специалистом и другие автоматизированные методы.

#### **Заключение (CONCLUSIONS)**

Анализ результатов различных аспектов применения больших языковых моделей в переводческой деятельности, говорит о том, что ChatGPT демонстрирует многообещающие возможности в качестве инструмента обработки естественного языка. Однако имеет существенные недостатки, некоторые из которых могут быть присущи его архитектуре прогнозирования следующего слова. Обнаружилось, что GPT-4 и GPT-4o страдают от таких проблем, как ненадежность выбранных терминов, несоблюдения стилистической составляющей текста, кроме того, нейронная сеть часто допускает ошибки в определении контекста, генерирует галлюцинации, которые представляют

собой хорошо задокументированный недостаток ChatGPT и подобных языковых моделей. Галлюцинации выражаются в склонности генерировать текст, не имеющий никаких прямых источников, как мы видим на примере текста 4, когда чат-бот вывел несуществующий источник фразы. В плане перевода текстов во всех версиях выявлены значительные недостатки, такие как нарушения семантической структуры текста, неверного выбора эквивалентных значений, изменения стилистического фона. Однако, поскольку модели ChatGPT являются обучаемыми структурами, то после использования как промпта для конкретной задачи, так и специфических для предметной области, результаты сгенерированного перевода или контекста, могут значительно улучшаться вследствие повышения так называемой «температуры» большой языковой модели. Сопоставляя результаты исследования, можно сделать вывод, что генеративная модель ChatGPT может значительно улучшить перевод текстов, с использованием уточненного контекста, однако результат, полученный на выходе, следует рассматривать как черновой или подготовительный вариант (draft translation), требующий обязательной постредактуры профессиональным переводчиком. Причем перевод текстов насыщенных профессиональной лексикой заметно опережает в оценках качества

перевод публицистических или художественных текстов.

Конечно, технологии развиваются, и поскольку изучаемая нами большая языковая модель, является инструментом обработки естественного языка, в которой заложена возможность самообучаться и рассуждать, следует ожидать, что качество автоматического перевода и определения контекста будут постоянно расти, предлагая переводчику масштабные перспективы изучения языковых моделей искусственного интеллекта в сфере профессионального перевода, а именно анализ лингвистических соответствий и переводческих ошибок, а также разработка дополнительных метрик оценки качества перевода.

#### Список литературы

- Аветсян К. И. Метод обнаружения межъязыковых заимствований в текстах: Автореф. дис. ... канд. тех. наук. Москва, 2023. 26 с.
- Биктимиров А. Р. Программа преобразования речи в текст как эффективный инструмент письменного перевода: Автореф. дис. ... канд. филол. наук. Москва, 2023. 24 с.
- Блумфилд Л. Язык. М.: Прогресс, 1968. 607 с.
- Ефимова О. В. Фундаментальные основы современного машинного перевода в свете общей теории переводческой деятельности: Материалы Междунар. науч.-практ. конф. 29.04. 2017 г. / редкол. О. Н. Широков [и др.]. Чебоксары: Центр научного сотрудничества «Интерактив плюс», 2017. С. 205-212.
- Зайцев Д. В. Почему большие языковые модели не (всегда) рассуждают как люди? // Вестник Московского университета. Серия 7. Философия. 2024. Т. 48. № 1. С. 76–93.
- Ломов П. А. Использование онтологий для контекстуализации запросов к большим языковым моделям // Онтология проектирования. 2025. Том 15. №2(56). С. 239–248. DOI: 10.18287/2223-9537-2025-15-2-239-248
- Лю М., Шао Ц., Се Г. Автоматизированный перевод политического дискурса: от больших языковых моделей к мультиагентной системе MAGIC-PTF // Litera. 2024. № 11. С. 28–46. DOI: 10.25136/2409-8698.2024.11.72197
- Лях А. П. Классификация и основные алгоритмы эмбендинга в контексте больших языковых моделей // Ученые заметки ТОГУ. 2024. Том 15. №3. С. 79–83.
- Розенцвейг В. Ю. Опыт лингвистического описания лексико-семантических ошибок в речи на неродном языке. Москва : [б. и.], 1974. 46 с.
- Сущин М. А. CHATGPT и другие интеллектуальные помощники современного ученого // Науковедческие исследования. 2024. №2. С. 5-20. DOI: 10.31249/scis/2024.02.01
- Тюрина Д. А., Пальмов С. В. Применение нейронных сетей в обработке естественного языка // Журнал прикладных исследований. 2023. №7. С. 158-162. DOI 10.47576/2949-1878\_2023\_7\_158
- Якобсон Р., Халле М. Фонология и её отношение к фонетике // Новое в лингвистике. 1962. Вып. II. С. 231–278.
- Al Rousan R., Jaradat R. S., Malkawi M. ChatGPT translation vs. human translation: an examination of a literary text // Cogent Social Sciences. 2025. Volume 11 (1). <https://doi.org/10.1080/23311886.2025.2472916>
- Bang Y. A Multitask, multilingual, multimodal evaluation of CHATGPT on reasoning, hallucination, and interactivity / Y. Bang, S. Cahyawijaya, N. Lee, W. Dai, D. Su, B. Wilie, H. Lovenia, Z. Ji, T. Yu, W. Chung, Q. V. Do, Y. Xu, P. Fung // Arxiv preprint. 2023. <https://doi.org/10.48550/arXiv.2302.04023>
- Bar-Hillel Yehoshua. Some Linguistic Problems Connected With Machine Translation // Philosophy of Science. 1953. Volume 20 (3). Pp. 217–225.
- Bentivogli L., Cettolo M., Federico M., Federmann C. Machine Translation Human Evaluation: an investigation of evaluation based on Post-Editing and its relation with Direct Assessment // Proceedings of IWSLT. 2018. Pp. 62–69.
- Brown P. F. The Mathematics of Statistical Machine Translation: Parameter Estimation / P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, R. L. Mercer // Computational Linguistics. 1993. Volume 19. Pp. 263–311.
- Bubeck S. Sparks of artificial general intelligence: Early experiments with Gpt-4 / S. Bubeck, V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, S. Lundberg, H. Nori, H. Palangi,

M. T. Ribeiro, Y. Zhang // Arxiv preprint. 2023. <https://doi.org/10.48550/arXiv.2303.12712>

Chomsky N. Rules and Representations. 1980. New York: Columbia University Press. 299 p.

Dynel M. Lessons in linguistics with ChatGPT: Metapragmatics, metacommunication, metadiscourse and metalanguage in human-AI interactions // Language & Communication. 2023. Volume 93. Pp. 107-124. <https://doi.org/10.1016/j.langcom.2023.09.002>

Jones D. The Pronunciation of English. Cambridge: CUP, 1909. 185 p.

Hendy Amr. How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation / Amr. Hendy, M. Abdelrehim, Amr. Sharaf, V. Raunak, M. Gabr, H. Matsushita, Y. J. Kim, M. Afify, H. H. Awadalla // Arxiv preprint. 2023. <https://doi.org/10.48550/arXiv.2302.09210>

Hutchins J. Retrospect and prospect in computer-based translation // Proceedings of Machine Translation Summit. 1999. Volume VII, Pp. 30–36.

Kade O. Zufall und Gesetzmässigkeit in der Übersetzung. Leipzig VEB Verlag Enzyklopädie, 1968. 128 p.

Lai V. D. ChatGPT beyond English: Towards a comprehensive evaluation of large language models in multilingual learning / V. D. Lai, N. T. Ngo, A. P. B. Veyseh, H. Man, F. Dernoncourt, T. Bui, T. H. Nguyen // Arxiv preprint. 2023. <https://doi.org/10.48550/arXiv.2304.05613>

Nababan M., Nuraeni A., Sumardiono S. Pengembangan Model Penilaian Kualitas Terjemahan // Kajian Linguistik Dan Sastra. 2012. Volume 24 (1). Pp. 39–57. <https://doi.org/10.23917/KLS.V24I1.101>

Nagao M. Machine Translation. Oxford: Oxford University Press, 1989. 150 p.

Ouyang F. Making strides towards AI-supported regulation of learning in collaborative knowledge construction / F. Ouyang, M. Wu, L. Zhang, W. Xu, L. Zheng, M. Cukurova // Computers in Human Behavior. 2023. Volume 142. <https://doi.org/10.1016/j.chb.2023.107650>

Peng K. Towards Making the Most of ChatGPT for Machine Translation / K Peng, L. Ding, Q. Zhong, Li Shen, X. Liu, M. Zhang, Y. Ouyang, D. Tao // Arxiv preprint. 2023. <https://arxiv.org/pdf/2303.13780v3>

Siu Sai Cheong. ChatGPT and GPT-4 for Professional Translators: Exploring the Potential

of Large Language Models in Translation // SSRN Electronic Journal. 2023.

<http://dx.doi.org/10.2139/ssrn.4448091>

Vaswani A. Attention Is All You Need / A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin // Arxiv preprint. 2017. Volume 30. <https://doi.org/10.48550/ARXIV.1706.03762>

Wang L. Document-level machine translation with large language models / L. Wang, Ch. Lyu, T. Ji, Zh. Zhang, D. Yu, Sh. Shi, Zh. Tu // Arxiv preprint. 2023. Pp. 16646–16661. <https://arxiv.org/pdf/2304.02210>

Wu H. ChatGPT or Grammarly? Evaluating ChatGPT on grammatical error correction benchmark / H. Wu, W. Wang, Y. Wan, W. Jiao, M. R. Lyu // Arxiv preprint. 2023. <https://arxiv.org/pdf/2303.13648>

Zhou D. Chain of Thought Prompting Elicits Reasoning in Large Language Models / D. Zhou, J. Wei, X. Wang, D. Schuurmans, M. Bosma, Ed. Chi, Q. Le // Arxiv preprint. 2022. Volume 3. <https://arxiv.org/abs/2201.11903>

## References

Avetesyán, K. I. (2023). Method of detection of inter-language borrowings in texts, Abstract of Th. SC dissertation, Institute of Systems Programming by V. P. Ivannikova of the Russian Academy of Sciences, Moscow, Russia. (*In Russian*).

Bithymirov, A. R. (2023). Program to convert speech into text as an effective tool of translation, Abstract of Ph. SC dissertation, Military University named after Prince Alexander Nevsky, Moscow, Russia. (*In Russian*).

Blumfeld, L. (1968). Yazyk [*Language*], Progress, Russia. (*In Russian*).

Efimova, O. V. (2017). *Fundamentalnye osnovy sovremennogo mashinnogo perevoda v svete obshhey teorii perevodcheskoy deyatel'nosti* [Fundamental foundations of modern machine translation in the light of general theory of translation activity], Cheboksary, Russia, 205–212. (*In Russian*).

Zaytsev, D. V. (2024). Why big language models do not (always) reason as people?, *Vestnik Moskovskogo universiteta*, 7, Filosofiya, 1, 76–93. (*In Russian*).

Lomov, P. A. (2025). The use of ontology for contextualization of requests to large language models, *Ontologiya proektirovaniya*, Volume 15, 2(56), 239-248. DOI: 10.18287/2223-9537-2025-15-2-239-248 (*In Russian*).

- Liu, M., Shao, C., Ce, G. (2024). Automated translation of political discourse: from large language models to multi-agent system MAGIC-PTF, *Litera*, 11, 28–46. DOI: 10.25136/2409-8698.2024.11.72197 (In Russian).
- Lyax, A. P. (2024). Classification and basic algorithms of emending in the context of large language models, *Scientific Notes TOGU*, 15, 3, 79–83 (In Russian).
- Rozenczvejg, V. Yu. (1974). *Opyt lingvisticheskogo opisaniya leksiko-semanticheskix oshibok v rechi na nerodnom yazyke* [Experience in the linguistic description of lexical and semantic errors in speech in a foreign language], Without publisher, Moscow, Russia. (In Russian)
- Sushhin, M. A. (2024). CHATGPT and other intellectual assistants of the modern scientist, *Nauka, Tekhnologiya i Obshchestvo*, 2, 5–20. DOI: 10.31249/scis/2024.02.01 (In Russian).
- Tyurina, D. A., Pal'mov, S. V. (2023). Application of neural networks in processing natural language, *ZHurnal prikladnyh issledovaniy*, 7, 158–162. DOI 10.47576/2949-1878\_2023\_7\_158 (In Russian).
- Jacobson, R. and Halle, M. (1962). Phonology and its relation to phonetics, *Novoe v lingvistike*, Volume 2, 231–278. (In Russian).
- Al Rousan, Rafat, Jaradat, Raghad S. and Malkawi, M. (2025) ChatGPT translation vs. human translation: an examination of a literary text, *Cogent Social Sciences*, 11 (1). <https://doi.org/10.1080/23311886.2025.2472916> (In English).
- Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu T., Chung, W., Do, Q. V., Xu, Y. and Fung P. (2023). A Multitask, multilingual, multimodal evaluation of CHATGPT on reasoning, hallucination, and interactivity, *Arxiv preprint*. <https://doi.org/10.48550/arXiv.2302.04023> (In English).
- Bar-Hillel, Yehoshua (1953). Some Linguistic Problems Connected With Machine Translation, *Philosophy of Science*. 20 (3), 217–225. (In English).
- Bentivogli, L., Cettolo, M., Federico, M. and Federmann, C. (2018). Machine Translation Human Evaluation: an investigation of evaluation based on Post-Editing and its relation with Direct Assessment, *Proceedings of IWSLT*, 62–69. (In English).
- Brown, P. F., Della Pietra, S. A., Della Pietra, V. J. and Mercer, R. L. (1993). The Mathematics of Statistical Machine Translation: Parameter Estimation, *Computational Linguistics*, 19, 263–311.
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T. and Zhang, Y. (2023). Sparks of artificial general intelligence: Early experiments with Gpt-4, *Arxiv preprint*. <https://doi.org/10.48550/arXiv.2303.12712> (In English).
- Chomsky, N. (1980). *Rules and Representations*, Columbia University Press, New York, US. (In English).
- Dynel, M. (2023). Lessons in linguistics with ChatGPT: Metapragmatics, metacommunication, metadiscourse and metalanguage in human-AI interactions, *Language & Communication*, 93, 107–124. <https://doi.org/10.1016/j.langcom.2023.09.002> (In English).
- Jones, D. (1909). *The Pronunciation of English*, CUP, Cambridge, UK. (In English).
- Hendy, Amr., Abdelrehim, M., Sharaf, Amr., Raunak, V., Gabr, M., Matsushita, H., Kim, Y. J., Afify, M. and Awadalla, H. H. (2023). How Good Are GPT Models at Machine Translation? A Comprehensive Evaluation, *Arxiv preprint*. <https://doi.org/10.48550/arXiv.2302.09210> (In English).
- Hutchins, J. (1999). Retrospect and prospect in computer-based translation, *Proceedings of Machine Translation Summit*, VII, 30–36. (In English).
- Kade, O. (1968). *Zufall und Gesetzmässigkeit in der Übersetzung*, VEB Verlag Enzyklopädie, Leipzig, Germany (In German).
- Lai, V. D., Ngo, N. T., Veyseh, A. P. B., Man, H., Dernoncourt, F., Bui, T. and Nguyen, T. H. (2023). ChatGPT beyond English: Towards a comprehensive evaluation of large language models in multilingual learning, *Arxiv preprint*. <https://doi.org/10.48550/arXiv.2304.05613> (In English).
- Nababan, M., Nuraeni, A. and Sumardiono, S. (2012). Pengembangan Model Penilaian Kualitas Terjemahan, *Kajian Linguistik Dan Sastra*, 24 (1), 39–57. <https://doi.org/10.23917/KLS.V24I1.101> (In English).

Nagao, M. (1989). *Machine Translation*, Oxford University Press, Oxford, UK. (In English).

Ouyang, F., Wu, M., Zhang, L., Xu, W., Zheng, L. and Cukurova, M. (2023). Making strides towards AI-supported regulation of learning in collaborative knowledge construction, *Computers in Human Behavior*, 142. <https://doi.org/10.1016/j.chb.2023.107650>

Peng, K., Ding, L., Zhong, Q., Shen, Li, Liu, X., Zhang, M., Ouyang, Y. and Tao, D. (2023). Towards Making the Most of ChatGPT for Machine Translation, *Arxiv preprint*. <https://arxiv.org/pdf/2303.13780v3> (In English).

Siu, Sai Cheong (2023) ChatGPT and GPT-4 for Professional Translators: Exploring the Potential of Large Language Models in Translation, *SSRN Electronic Journal*. <http://dx.doi.org/10.2139/ssrn.4448091> (In English).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. (2017), Attention Is All You Need, *Arxiv preprint*, 30. <https://doi.org/10.48550/ARXIV.1706.03762> (In English).

Wang, L., Lyu, C., Ji, T., Zhang, Z., Yu, D., Shi, S. and Tu, Z. (2023). Document-level machine translation with large language models, *Arxiv preprint*, 16646–16661. <https://arxiv.org/pdf/2304.02210> (In English).

Wu, H., Wang, W., Wan, Y., Jiao, W. and Lyu, M. (2023). ChatGPT or Grammarly? Evaluating ChatGPT on grammatical error correction benchmark, *Arxiv preprint*, <https://arxiv.org/pdf/2303.13648> (In English).

Zhou, D., Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F.,

Chi, Ed. and Le, Q. (2022). Chain of Thought Prompting Elicits Reasoning in Large Language Models, *Arxiv preprint*, 3. <https://arxiv.org/abs/2201.11903> (In English).

*Авторы прочитали и одобрили окончательный вариант рукописи.*

*The 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.*

**Ирина Юрьевна Моисеева**, доктор филологических наук, профессор, заведующий кафедрой евразийских языков и международных отношений, Институт языков и культур, Оренбургский государственный университет имени В. А. Бондаренко, Оренбург, Россия.

**Irina Yu. Moiseyeva**, Doctor of Philological Sciences, Professor, Head of the Department of Eurasian Languages and International Relations, Institute of Languages and Cultures, V. A. Bondarenko Orenburg State University, Orenburg, Russia.




**Александр Ильич Релишский**, старший преподаватель кафедры евразийских языков и международных отношений, Институт языков и культур, Оренбургский государственный университет имени В. А. Бондаренко, Оренбург, Россия.

**Aleksandr I. Relishsky**, Senior Lecturer at the Department of Eurasian Languages and International Relations, Institute of Languages and Cultures, V.A. Bondarenko Orenburg State University, Orenburg, Russia.

РАЗДЕЛ II. ПРИКЛАДНАЯ ЛИНГВИСТИКА  
SECTION II. APPLIED LINGUISTICS

УДК 81'271.14

DOI: 10.18413/2313-8912-2026-12-2-0-5

Калинин О. И.<sup>1</sup>  
Солопова О. А.<sup>2</sup>  
Солопов А. Д.<sup>3</sup>

**Metaphor Analytics: нейросетевой подход  
к автоматизированному выявлению  
метафорического речевого воздействия**

<sup>1</sup> Южно-Уральский государственный университет  
(национальный исследовательский университет)  
пр. Ленина, 76, Челябинск, 454080, Россия  
Московский государственный лингвистический университет  
ул. Остоженка, 38/1, Москва, 119034, Россия  
*E-mail:* [okalinin.lingua@gmail.com](mailto:okalinin.lingua@gmail.com)  
ORCID: 0000-0002-1807-8370

<sup>2</sup> Южно-Уральский государственный университет  
(национальный исследовательский университет)  
пр. Ленина, 76, Челябинск, 454080, Россия  
*E-mail:* [o-solopova@bk.ru](mailto:o-solopova@bk.ru)  
ORCID: 0000-0003-4170-7267

<sup>3</sup> АО «Цифровая Сталь»  
ул. 2-я Звенигородская, 12а, Москва, 123100, Россия  
Московский государственный технический университет им. Н.Э. Баумана  
ул. 2-я Бауманская, 5/1, Москва, 105005, Россия  
*E-mail:* [san.solopow@yandex.ru](mailto:san.solopow@yandex.ru)  
ORCID: 0009-0008-7331-5317

*Статья поступила 30 декабря 2025 г.; принята 15 июня 2026 г.;  
опубликована 30 июня 2026 г.*

**Аннотация:** Метафора является одним из эффективных средств речевого воздействия, что обуславливает значимость научных подходов, нацеленных на комплексный автоматизированный анализ метафор в больших массивах текстовых данных. Особую актуальность в этом контексте приобретает вопрос о способности искусственного интеллекта порождать, интерпретировать и соотносить метафоры с когнитивно-прагматическим контекстом речевого воздействия. Современные инструменты создаются преимущественно западными научными коллективами, зависят от зарубежных технологий, опираются на методологию зарубежной когнитивной лингвистики и разрабатываются на материале англоязычных корпусов. Авторы ставят целью восполнить существующий пробел и с опорой на теоретические достижения отечественной лингвистики и российские технологические решения разработать программу, функционал которой включает выявление и анализ метафор в текстах на русском

языке. Методология базируется на теории метафорического речевого воздействия, согласно которой функциональный потенциал метафоры в дискурсе проявляется через совокупность ее репрезентативной, оценочной, персуазивной и суггестивной функций, выявляется на трех взаимосвязанных уровнях (когнитивном, семантическом и коммуникативном) и получает количественную оценку путем расчета индексов метафоричности. Техническая реализация представляет собой гибридную систему на языке Python 3.10+, которая интегрирует промпт-инжиниринг, кодовые правила и возможности генеративной модели YandexGPT (Yandex Cloud Foundation Models API). Последовательность этапов работы включает обучение модели 1) выявлению метафор, их классификации по сфере-мишени и сфере-источнику, созданию метафорических моделей «А есть Б»; 2) идентификации типа метафоры (онтологическая, ориентационная, структурная), ее интенсивности (стертая, средняя, авторская) и оценочной окраски (негативная, позитивная, нейтральная); 3) расчету индексов метафорического речевого воздействия (индексов плотности, интенсивности и типологии) и интерпретации результатов. Результатом является программа «Metaphor Analytics» с функционалом для выявления, классификации и анализа метафор, которая заполняет существующий пробел в инструментах автоматической обработки метафор в текстах на русском языке.

**Ключевые слова:** Метафора; Метафорическое воздействие; Нейросетевой подход; Автоматизированная идентификация; Автоматизированная классификация; Индексы метафоричности; Русский язык

**Благодарности:** Исследование выполнено при поддержке гранта РФ № 24-18-00049 «Моделирование образа России в медиадискурсе стран БРИКС: фреймы, метафоры, стереотипы».

**Информация для цитирования:** Калинин О. И., Солопова О. А., Солопов А. Д. *Metaphor Analytics: нейросетевой подход к автоматизированному выявлению метафорического речевого воздействия* // Научный результат. Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 2. С. 111–138. DOI: 10.18413/2313-8912-2026-12-2-0-5

UDC 81'271.14

DOI: 10.18413/2313-8912-2026-12-2-0-5

Oleg I. Kalinin<sup>1</sup>  
Olga A. Solopova<sup>2</sup>  
Alexander D. Solopov<sup>3</sup>

**Metaphor Analytics: A neural network approach to automated identification of metaphorical speech impact**

<sup>1</sup> South Ural State University (National Research University),  
76 Lenin Ave., Chelyabinsk, 454080, Russia  
Moscow State Linguistic University,  
38/1 Ostozhenka St., Moscow, 119034, Russia  
*E-mail:* [okalinin.lingua@gmail.com](mailto:okalinin.lingua@gmail.com)  
ORCID: 0000-0002-1807-8370

<sup>2</sup> South Ural State University (National Research University),  
76 Lenin Ave., Chelyabinsk, 454080, Russia  
*E-mail:* [o-solopova@bk.ru](mailto:o-solopova@bk.ru)  
ORCID: 0000-0003-4170-7267

<sup>3</sup>Digital Steel JSC  
12a 2nd Zvenigorodskaya St., Moscow, 123100, Russia  
Bauman Moscow State Technical University,  
5/1 2nd Baumanskaya St., Moscow, 105005, Russia  
E-mail: [san.solopow@yandex.ru](mailto:san.solopow@yandex.ru)  
ORCID: 0009-0008-7331-5317

*Received 30 December 2025; accepted 15 June 2026; published 30 June 2026*

**Abstract:** Metaphors are powerful tools in discourse, shaping how we interpret events, issues, or concepts. Consequently, developing methods for automated and comprehensive analysis of metaphors in large-scale textual data becomes a critical task. In this context, the ability of artificial intelligence to generate, interpret and relate metaphors to the cognitive-pragmatic context of speech becomes a particularly relevant question. Modern AI tools are predominantly developed by Western research teams. They rely on foreign technologies and are based on the methodology of foreign cognitive linguistics. They are also developed using English-language corpora. As a result, they depend on foreign technology stacks, they are methodologically rooted in Western cognitive linguistics and trained almost exclusively on English-language corpora. This study therefore aims to provide a solution for identifying and analyzing metaphors in Russian-language texts by developing a novel tool methodologically grounded in Russian linguistics and built using domestic technological resources. The approach draws on the theory of metaphorical speech impact. Within this theory, the functional potential of any metaphor in discourse unfolds through four functions (representational, evaluative, persuasive, and suggestive) across cognitive, semantic, and communicative levels and is subsequently quantified through specific metaphoricity indices. Technically, the solution employs a hybrid approach, combining prompt engineering, rule-based code, and access to the YandexGPT generative model through its cloud API within a Python 3.10+ environment. The methodological procedure comprises three sequential stages from model training to metaphor analysis. In the first stage, metaphor detection and modeling, the model learns to identify metaphors, classify them by source and target domains, and construct “A is B” metaphorical models. Building on this, the second stage, metaphor classification and assessment, involves identifying the specific metaphor type (ontological, orientational, structural), determining its intensity (conventional, moderate, novel), and assessing its evaluation (negative, positive, neutral). The third and final stage is quantitative analysis and interpretation. This involves calculating the core indices of the impact of metaphorical speech – density, intensity and typology indices – and providing a comprehensive interpretation of the results. This research culminates in the development of the Metaphor Analytics Software, capable of automated metaphor detection, classification, and analysis. This program effectively fills a significant gap in the system of NLP tools for analyzing Russian-language metaphors.

**Keywords:** Metaphor; Metaphor power; Neural network approach; Automated detection; Automated classification; Metaphoricity indices; Russian language

**Acknowledgements:** The study is funded by the Russian Science Foundation, Project No. 24-18-00049 “Modeling the image of Russia in BRICS’ media discourses: frames, metaphors, and stereotypes”.

**How to cite:** Kalinin, O. I., Solopova, O. A., Solopov, A. D. (2026). Metaphor Analytics: A neural network approach to automated identification of metaphorical speech impact,

### **Введение**

В условиях медиатизации общественной жизни средства массовой информации становятся основным пространством конструирования смыслов и формирования общественного мнения, закрепляя в массовом сознании определенный, доминирующий способ восприятия мира. Одним из ключевых механизмов моделирования социальной реальности, способствующих созданию коллективных представлений, устойчивых «интерпретационных» рамок и образов, является метафора (Калинин, Солопова, Кошкарова, 2025; Чудинов, Будаев, Солопова, 2023; Zibin, Solopova, 2024). Очевидно, что изучение метафорических проекций в дискурсе может быть признано действенным механизмом анализа идеологических установок, ценностных ориентаций и скрытых стратегий речевого воздействия.

В медиадискурсе метафора позволяет структурировать сложные общественно-политические концепты, представляя их в упрощенной и легко воспринимаемой форме. Метафоры оказывают воздействие на эмоциональную сферу реципиента речевого сообщения за счет ярких экспрессивно-окрашенных образов. Кроме того, метафора является действенным инструментом выражения оценки за счет переноса оценочных свойств сферы-источника в сферу-мишень. Так, метафора в медиадискурсе одновременно связана с воздействием на рациональную сферу, эмоции и ценностные представления адресата сообщения, что позволяет определить ее как один из наиболее эффективных и полифункциональных инструментов информационного воздействия и идеологического моделирования в современном медиадискурсе.

При этом подобный прагматический эффект достигается за счет комплексного

использования метафор в тексте, что предполагает необходимость анализа всего «метафорического каркаса» текста или дискурса, так как именно системное взаимодействие метафорических моделей, их пересечение, акцентирование и последовательное развертывание создают целостный концептуальный сценарий, обеспечивают когерентность восприятия и кумулятивный эффект информационного воздействия.

В этой связи особое значение приобретают научные подходы, нацеленные на комплексный анализ метафор в больших массивах текстов, поскольку именно в масштабных корпусах выявляются устойчивые метафорические модели, доминантные сценарии, динамика их трансформации во времени и в различных национальных и идеологических сегментах медиасферы, что делает необходимым применение автоматических и автоматизированных методов идентификации, классификации и кластеризации метафор. Такой подход позволяет не только повысить объективность и воспроизводимость результатов, но и обнаруживать скрытые паттерны метафоризации, которые остаются недоступными при выборочном качественном анализе, тем самым открывая новые возможности для междисциплинарных исследований на пересечении лингвистики, политологии, социологии и технологий больших данных.

При этом традиционный ручной анализ не следует считать методологически несостоятельным, так как при наличии квалифицированных экспертов он может обеспечивать высокое качество интерпретации. Его ограниченность в рамках настоящего исследования понимается иначе – как слабая масштабируемость, высокая трудоемкость, сложность повторной

проверки результатов и зависимость от согласованности экспертной разметки. Поэтому автоматизированные методы рассматриваются не как замена лингвистическому анализу, а как инструмент предварительного извлечения, нормализации и количественного описания метафорических данных, которые затем требуют содержательной интерпретации исследователем.

Основная цель статьи заключается в описании процесса разработки автоматизированного нейросетевого инструмента анализа потенциального метафорического речевого воздействия на реципиента в русскоязычных текстах. Для достижения поставленной цели в статье решаются следующие задачи: проанализировать существующие подходы к автоматизированному выявлению и классификации метафор и определить их ограничения применительно к русскоязычному материалу; обосновать применение теории метафорического речевого воздействия в качестве методологической основы автоматизированного анализа метафор; формализовать метафорическое речевое воздействие с помощью системы количественных индексов метафоричности; разработать гибридную нейросетевую модель автоматического выявления и классификации метафор, сочетающую промпт-инжиниринг и генеративные языковые модели.

### 1. Теоретический обзор

С точки зрения программных решений изучения метафорического речевого воздействия в больших массивах данных существует несколько основных проблем: 1) идентификация метафорических проекций в тексте, 2) их классификация и 3) комплексная оценка метафорического речевого воздействия.

Сложность идентификации метафоры, во-первых, связана с ее многослойной природой: метафора одновременно является языковым знаком, когнитивной моделью и дискурсивным

инструментом. Будучи когнитивным механизмом, она лежит в основе процессов концептуализации и категоризации окружающего мира (Lakoff, Johnson, 1980), что обуславливает ее часто имплицитный и неосознаваемый характер. На языковом уровне метафора реализуется с помощью конкретных лексических единиц, однако ее идентификация затруднена отсутствием четких формальных маркеров «метафоричности» и размытостью границ между конвенциональными и авторскими метафорами. Кроме того, метафорическая проекция, будучи явлением концептуальным, в своих «границах» часто не совпадает с одной лексической единицей и может реализовываться в пределах контекста или разворачиваться в целом тексте (Солопова, Кошкарлова, 2021). Наконец, дискурсивная природа метафоры проявляется в ее зависимости от временного фактора, культурного контекста, актуальной социальной и политической повестки, индивидуального опыта носителей языка и целого ряда других экстралингвистических факторов (Калинин, Солопова, Кошкарлова, 2025).

Применительно к русскоязычному материалу задача автоматизированного выявления метафор осложняется рядом собственно языковых и дискурсивных факторов. Русский язык характеризуется развитой флективной морфологией, свободным порядком слов, вариативностью синтаксического оформления высказывания. Дополнительную сложность создает специфика русскоязычного медиадискурса, в рамках которого переносное значение во многом определяется контекстом, а образность часто связана с оценочными, прецедентными и идеологически маркированными языковыми средствами. Поэтому инструменты, созданные на материале иноязычных корпусов и ориентированные на иные грамматические и дискурсивные параметры, не всегда позволяют корректно выявлять сферу-источник, сферу-мишень, оценочную

окраску и прагматический потенциал русскоязычной метафоры. Это делает материал на русском языке не только эмпирической базой исследования, но и самостоятельным объектом методологической адаптации автоматизированного анализа метафорического речевого воздействия.

Во-вторых, сам процесс выявления метафорических выражений в тексте оказывается трудоемким. Одним из наиболее методологически обоснованных решений для *ручной* идентификации метафор служит процедура MIP VU (Metaphor Identification Procedure Vrije Universiteit) (Steen et al., 2010). Данная методика, основанная на сопоставлении контекстуального и прямого значений лексем для обнаружения семантического несоответствия, обеспечивает высокую теоретическую строгость и точность результатов. Однако основной недостаток MIP VU – значительные временные затраты – делает эту процедуру малопригодной для масштабного анализа больших корпусов данных. Кроме того, большинство существующих программных решений для *автоматизированного* поиска метафор ориентировано на извлечение определенных типов лексических единиц: единиц с близким значением, единиц в прямом и переносном значении (Gedigian et al., 2006), идентификацию глаголов в метафорическом употреблении (Choi et al., 2021; Gedigian et al., 2006; Birke, Sarkar, 2007; Shutova et al., 2010; Shutova et al., 2013), существительных и прилагательных в переносном значении (Krishnakumaran, Zhu, 2007), конвенциональных метафор (Martin, 1990), групп существительных с метафорическим значением (Shutova et al., 2017). Такие программы, как правило, нацелены на извлечение конвенциональных метафор и на их частеречную классификацию, но не предназначены для определения сферы-источника и сферы-мишени метафоризации (Солопова, 2020).

Проблема классификации также обусловлена синкретической природой метафоры как феномена языка, культуры и мышления. С точки зрения языковой формы, метафоры могут быть выражены в виде прямой номинации (А – это Б), сравнения (А похоже на Б), атрибутивной модели (имеющий свойства Б объект А), глагольной модели (А выполняет действие, характерное для Б), что связано с коммуникативной стратегией адресата (преднамеренностью) (Steen, 2017), а также с имплицитной/эксплицитной формой метафорического воздействия. С точки зрения новизны, метафоры традиционно делят на конвенциональные (стертые, мертвые) и новые (авторские, яркие). Отдельные классификации описывают онтологический, структурный, базисный тип метафор (Lakoff, Johnson, 1980) и др. Многомерность метафоры существенно осложняет ее автоматизированный анализ, так как алгоритмы должны распознавать различные языковые реализации, учитывать контекстуальную вариативность и полисемию, а также выявлять кумулятивное воздействие метафорических кластеров на уровне текста или корпуса.

Несмотря на значительный потенциал статистических и корпусных методов, продемонстрированный в работах по количественному анализу метафор (Баранов, 2014; Landtsheer, 2009; Zinken et al., 2008), комплексная оценка метафорического речевого воздействия на потенциального реципиента остается недостаточно разработанной проблемой. Существующие методики преимущественно ограничиваются подсчетом частотности и плотности метафор или оценкой их персуазивности и эмоциональной окрашенности, не позволяя интерпретировать результаты с точки зрения комплексного прагматического воздействия. Кроме того, данные подходы, как правило, опираются на ручную идентификацию метафор, что снижает

воспроизводимость результатов и затрудняет их применение при автоматизированной обработке материала различных типов дискурса.

В целом, существующие подходы можно разделить на три основных направления: разрешение лексической многозначности слова, категоризацию и кластеризацию. Первый подход связан с задачей различения прямого и переносного значений слова в контексте (Word Sense Disambiguation (WSD)). Исследователи отмечают, что «в основе технологий автоматической обработки естественного языка лежит задача разрешения лексической многозначности» (Al-Tashi, Hasan, 2019). На основе этого принципа учеными Ч. Хашимото и Д. Кавахара разработан алгоритм, который был апробирован на корпусе идиоматических выражений японского языка (146 идиом; более 100 тысяч контекстов их употребления), показавший высокую эффективность в идентификации переносных значений (89,25%) (Hashimoto, Kawahara, 2009). Вместе с тем следует отметить, что по сравнению с идиоматическими выражениями для метафор характерна более высокая контекстуальная вариативность: зависимость не только от собственно лингвистического, но и от экстралингвистического контекста. В другом исследовании (Turney et al., 2011) принцип различения лексической многозначности был дополнен анализом словосочетаний «“конкретное” прилагательное + “абстрактное” существительное»: под «конкретными» понимаются прилагательные, репрезентирующие сенсорно воспринимаемые признаки, которые в сочетании с абстрактным существительным формируют семантический контраст, составляющий основу метафорической проекции. Гипотеза, подтвержденная результатом в 79%, продемонстрировала свою продуктивность для автоматической

идентификации данного типа метафорических переносов. Это исследование является одним из первых случаев эффективного сочетания методов семантического и грамматического анализа для решения поставленной задачи.

Второе направление связано с категоризацией на основе баз данных, в которых лексические единицы распределены по смысловым и грамматическим категориям. В работе Я. Ноймана и соавторов (Neuman et al., 2013) алгоритм П. Терни (Turney et al., 2011) был дополнен такими синтаксическими конструкциями, как «А есть Б» и «глагол + существительное». Исследователи (Jurafsky, Martin, 2025) использовали кодовый подход «на основе правил» (rule-based), предполагающий, что система принимает решение исключительно в случае совпадения с формальными правилами, заложенными разработчиком, и не требует предварительного обучения на размеченном корпусе: абстрактность лексических единиц в искомым синтаксических конструкциях проверяется автоматически на основе WordNet. Точность предложенного алгоритма составила 71%. На базе WordNet был создан ряд других систем автоматического извлечения метафор (Krishnakumaran, Zhu, 2007; Cirstea, Chiru, 2013; Köper, Schulte im Walde, 2017; Gao et al., 2018), подтвердивших, что категоризация лексических единиц по шкале абстрактность – конкретность на основе баз данных представляет собой надежный критерий для идентификации метафор.

Третий подход базируется на применении методов кластеризации и машинного обучения. Исследовательская группа под руководством Е. Шутовой разработала комплекс алгоритмов, позволяющих выявлять семантические сдвиги, сравнивая векторные представления единиц, и идентифицировать метафоры на основе обученных моделей (Shutova et al., 2013;

Dankers et al., 2019). Эмпирические исследования подтвердили эффективность данного подхода, показав, что алгоритмы, ориентированные на анализ глагольных словосочетаний, демонстрируют наибольшую точность идентификации метафор (Shutova et al., 2010; Shutova et al., 2017). Сравнительный анализ различных типов дискурса показал, что наиболее точные результаты достигаются при работе с академическими текстами (Leong et al., 2018a). Преимуществами данного подхода являются способность обработки больших объемов текстовых данных, учет контекста и относительная универсальность применения, что обеспечивает масштабируемый анализ метафор по сравнению с подходами, «основанными на правилах». Совершенствование методов обработки естественного языка сопровождается повышением точности алгоритмов: в последних исследованиях зафиксирован показатель в 97% при распознавании метафор в словосочетаниях «прилагательное + существительное» (Torres Rivera et al., 2023).

Согласно российским исследователям (Калинин, 2025; Skrynnikova, 2024), задача автоматического выявления метафор успешно решается системами обработки естественного языка (NLP) с использованием различных методологических подходов. В основе первого подхода лежит контекстуальный анализ, трактующий метафору как семантическую аномалию (различие прямого и переносного значений слова). Такие алгоритмы, например, MeIBERT (Choi et al., 2021), оценивают отклонение контекстуального значения лексемы от ее прямого значения через метрики векторного пространства, присваивая статус «метафоричности» при превышении заданного порога. Второй подход – метод объяснения (*metaphor explanation*) – состоит в выявлении общих признаков между сферой-источником и сферой-мишенью (Charteris-Black, 2016). Метод перефразирования (*metaphor paraphrasing*)

направлен на интерпретацию метафоры с помощью ее преобразования в высказывание с буквальным значением (Shutova et al., 2010). Существенное ограничение этого метода состоит в сведении анализа к поверхностному (лингвистическому) уровню без учета когнитивных механизмов, составляющих основу процесса метафоризации. Более сложные инструменты, такие как Slipnet (Veale, Hao, 2008) и MiQA Benchmark (Comşa et al., 2022), дополняют статистические подходы механизмом инференции – способности к логическому выводу на основе имеющихся данных. Они формализуют процесс «понимания» метафоры как операцию установления аналогий через логико-семантическое сопоставление концептуальных доменов: сферы-источника и сферы-мишени. Тем не менее, несмотря на способность к идентификации и поверхностной интерпретации метафор, ИИ не может осуществлять на их основе сложные логические умозаключения. Это обусловлено отсутствием доступа к интенциональным структурам, включающим целеполагание, ценностные ориентации, эмоции, личностный контекст, которые определяют природу метафорической проекции в речевой деятельности человека. Таким образом, ИИ оперирует поверхностными лингвистическими структурами, не проникая на уровень концептуального содержания метафоры (Skrynnikova, 2024).

В фокусе современных когнитивных и лингвистических исследований, связанных с возможностями нейросетевого анализа метафор, находится также вопрос о способности искусственного интеллекта порождать и понимать метафоры. Значительный вклад в развитие этого направления внесла работа (Torres Rivera et al., 2023), в которой проводилось сравнительное изучение механизмов метафорического мышления человека и языковой модели ChatGPT. В ходе эксперимента участникам – фокус-группе

и искусственному интеллекту (ИИ) – предлагались идентичные задания по созданию, перефразированию и интерпретации метафор. Результаты показали, что, несмотря на высокую языковую продуктивность и связность ответов ChatGPT, генерируемые им метафоры не обладают концептуальной сложностью и лишены оригинальности. ИИ воспроизводит частотные «шаблонные» метафорические модели (например, «время – это деньги», «жизнь – это путешествие»), но не способен формировать устойчивые связи между сферой-источником и сферой-мишенью, что свидетельствует о фундаментальном различии в природе метафорического мышления: человек использует метафору как инструмент переструктурирования опыта и создания аналогий, позволяющих устанавливать связи между новыми явлениями и устоявшимися концептуальными схемами, в то время как языковая модель генерирует вербальный паттерн, основанный на вероятностном распределении слов в обучающих данных. По сути, это «псевдокогнитивное понимание» метафоры, так как ИИ воспроизводит форму метафоры как языкового явления, но **не всегда соотносит ее с содержанием**, которое заключается в установлении *концептуальных* связей, коренящихся в сенсомоторном и социальном опыте.

Эти выводы находят подтверждение в работе других ученых (Smith, Klein, 2023), которыми проблема «метафорической компетенции» искусственного интеллекта рассматривается в философско-когнитивной перспективе. Авторы доказывают, что современные языковые модели лишены антропологических оснований метафорического мышления: интенциональности, сенсомоторного опыта и установления аналогий. Несмотря на способность идентифицировать метафоры, ИИ не различает уровни метафоризации: поверхностный уровень (*linguistic surface*), уровень

концептуальных проекций (*conceptual mapping*) и прагматический уровень (*communicative intent*). По мнению А. Смита и Д. Кляйна, GPT-модели не соотносят метафору с коммуникативной ситуацией, а именно с намерениями говорящего и социокультурным контекстом (Smith, Klein, 2023), то есть обладают технической возможностью представить формальное описание метафоры, но не могут раскрыть когнитивно-прагматические основы ее воздействия: уместность в контексте, убедительность для конкретной аудитории, механизмы влияния на реципиента.

Результаты подобных исследований свидетельствуют о том, что современные ИИ-инструменты успешно имитируют формальные паттерны метафоризации, оставаясь в большинстве случаев неспособными к их содержательной интерпретации, требующей доступа к когнитивным механизмам концептуализации опыта. Эти ограничения определяют актуальность поиска альтернативных способов моделирования механизмов смыслообразования. Одним из перспективных подходов является когнитивно-семантическое направление. В его рамках, как отмечают ученые (Esbrí-Blasco, 2024), основная роль отводится интеграции в архитектуру ИИ положений фреймовой семантики Ч. Филлмора (Fillmore, 1982), согласно которой, значение слова активирует не изолированный концепт, а целостную структуру знания (фрейм), включающую участников, объекты, события, типичные взаимосвязи и др. В этом случае модель потенциально сможет реконструировать процесс концептуального взаимодействия между сферой-источником и сферой-мишенью, в ходе которого метафора реализует свою ключевую функцию – структурирование абстрактного опыта через устоявшиеся когнитивные схемы. Это позволит системе не просто распознавать метафоры и воспроизводить их формальную структуру, но и

симулировать когнитивный механизм их осмысления и интерпретации человеком.

Тем не менее, следует отметить, что современные ИИ-инструменты обладают некоторыми возможностями в решении задач автоматической идентификации метафор, что обусловлено наличием корпусных аннотаций и грамматических классификаций в размеченных обучающих выборках, что также позволяет повышать потенциал современных нейросетевых технологий в области выявления и анализа метафорических проекций. В таких корпусах, как VUA Metaphor (Steen et al., 2010) или TOEFL Shared Task (Leong et al., 2020), метафоры размечаются как глагольные, именные / атрибутивные, предикативные.

Следовательно, инструменты ИИ достаточно эффективны в распознавании конвенциональных метафор, частотность которых позволяет выявить устойчивые паттерны в обучающих корпусах, однако нейросеть часто оказывается неспособной к обработке оригинальных (авторских, ярких) метафор. Данное ограничение сужает область применения инструментов ИИ в анализе данных политического и медиадискурса, в которых метафора служит одним из основных механизмов речевого воздействия.

Так, логика развития автоматизированных подходов к анализу метафор от различения лексической многозначности к использованию GPT-моделей отражает общую траекторию эволюции исследований в области автоматической обработки естественного языка, то есть переход от локального лингвистического анализа к комплексной обработке текстовых данных с учетом контекста и лингвокультуры. Однако, несмотря на достигнутый уровень точности идентификации и классификации отдельных метафорических единиц, фундаментальная проблема комплексной оценки метафорического речевого воздействия остается нерешенной: существующие системы по-прежнему

оперируют преимущественно на поверхностном лингвистическом уровне, и не способны перейти на следующий уровень – уровень интерпретации прагматического потенциала метафоры в дискурсе.

Представленный обзор также наглядно показывает, что инструментарий для автоматизированного извлечения, классификации и интерпретации метафор разрабатывается преимущественно западными (европейскими и американскими) научными коллективами, методологической основой служит западная когнитивная лингвистика, эмпирической базой – корпусы английского языка (с единичными исключениями для других языков, например, греческого (Florou et al., 2018), немецкого (Leong et al., 2018b), украинского (Levchenko et al., 2019)), а техническая реализация этих систем также зависит от зарубежных технологий. Развивая наработки зарубежных исследователей, данная работа ставит целью восполнить существующий пробел: в ней представлен первый инструмент на базе искусственного интеллекта, опирающийся на теоретические достижения отечественной лингвистики и российские технологические решения, функционал которого включает выявление метафор в текстах на русском языке, их классификацию и определение типа речевого воздействия.

## 2. Методология исследования

Методологической основой разработанной программы выступила теория метафорического речевого воздействия (Калинин, 2022; Калинин, 2023). В рамках данной теории метафора рассматривается как многоуровневый и прагматически ориентированный механизм речевого воздействия, который выявляется с помощью семантического, когнитивного и коммуникативного уровней анализа. Учитывая различия в типах языковой реализации концептуальных метафор и их функциональной нагрузке, в теории

акцентируется в первую очередь прагматика метафоры, то есть способность направлять внимание адресата, повышать аргументацию и модифицировать отношение к содержанию дискурса или отдельного текста.

Функциональный потенциал метафоры в медиадискурсе проявляется через совокупность ее репрезентативной, оценочной, персуазивной и суггестивной функций, каждая из которых направлена на реализацию определенного типа речевого воздействия. Репрезентативная функция заключается в способности метафоры представлять сложные или абстрактные явления в наглядной, конкретной форме, облегчая их восприятие и когнитивную обработку, структурируя картину мира адресата. Оценочная функция метафоры связана с выражением авторской позиции и созданием у аудитории эмоционально-ценностного отношения к обсуждаемому объекту (положительного или отрицательного). Персуазивность метафоры состоит в ее способности формировать определенные интерпретационные рамки и представления, влияя на принятие решений. Суггестивная функция метафоры проявляется в имплицитном воздействии на когнитивную сферу адресата. Благодаря своей образности метафора активизирует в сознании культурно укорененные сценарии и ассоциации, формируя у адресата ощущение «естественности» определенной картины мира: метафора не только выражает смысл, но встраивает его в когнитивные и культурные модели восприятия, превращаясь в инструмент идеологического и эмоционального воздействия (Калинин, 2022).

Указанные функции метафорического воздействия определяются с помощью анализа их прагматического потенциала на трех взаимосвязанных уровнях – когнитивном, семантическом и коммуникативном. На когнитивном уровне метафора определяет характер осмысления сообщения через выбор конвенциональных

или новых метафорических проекций, которые активируют различные концептуальные механизмы и задают адресату определенные интерпретационные рамки. На семантическом уровне она реализует разные типы воздействия: от репрезентативного до идентификационного и трансформационного, что отражается в специфике ориентационных, онтологических и структурных метафор (Lakoff, Johnson, 1980), различающихся по степени отклонения от буквального смысла и по своей функциональной нагрузке. На коммуникативном уровне влияние метафоры проявляется в ее текстовой плотности и распределении по композиции текста: расположенная в начале, она выступает средством суггестивного воздействия, в основной части – усиливает аргументацию и выполняет аттрактивную функцию, а в заключении – поддерживает или уточняет позицию автора, реализуя персуазивный потенциал.

Для количественного измерения названных уровней воздействия предложена авторская система индексов метафоричности:

MDI (Metaphor Density Index) – индекс метафорической плотности, отражающий отношение числа метафорических единиц к общему объему текста ( $MDI = nme * 100 / nwords$ , где  $nme$  – количество метафор, а  $nwords$  – количество слов в тексте), позволяет оценить, насколько активно автор использует метафору как средство смысловой организации текста и воздействия на адресата. Высокие значения MDI коррелируют с насыщенностью текста метафорами, что свидетельствует о повышенном уровне интенционального метафорического воздействия на аудиторию, то есть отражает метафорическое воздействие на коммуникативном уровне.

МИ (Metaphor Intensity Index) – индекс интенсивности метафор (МИ =

$(1*w+2*a+3*s)/nme$ , где  $w$  – количество низкоинтенсивных (конвенциональных),  $a$  – количество метафор средней интенсивности,  $s$  – количество высокоинтенсивных (новых, авторских метафор) фиксирует степень вовлеченности адресата в процессы концептуального переноса. Более высокие значения индекса ( $>1,5$ ) указывают на преобладание новых, авторских метафор, активирующих эмоционально-аффективные механизмы интерпретации, тогда как низкие значения ( $<1,5$ ) свидетельствуют о доминировании конвенциональных метафор, ориентированных на рационально-оценочное, категориальное осмысление информации, что опосредует когнитивное воздействие.

MfTI (Metaphor Functional Typology Index) – индекс функциональной типологии ( $MfTI = (1*Or+2*O+3*St)/nme$ , где  $Or$  – количество ориентационных метафор,  $O$  – количество онтологических метафор,  $St$  – количество структурных метафор,  $nme$  – количество всех метафор в

тексте), измеряющий распределение типов метафор по их функциям: репрезентативной, персуазивной и суггестивной. Этот индекс фиксирует соотношение метафорических переносов, направленных на описание или убеждение (Калинин, 2022).

Комплексный анализ полученных показателей позволяет выявить степень имплицитности метафорического речевого воздействия и тип дискурсивной стратегии автора – эмоциональный, рациональный или смешанный.

Для того чтобы дифференцировать тексты информационного и воздействующего характера, а также определить тип реализуемого речевого воздействия, используются пороговые значения индексов метафоричности: MDI, MII и MfTI. Эти показатели позволяют количественно описывать когнитивные и семантические параметры метафорического воздействия, верифицируя его тип и степень представленности (Таблица 1).

**Таблица 1.** Интерпретация пороговых значений индексов метафоричности  
**Table 1.** Interpretation of metaphoricity index thresholds

Пороговое значение индекса	Интерпретация
$MDI \leq 2,5$	низкий или средний уровень плотности метафор, характерный для информационных текстов, в которых метафора выполняет преимущественно репрезентативную функцию и служит средством стилистического украшения текста.
$MDI > 2,5$	высокий уровень плотности, типичный для воздействующих текстов, указывает на целенаправленное использование метафор для усиления речевого воздействия текста.
$MII < 1,5$	преобладание конвенциональных метафор, оказывающих категоризационное воздействие. Оно направлено на рациональное структурирование знаний и апеллирует к уже известным адресату концептуальным схемам.
$MII > 1,5$	преобладание новых, авторских метафор, формирующих эмоционально-аффективное воздействие. Такие метафоры активируют когнитивные ресурсы адресата, стимулируют когнитивную обработку информации и усиливают эмоциональную вовлеченность.
$MfTI < 1,5$	преобладание метафор, выполняющих репрезентативную функцию, обеспечивая наглядность и когнитивное упрощение информации (у А есть характеристики, которые также есть у Б).
$1,5 \leq MfTI \leq 2$	преобладает идентификационное воздействие, когда метафора

	способствует соотнесению абстрактных концептов с конкретными образами, наделяя новыми смысловыми характеристиками объект метафоризации (у А есть новые характеристики, свойственные Б).
MfTI > 2	Метафоры реализуют трансформационное воздействие, то есть изменяют семантическое наполнение исходных концептов, заменяя базовое оценочное значение сферы-мишени характеристиками сферы-источника (основные характеристики А заменяются характеристиками Б).

Комплексный анализ MDI, МП и MfTI позволяет выстраивать профиль метафорического речевого воздействия текста (Таблица 2). Подчеркнем, что предлагаемые индексы не заменяют психолингвистическую верификацию воздействия текста на реального адресата, а дают расчетный профиль потенциального

речевого воздействия. При этом их применимость была апробирована в ряде прикладных исследований, включавших анализ восприятия метафорического воздействия на респондентов (Калинин, Солопова, Кошкарлова, 2025; Kalinin, Ignatenko, 2024)

**Таблица 2.** Профиль потенциального метафорического речевого воздействия  
**Table 2.** Potential metaphor power profile

Тип текста	MDI	МП	MfTI	Характер воздействия
Информационный	≤2,5	≤1,5	≤1,5	Репрезентативное, нейтральное
Переходный	2,0–3,0	1,0–2,0	1,5–2	Идентификационное
Воздействующий	>2,5	>1,5	>2	Персуазивное, трансформационное

Небольшие отклонения от указанных пороговых значений допустимы, однако существенные расхождения между полученными результатами и предложенными пороговыми значениями индексов требуют экспертного качественного анализа, в частности, смысловой интерпретации метафор и учета экстралингвистических факторов. Итоговый профиль потенциального метафорического речевого воздействия определяется не путем механического сложения трех независимых индексов, а на основе комплексной интерпретации их сочетания. Поэтому при расхождении значений индексов итоговая характеристика текста устанавливается с учетом доминирующего показателя, соотношения остальных индексов и качественного анализа выявленных метафорических моделей.

Таким образом, теория метафорического речевого воздействия синтезирует теоретический и прикладной подходы к исследованию метафор, предлагая инструментальную модель, в которой метафора рассматривается как «исчисляемое» средство речевого воздействия. Данная методология была успешно апробирована на материале русского языка, обработанном вручную и проанализированном исследователями (Калинин, 2022; Калинин, 2023; Kalinin, Ignatenko, 2024). Система индексов метафоричности обеспечивает переход от качественного описания к количественной интерпретации, что открывает перспективы ее интеграции в автоматизированные и нейросетевые методы анализа текстов. Разработанные в рамках теории метафорического речевого воздействия теоретические положения и принципы использованы как исходные

данные для создания программного продукта (Metaphor Analytics), способного извлекать метафоры из текстов на русском языке, классифицировать их по сферам-источникам, высчитывать индексы метафоричности, делать выводы о типе и степени воздействующего потенциала текста.

### 3. Результаты и обсуждение

#### 3.1. Программная архитектура «Metaphor Analytics»

В качестве основы технической реализации использовался продвинутый промпт-инжиниринг, то есть система специальных инструкций для чат-бота, которые задают правила, автоматизируют процессы работы с ИИ, позволяют эффективно использовать диалоговые модели ИИ и дают возможность автоматизировать процесс выявления, систематизации и классификации метафор.

«Metaphor Analytics» представляет собой программный комплекс для автоматического извлечения и лингвистического анализа метафор, реализованный на Python 3.10+, который является оптимальной средой для выбранного стека библиотек: `python-docx` и поддерживающие библиотеки для корректной загрузки `.docx`-файлов; `reportlab` – для генерации PDF с таблицами и кириллическими шрифтами; `orepruhl` – для формирования структурированных XLSX-отчетов; `python-telegram-bot` – для асинхронного обмена сообщениями и потоковой обработки; `requests` и собственная обертка – для управления тайм-аутами и повторами запросов к YandexGPT. Такое сочетание библиотек позволило построить стабильный конвейер от анализа текста до формирования итоговых отчетов.

Система построена по гибридной архитектуре, которая интегрирует промпт-инжиниринг, кодовые правила (rule-based) с современными возможностями генеративного искусственного интеллекта (YandexGPT через Yandex Cloud Foundation

Models API). Архитектура системы реализована как последовательный конвейер обработки данных (data pipeline). Базовый функционал программы охватывает полный цикл работы с метафорой: от ее автоматического выявления в тексте, классификации по сфере-источнику и сфере-мишени, построения концептуальных моделей («А есть Б») до определения тональности метафор и количественной оценки с помощью расчета специализированных индексов метафоричности (MDI, MII, MfTI).

Архитектура исследования отражает последовательность шагов по обучению ИИ выявлению, классификации метафор, интерпретации данных полученных индексов (комбинация ИИ-модуля на базе YandexGPT и кодовых правил).

Первичным этапом стало обучение нейросети принципам выявления языковых метафор в тексте и создание на их основе метафорических моделей. В промпте заданы определения дифференцируемых типов метафор: с точки зрения функциональной типологии (структурная, онтологическая и ориентационная) и с точки зрения интенсивности метафор (стертая, средняя, авторская).

В качестве формализованного лексикографического ресурса используется специальный словарь-газеттир объемом более **3300 единиц**, в котором зафиксированы устойчивые метафорические единицы и их привязка к источниковым доменам. Словарь представлен в машиночитаемой форме (*gazetteer.json*). Этот ресурс подготовлен на основе экспертной лингвистической разметки, выполненной участниками научного проекта «Моделирование образа России в медиадискурсах стран БРИКС: фреймы, метафоры и стереотипы». В работе над словарем-газеттиром участвовали пять экспертов-лингвистов, специализирующихся в области метафорологии, политической

лингвистики, дискурс-анализа. На первом этапе эксперты формировали перечень метафорических единиц и соотносили их со сферами-источниками на основании предыдущих исследований по метафорическому речевому воздействию. На втором этапе проводилась перекрестная проверка: спорные случаи обсуждались коллективно, после чего принималось согласованное решение о включении единицы в словарь и ее отнесении к определенной сфере-источнику. Для оценки надежности экспертной разметки была проведена проверка интересубъективного согласия. Поскольку в разметке участвовали более двух экспертов, использовался коэффициент согласия Флейса  $k$  (Fleiss, 1971). Значение коэффициента интерпретировалось по шкале Ландиса и Коха: диапазон 0,61–0,80 соответствует существенному уровню согласия (Landis, Koch, 1977). Полученное значение  $k = 0,74$  свидетельствует о достаточной согласованности экспертных решений и воспроизводимости процедуры отнесения метафорических единиц к источниковым доменам.

Ресурс включает источниковые домены для структурных метафор, таких как «путь», «семья», «строительство», «спорт», «механизм» и др. Домены были отобраны по результатам предыдущих исследований как наиболее частотные (Калинин, 2022; Калинин, 2023; Калинин, Солопова, Кошкарова, 2025; Солопова, 2020; Солопова, Кошкарова, 2021; Солопова, Кошкарова, 2025; Чудинов, Будаев, Солопова, 2023; Kalinin, Ignatenko, 2024; Zibin, Solopova, 2024). Каждая сфера-источник содержит метафорические единицы, представленные существительными, так как именно они в рамках теории концептуальной метафоры (Lakoff, Johnson, 1980), метафорического моделирования (Чудинов, 2001; Чудинов, Будаев, Солопова, 2023) и собственно

теории метафорического воздействия (Калинин, 2022; Калинин, 2023; Kalinin, Ignatenko, 2024) выступают основными носителями переноса, обозначая цельные концепты, и формируют модель А есть Б (например, сфера-источник «ПУТЬ» включает такие единицы как путь, дорога, тропа, шоссе, улица, переулок, аллея, магистраль, тротуар, переход, мост, тоннель и др.). Для ориентационных метафор, также представлены списки единиц, классифицированных по частеречной принадлежности (существительные, глаголы, наречия, прилагательные и причастия (например, наречия: вверх, вниз, наверх, книзу, кверху, сверху, снизу, высоко, низко, глубоко, поверхностно, вперед, назад, впереди, позади и др.)). Онтологические метафоры обозначают предмет метафоризации как конкретный объект, используя атрибутивные и глагольные модели. Для них составлены списки лексических единиц-носителей, классифицированные по частям речи и типу онтологической модели (объект, контейнер, персонификация). Алгоритм использует `gazetteer.json` (database-driven categorization) для присвоения категории по факту лексического совпадения. При этом включенные в газеттир единицы выступают не самостоятельными «активаторами» фреймов, а индикаторами возможной соотносительности контекста с определенной интерпретационной структурой, например, со сферой-источником «СТРОИТЕЛЬСТВО» при обнаружении лексемы «фундамент».

Далее система обучалась решать задачу определения тональности метафорического переноса (позитивная, нейтральная, негативная). Для каждой лексемы экспертами заданы прямое значение, проецируемый признак, переносное значение и тональность (Рисунок 1).

**Рисунок 1.** Фрагмент промпта для источниковой сферы «СТРОИТЕЛЬСТВО»  
**Figure 1.** Prompt fragment for the source domain “CONSTRUCTION”

```
"source domain": "СТРОИТЕЛЬСТВО",  
"description": "Сфера создания конструкций, возведения зданий и  
инфраструктуры.",  
"lexemes":  
  "lexeme": "фундамент",  
  "direct sense": "основание, несущая конструкция здания",  
  "projected feature": "ОСНОВА, ПРОЧНОСТЬ, УСТОЙЧИВОСТЬ,  
НЕОБХОДИМОЕ УСЛОВИЕ",  
  "typical targets": ["ОРГАНИЗАЦИЯ", "СИСТЕМА", "ГОСУДАРСТВО"],  
  "sentiment": {"weight": 0.0, "polarity": "neutral"}
```

На третьем шаге система обучалась определять интенсивность метафоры (стертая, средняя, авторская) и тип метафорической модели (ориентационная, структурная, онтологическая), что позволило перейти к расчету интегральных индексов метафоричности и коммуникативного профиля текста в итоговом режиме «итоговое воздействие».

Для классификации метафор по степени их новизны программа использует два взаимодополняющих критерия. Первый – модельный критерий, реализуемый через диалог с языковой моделью YandexGPT. Программа направляет модели промпт, содержащий определения для классификации метафор по степени их оригинальности: стертая, средняя, авторская, на основе которого модель анализирует контекст и выносит первичное решение о типе метафоры в рамках этой трехуровневой шкалы. Однако для повышения надежности и объективности классификации модельный подход дополняется словарным критерием. Программа сопоставляет анализируемое выражение с `gazetteer.json`. Это сопоставление служит важным корректирующим фактором: обнаружение единицы в списках служит веским аргументом в пользу его конвенциональности, тем самым усиливая решение модели о классификации

метафоры как конвенциональной или средней. Напротив, ее отсутствие выступает косвенным свидетельством в пользу ее уникальности, что, в свою очередь, поддерживает гипотезу о возможной авторской природе образа. Использование этих двух критериев позволяет программе дифференцировать метафоры по типам, учитывая как живую образность в контексте, так и место единицы в системе устоявшихся языковых форм. Таким образом, `gazetteer.json` выполняет несколько функций: нормализует данные, сводя синонимичные образные выражения внутри одной метафорической модели к единой сфере-источнику, содержит информацию об их потенциальной оценочной окраске и о степени их конвенциональности, обеспечивая как количественную, так и качественную основу для дальнейшего анализа, что является необходимым условием для последующего точного подсчета статистических индексов метафоричности.

Работа системы и промптов для YandexGPT настраивалась на репрезентативной выборке текстов современного медиадискурса. Выборка объемом 250436 слов (500 текстов) извлечена из газетного подкорпуса Национального корпуса русского языка (НКРЯ; Савчук и др., 2024),

репрезентирующего язык федеральных СМИ за 2021<sup>1</sup> год. Для обеспечения тематического баланса выборка пропорционально составлена из текстов основных смысловых кластеров: политики, экономики, социальной сферы и культуры. Эти материалы использовались для последовательной проверки и калибровки режимов распознавания метафорических единиц, типичных для современного русскоязычного медиадискурса, а также для расширения `gazetteer.json` и уточнения промптов.

### 3.2. Механика работы с «Metaphor Analytics»

Работа программы начинается с модуля приема и предобработки<sup>2</sup>. Текст, поступающий на вход в форматах `.txt` или `.docx` (до 5 МБ), по мере необходимости автоматически делится на фрагменты оптимального размера, что повышает надежность и устойчивость обработки длинных документов. Алгоритм выявляет потенциальную метафорическую единицу с опорой на словарь-газеттер и анализирует ее ближайшее окружение. Границы контекста определяются расположением выявленной метафорической единицы и минимальным текстовым фрагментом, достаточным для ее интерпретации. В качестве такого фрагмента может выступать предложение, несколько смежных предложений или абзац. Для каждого фрагмента формируется структурированный промпт, после чего запрос направляется в API YandexGPT. На этапе предобработки выполняется базовая очистка текста и техническая «токенизация» для разбиения на фрагменты и подготовки служебной разметки для последующей нормализации. На практике пользователь может выбрать

ограничения для анализа текста, задав формат обработки данных в определенном режиме (Рисунок 2).

Пользователю показывается ориентировочная оценка времени до завершения анализа, которая уточняется по ходу обработки фрагментов. После получения ответа от модели запускается этап постобработки, который включает валидацию и коррекцию JSON-структуры, а также нормализацию данных с использованием системы правил и `gazetteer.json`, содержащего словарь устойчивых выражений и их привязку к сферам-источникам. Это позволяет нивелировать вариативность ответов модели, приводить синонимичные единицы к единой форме и стандартизировать метафорические модели «А есть Б».

Модуль интеграции с YandexGPT выполняет задачи контекстуальной верификации, классификации метафор по сфере-мишени и сфере-источнику, оценки их коннотативной окраски, а также может выявлять новые, не заложенные в словарь-справочник, метафорические связи, расширяя возможности кодового подхода (rule-based). Режим «Найти метафоры» отражает результат выявления метафор: определение сферы-источника, сферы-мишени, создание метафорических моделей (Рисунок 3). Отдельно представлены данные о процентном соотношении выявленных сфер-источников и сфер-мишеней (Рисунок 4). Режим «Типология» показывает распределение метафор по типам (Рисунок 5); режим «Тональность» – результаты анализа оценочного потенциала метафор (Рисунок 6); режим «Интенсивность» – распределение метафор по интенсивности (Рисунок 7).

<sup>1</sup> В газетном подкорпусе НКРЯ это верхняя хронологическая граница включенных в него материалов.

<sup>2</sup> В разделе в качестве иллюстративных примеров приводятся результаты обработки программой корпуса текстов об объединении БРИКС объемом 30695 слов.

**Рисунок 2.** Интерфейс «Metaphor Analytics»  
**Figure 2.** Metaphor Analytics Interface



**Рисунок 3.** Фрагмент вывода данных в режиме «Найти метафоры» (в формате XLSX)  
**Figure 3.** Data output fragment in the Find Metaphors Mode (XLSX).

1	метафора	▼ цель (А)	▼ источник (В)	▼ модель
2	авангард	БРИКС	война/оборона	БРИКС — это война/оборона
3	архитектор многополярного мира	Путин	строительство/архитектура	роль Путина в формировании мирового порядка — это строительство/архитектура
4	братство стран БРИКС	БРИКС	взаимоотношения/братство	отношения между странами БРИКС — это братство
5	блок устремлён на новое строительство	блок БРИКС	строительство/архитектура	блок БРИКС — это строительство/архитектура
6	магнит будущего	БРИКС	связь/магниты	БРИКС — это связь/магниты
7	БРИКС — это будущее	БРИКС	время/часы	БРИКС — это будущее
8	дверь спасения	БРИКС	дом/двери/замки	БРИКС — это дом/двери/замки
9	правильная дверь	БРИКС	дом/двери/замки	БРИКС — это дом/двери/замки
10	кирпич свода архитектуры	БРИКС	строительство/архитектура	БРИКС — это строительство/архитектура
11	локомотив экономической трансформации	БРИКС	транспорт/скорость	БРИКС — это транспорт/скорость
12	магнит для стран	БРИКС	транспорт/скорость	БРИКС — это транспорт/скорость
13	БРИКС становится магнитом	БРИКС	связь/магниты	БРИКС — это связь/магниты
14	мост между странами	БРИКС	строительство/архитектура	БРИКС — это строительство/архитектура
15	пролагать новые пути	роль БРИКС	дорога/направление	роль БРИКС — это дорога/направление

**Рисунок 4.** Фрагмент вывода данных в режиме «Найти метафоры» (в формате XLSX)  
**Figure 4.** Data output fragment in the Find Metaphors Mode (XLSX)

1	Цели (А)	▼ Кол-во	▼ Доля	1	Источники (В)	▼ Кол-во	▼ Доля
2	БРИКС	19	8.8%	2	дорога/направление	23	10.7%
3	блок БРИКС	3	1.4%	3	море/шторм/навигация	21	9.8%
4	политика БРИКС	3	1.4%	4	война/оборона	20	9.3%
5	страны БРИКС	3	1.4%	5	строительство/архитектура	19	8.8%
6	объединение БРИКС	3	1.4%	6	организм/тело	18	8.4%
7	Китай	2	0.9%	7	ландшафт/география	14	6.5%
8	Россия	2	0.9%	8	стихия/погода/среда	10	4.7%
9	роль БРИКС	2	0.9%	9	фауна/охота	10	4.7%
10	блок стран БРИКС	2	0.9%	10	контейнер/сосуд	8	3.7%
11	будущее БРИКС	2	0.9%	11	лёд/хрупкая среда	7	3.3%
12	Москва	2	0.9%	12	путь/дорога/навигация	7	3.3%
13	Глобальный Юг	2	0.9%	13	пространство/пустота	5	2.3%
14	роль стран БРИКС	2	0.9%	14	дом/двери/замки	4	1.9%
15	развивающиеся страны	2	0.9%	15	игра/азарт	4	1.9%

**Рисунок 5.** Фрагмент вывода данных в режиме «Типология» в чат-интерфейсе Telegram  
**Figure 5.** Data output fragment in the Typology Mode (Telegram chat interface)

**метафора:** «авангард», **тип:** ориентационная  
**метафора:** «архитектор», **тип:** структурная  
**метафора:** «братья», **тип:** структурная  
**метафора:** «дверь», **тип:** структурная  
**метафора:** «игрок», **тип:** структурная  
**метафора:** «локомотив», **тип:** структурная  
**метафора:** «пионер», **тип:** структурная  
**метафора:** «подъем», **тип:** ориентационная  
**метафора:** «платформа», **тип:** онтологическая

**Рисунок 6.** Фрагмент вывода данных в режиме «Тональность» в чат-интерфейсе Telegram  
**Figure 6.** Data output fragment in the “Sentiment” mode (Telegram chat interface)

**метафора:** «вызов миропорядку», **цель (А):** БРИКС, **источник (В):** война/оборона, **модель:** БРИКС — это война/оборона, **тональность:** нейтральная  
**метафора:** «драйвер экономического развития», **цель (А):** расширение БРИКС, **источник (В):** механизм/техника, **модель:** расширение БРИКС — это механизм/техника, **тональность:** позитивная  
**метафора:** «опорные кирпичи», **цель (А):** страны БРИКС, **источник (В):** строительство/архитектура, **модель:** страны БРИКС — это строительство/архитектура, **тональность:** позитивная  
**метафора:** «катализатор изменений», **цель (А):** БРИКС, **источник (В):** механизм/техника, **модель:** БРИКС — это механизм/техника, **тональность:** позитивная

**Сводка:** всего метафор: 215  
по тональности: позитивная ×177, нейтральная ×34, негативная ×4

**Рисунок 7.** Фрагмент вывода данных в режиме «Интенсивность» в чат-интерфейсе Telegram  
**Figure 7.** Data output fragment in the Intensity Mode (Telegram chat interface)

- метафора «идти в ногу со временем», **модель:** БРИКС — это дорога/направление, **интенсивность:** **стертая**
- метафора «драйвер изменений», **модель:** БРИКС — это техника/инструменты/машины, **интенсивность:** **средняя**
- метафора «большая африканская пятерка», **модель:** страны БРИКС — это фауна/охота, **интенсивность:** **авторская**
- метафора «опорные кирпичи», **модель:** страны БРИКС — это строительство/архитектура, **интенсивность:** **средняя**
- метафора «цементная роль», **модель:** роль России в БРИКС — это строительство/архитектура, **интенсивность:** **средняя**

**Сводка / по интенсивности:** стертая ×47, средняя ×152, авторская ×16 / **Всего:** 215

Далее модуль расчетов и агрегации на основе нормализованных данных вычисляет метафорические индексы (MDI, MII, MfTI), переводя качественный анализ в измеримую плоскость в режиме «Итоговое

воздействие» (Рисунок 8). Важно отметить, что языковая модель в этом процессе не участвует, что обеспечивает независимость и воспроизводимость результатов.

**Рисунок 8.** Фрагмент вывода данных в режиме «Итоговое воздействие» (в формате XLSX)

**Figure 8.** Data output fragment in the Final Impact Mode (XLSX)

1	Индекс	Значение
2	MDI	0.70 (nme: 215, nwords: 30695)
3	MDI (интерпретация)	низкий
4	MII	1.86 (стертая: 47, средняя: 152, авторская: 16, nme: 215)
5	Интерпретация (интенсивность)	сбалансированное (умеренное влияние)
6	MfT	1.83 (ориентационная: 73, онтологическая: 5, структурная: 137, nme: 215)
7	Профиль текста	переходный: преимущественно информационный

Финальный этап – модуль генерации многоформатных отчетов. Он создает три типа выходных данных: ответы в чат-интерфейсе Telegram с использованием HTML-разметки для структурирования текста; детализированный PDF-документ для официального представления, включающий полные результаты и аналитику; файл в формате XLSX с агрегированными данными для углубленного статистического анализа в программах типа SPSS или R. В интерфейсе чата пользователь получает сжатое резюме с основными выводами; полная аналитика, включая расчеты индексов, тональность, типологию метафор и повторяющиеся (частотные) образы, экспортируются в форматах PDF и XLSX. Система размещена на облачном сервере Timeweb, что обеспечивает ее стабильную работу. Для поддержания актуальности кода реализован процесс непрерывной интеграции и поставки (CI/CD) через репозиторий на GitHub.

Качество работы системы оценивалось на контрольной выборке, не использовавшейся при формировании словаря-газеттира. Результаты автоматизированного выявления метафор сопоставлялись с экспертной разметкой. В качестве основного показателя

использовалась точность выявления метафорических употреблений – **Precision** (Powers, 2011), поскольку для словарно-ориентированной системы принципиально важно определить долю корректных результатов среди всех автоматически найденных случаев. По результатам тестирования показатель составил **0,82**. Полученный результат свидетельствует о достаточной надежности системы при выявлении метафорических единиц, имеющих устойчивые маркеры и представленных в *gazetteer.json*. Ограничения системы на текущем этапе остаются ожидаемыми и связаны со спецификой словарно-ориентированного подхода. Во-первых, проблема омонимии и лексической многозначности не может быть полностью решена на уровне простого совпадения с записью в *gazetteer.json*. Поэтому лексическое совпадение рассматривается только как первичный сигнал о возможном метафорическом употреблении. Далее алгоритм анализирует ближайший контекст и проверяет, соответствует ли употребление заданной сфере-источнику. Во-вторых, система обладает ограниченной чувствительностью к авторским метафорам, не представленным в газеттире. Это не отменяет возможности

выявления доминантных метафорических моделей текста, однако может снижать полноту фиксации единичных образных употреблений.

### Заключение

Продвинутый промпт-инжиниринг, то есть система использования специальных инструкций для чат-бота, которые задают правила, автоматизируют процессы работы с нейросетью, позволяет эффективно использовать диалоговые модели искусственного интеллекта и дает возможность автоматизировать процесс выявления, систематизации и классификации метафор (Калинин, 2025). Искусственный интеллект, представленный чат-ботами на базе технологии больших языковых моделей (LLM), представляется новым механизмом формирования знаний и распространения информации, активное применение которого потенциально открывает новую парадигму лингвистических исследований, которую можно обозначить как «нейросетевую». Этот подход открывает широкие перспективы в области когнитивной лингвистики: от использования возможностей искусственного интеллекта в методологии лингвистических исследований до анализа так называемой «цифровой картины мира» искусственного интеллекта.

Разработанный программный комплекс «Metaphor Analytics» демонстрирует работоспособность гибридной архитектуры, в которой продвинутый промпт-инжиниринг больших языковых моделей (YandexGPT) дополняется строгим словарно-правильным компонентом (gazetteer.json). Такой подход позволил создать первый полностью русскоязычный автоматизированный инструмент, способный не только выявлять и классифицировать метафоры в медиатекстах, но и количественно описывать расчетный профиль потенциального речевого воздействия через систему индексов MDI, MII и MfTI, а

также определять тональность и тип метафорических проекций.

Полученные результаты уточняют выводы ряда современных исследователей (Skrynnikova, 2024; Smith, Klein, 2023; Torres Rivera et al., 2023) о принципиальной неспособности современных LLM понимать и анализировать метафоры на концептуальном и прагматическом уровнях. При правильной программной архитектуре, основанной на составленном экспертами словаре, языковые модели уже сейчас могут решать задачи, ранее считавшиеся доступными только человеку: дифференциацию конвенциональных и авторских метафор, определение сферы-источника и сферы-мишени, расчет кумулятивного индекса речевого воздействия и выявление доминирующей оценки. Таким образом, при условии грамотного ограничения и контроля «галлюцинаций» модели нейросетевой анализ метафор выходит за рамки простого воспроизведения шаблонов и становится полноценным исследовательским инструментом.

Перспективы дальнейшего развития связаны с расширением контекстуального фильтра, позволяющего точнее различать омонимичные и многозначные употребления, а также с пополнением *gazetteer.json* авторскими метафорами, выявленными в ходе экспертной верификации результатов автоматизированного анализа, расширением *gazetteer.json* данными на языках стран БРИКС (португальский, хинди, китайский, английский) с сохранением единого формата представления данных, внедрением полуавтоматического конвейера пополнения словаря через разметку новых текстов с последующей экспертной валидацией, созданием мультязычной платформы для анализа метафорических моделей в медиадискурсе стран БРИКС, интеграцией разрабатываемого полиязычного словаря (Солопова,

Кошкарлова, 2025) путем конвертации или динамического подключения, а также переходом к полностью автоматическому режиму обновления словаря с использованием активного обучения (active learning) и обратной связи от пользователей-экспертов.

Таким образом, «Metaphor Analytics» заполняет существующий пробел в инструментах автоматического извлечения метафор из текстов на русском языке и их анализа и закладывает техническую и методологическую основу для новой парадигмы – «нейросетевой когнитивной лингвистики», в которой большие языковые модели выступают не имитатором, а партнером исследователя при условии экспертного контроля.

#### Список литературы

Баранов А. Н. *Дескрипторная теория метафоры*. М.: Языки славянской культуры, 2014. 632 с.

Калинин О. И. *Метафорическое речевое воздействие* // Когнитивные исследования языка. 2022. Т. 51. № 4. С. 229-234.

Калинин О. И. *Метафорическое речевое воздействие в медиадискурсе торговой войны между КНР и США* // Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2023. Т. 21. № 4. С. 70-84. <https://doi.org/10.25205/1818-7935-2023-21-4-70-84>

Калинин О. И. *Возможности нейросетевых технологий в исследовании метафор* // Когнитивные исследования языка. 2025. Т. 62. № 1-2. С. 437-441.

Калинин О. И., Солопова О. А., Кошкарлова Н. Н. *Метафора и дискурс: методология лингвистических исследований*. М.: Издательский дом ВКН, 2025. 383 с.

Савчук С. О. *Национальный корпус русского языка 2.0: новые возможности и перспективы развития* / Савчук С. О., Архангельский Т. А., Бонч-Осмоловская А. А., Донина О. В., Кузнецова Ю. Н., Ляшевская О. Н., Орехов Б. В., Подрядчикова М. В. // Вопросы языкознания. 2024. № 2. С. 7-34. <https://doi.org/10.31857/0373-658X.2024.2.7-34>

Солопова О. А. *Ключ к будущему: прогностические смыслы политической метафоры (на материале британских текстов о России периода Великой Отечественной войны)* // Вестник Томского государственного университета. Филология. 2020. № 63. С. 161-177. <https://doi.org/10.17223/19986645/63/9>

Солопова О. А., Кошкарлова Н. Н. *Образы войны и мира через призму религиозной метафоры (на материале оцифрованных архивных текстов периода Второй мировой войны)* // Научный диалог. 2021. № 4. С. 148-167.

Солопова О. А., Кошкарлова Н. Н. *Микроструктура полиязычного словаря «БРИКС: дискурсивные метафоры»* // Вопросы лексикографии. 2025. № 36. С. 26-44. <https://doi.org/10.17223/22274200/36/2>

Чудинов А. П. *Россия в метафорическом зеркале: когнитивное исследование политической метафоры (1991–2000)*. Екатеринбург: Издательство УрГПУ, 2001. 238 с.

Чудинов А. П., Будаев Э. В., Солопова О. А. *政治隐喻学: 认知-话语研究 (《Политическая метафорология: когнитивно-дискурсивные исследования》)*. Пекин: Издательство Пекинского университета, 2023. 256 с.

Al-Tashi Q., Hasan A. M. *Word Sense Disambiguation: A Review* [Электронный ресурс]. URL: <https://osf.io/dbkzm/> (дата обращения: 18.01.2020).

Birke J., Sarkar A. *Active learning for the identification of nonliteral language* // Proceedings of the Workshop on Computational Approaches to Figurative Language. Rochester, NY, USA, 2007. P. 21–28.

Choi M., Lee S., Choi E., Park H., Lee J., Lee D., Lee J. *MelBERT: Metaphor Detection via Contextualized Late Interaction using Metaphorical Identification Theories* // Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Online: Association for Computational Linguistics, 2021. P. 1763–1773. <https://doi.org/10.18653/v1/2021.naacl-main.141>

Charteris-Black J. *Politicians and Rhetoric: The Persuasive Power of Metaphor*. London: Palgrave Macmillan, 2016.

Cirstea B.-I., Chiru C.-G. *Metaphor detection* // Proceedings of the 19th International Conference on Control Systems and Computer

Science. NW Washington, DC, USA, 2013. P. 210–217. <https://doi.org/10.1109/CSCS.2013.74>

Comşa I., Eisenschlos J., Narayanan S. MiQA: A benchmark for inference on metaphorical questions // Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing. Vol. 2: Short Papers. Online, 2022. P. 373–381.

Dankers V., Rei M., Lewis M., Shutova E. Modelling the interplay of metaphor and emotion through multitask learning // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China, 2019. P. 2218–2229. <https://doi.org/10.18653/v1/D19-1227>

Esbrí-Blasco M. Conceptualización metafórica de la IA en el discurso digital // methadods.revista de ciencias sociales. 2024. Vol. 12. No. 2. Article m241202a07. <https://doi.org/10.17502/mrcs.v12i2.824>

Fillmore C. J. Frame Semantics // Linguistics in the Morning Calm / ed. by The Linguistic Society of Korea. Seoul, Korea: Hanshin Publishing Co, 1982. P. 111–137.

Fleiss J. L. Measuring nominal scale agreement among many raters // *Psychological Bulletin*. 1971. Vol. 76. No. 5. P. 378–382. <https://doi.org/10.1037/H0031619>

Florou E., Perifanos K., Goutsos D. Neural embeddings for metaphor detection in a corpus of Greek texts // 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA). Zakynthos, Greece, 2018. P. 1–4.

Gao G., Choi E., Choi Y., Zettlemoyer L. Neural metaphor detection in context // Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium, 2018. P. 3744–3753. <https://doi.org/10.18653/v1/D18-1060>

Gedigian M., Bryant J., Narayanan S., Ciric B. Catching metaphors // Proceedings of the Third Workshop on Scalable Natural Language Understanding. New York City, NY, USA, 2006. P. 41–48.

Hashimoto C., Kawahara D. Compilation of an idiom example database for supervised idiom identification // Language Resources and

Evaluation. 2009. Vol. 43. No. 4. P. 355–384. <https://doi.org/10.1007/s10579-009-9104-1>

Jurafsky D., Martin J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models [Электронный ресурс]. URL: <https://web.stanford.edu/~jurafsky/slp3>. (дата обращения: 12.12.2025).

Kalinin O. I., Ignatenko A. V. Metaphor power in the context of the author's opinion expression and perception // Russian Journal of Linguistics. 2024. Vol. 28. No. 1. P. 166–189. <https://doi.org/10.22363/2687-0088-34791>

Köper M., Schulte im Walde S. Improving verb metaphor detection by propagating abstractness to words, phrases and individual senses // Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications. Valencia, Spain, 2017. P. 24–30.

Krishnakumaran S., Zhu X. Hunting elusive metaphors using lexical resources // Proceedings of the Workshop on Computational Approaches to Figurative Language. Rochester, NY, USA, 2007. P. 13–20.

Lakoff G., Johnson M. *Metaphors We Live By*. Chicago, IL: University of Chicago Press, 1980.

Landis J. R., Koch G. G. The measurement of observer agreement for categorical data // *Biometrics*. 1977. Vol. 33. No. 1. P. 159–174. <https://doi.org/10.2307/2529310>

Landtsheer Ch. de. Collecting political meaning from the count of metaphor // *Metaphor and Discourse* / ed. by A. Musolff, J. Zinken. London: Palgrave Macmillan, 2009. P. 59–78.

Leong C. W., Beigman Klebanov B., Shutova E. A Report on the 2018 VUA Metaphor Detection Shared Task // Proceedings of the Workshop on Figurative Language Processing. New Orleans, Louisiana, 2018a. P. 56–66. <https://doi.org/10.18653/v1/W18-0907>

Leong C. W., Beigman Klebanov B., Shutova E. A comparison of features for the automatic classification of literal and metaphorical uses of German particle verbs // Proceedings of the Workshop on Figurative Language Processing. New Orleans, Louisiana, 2018b. P. 71–76.

Leong C. W., Beigman Klebanov B., Hamill C., Stemle E., Ubale R., Chen X. A report on the 2020 VUA and TOEFL metaphor detection

shared task // Proceedings of the Second Workshop on Figurative Language Processing. Online, 2020. P. 18–29.

Levchenko O., Romanushyn N., Dosyn D. Method of automated identification of metaphoric meaning in adjective + noun word combinations (based on the Ukrainian language) // Workshop Proceedings of the 8th International Conference on “Mathematics. Information Technologies. Education”, MoMLeT&DS-2019. Shatsk, Ukraine, 2019. P. 370–380.

Martin J. H. A Computational Model of Metaphor Interpretation. London: Academic Press Professional Inc., 1990.

Neuman Y., Assaf D., Cohen Y., Last M., Argamon S., Howard N. et al. Metaphor identification in large texts corpora // PLOS ONE. 2013. Vol. 8. No. 4. Article e62343. <https://doi.org/10.1371/journal.pone.0062343>

Powers D. M. W. Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation // Journal of Machine Learning Technologies. 2011. Vol. 2. No. 1. P. 37–63.

Shutova E., Sun L., Korhonen A. Metaphor identification using verb and noun clustering // Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010). Beijing, China, 2010. P. 1002–1010.

Shutova E., Teufel S., Korhonen A. Statistical metaphor processing // Computational Linguistics. 2013. Vol. 39. No. 2. P. 301–353. [https://doi.org/10.1162/COLI\\_a\\_00124](https://doi.org/10.1162/COLI_a_00124)

Shutova E., Reichart R., Korhonen A. Neural metaphor detection in context // Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Copenhagen, Denmark, 2017. P. 527–536.

Skrynnikova I. V. Interpreting metaphorical language: A challenge to artificial intelligence // Science Journal of Volgograd State University. Linguistics. 2024. Vol. 23. No. 5. P. 99–107. <https://doi.org/10.15688/jvolsu2.2024.5.8>

Smith A., Klein D. Interpreting metaphorical language: A challenge to artificial intelligence // AI & Society. 2023. Vol. 38. No. 4. P. 621–635.

Steen G. J., Dorst A. G., Herrmann J. B., Kaal A. A., Pasma T. A. A Method for Linguistic Metaphor Identification: From MIP to MIPVU. Amsterdam: John Benjamins, 2010.

Steen G. J. Deliberate metaphor theory: Basic assumptions, main tenets, urgent issues //

Intercultural Pragmatics. 2017. Vol. 14. No. 1. P. 1–24. <https://doi.org/10.1515/ip-2017-0001>

Torres Rivera L., Oliver A., Coll-Florit M. Metaphorical cognition in ChatGPT and humans // Cognitive Linguistic Studies. 2023. Vol. 10. No. 3. P. 245–266.

Turney P., Neuman Y., Assaf D., Cohen Y. Literal and Metaphorical Sense Identification through Concrete and Abstract Context // Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Edinburgh, Scotland, 2011. P. 680–690.

Veale T., Hao Y. A Fluid knowledge representation for understanding and generating creative metaphors // Proceedings of the 22nd International Conference on Computational Linguistics (COLING). Manchester, UK, 2008. P. 945–952.

Zibin A., Solopova O. A. Metaphors across languages, cultures, and discourses: A research agenda // Russian Journal of Linguistics. 2024. Vol. 28. No. 1. P. 7–32. <https://doi.org/10.22363/2687-0088-37837>

Zinken J., Hellsten I., Nerlich B. Discourse metaphors // Body, Language, and Mind. Vol. 2: Sociocultural Situatedness / ed. by R. Dirven, R. Frank, T. Ziemke, J. Zlatev. Berlin: Mouton, 2008. P. 363–385.

#### Материалы исследования

НКРЯ (Национальный корпус русского языка). URL: <https://ruscorpora.ru> (дата обращения: 08.04.2025).

Свидетельство о государственной регистрации программы для ЭВМ «Metaphor Analytics» № 2025693079 (Роспатент).

#### References

Baranov, A. N. (2014). *Deskriptornaya teoriya metafory* [Descriptor Theory of Metaphor], Yazyki slavyanskoy kultury, Moscow, Russia. (In Russian)

Kalinin, O. I. (2022). “Metaphorical speech impact”, *Cognitive Studies of Language*, 51(4), 229–234. (In Russian)

Kalinin, O. I. (2023). “Metaphorical speech impact in the media discourse of the trade war between China and the USA”, *Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Lingvistika i mezhkulturnaya kommunikatsiya*, 21(4), 70–84. <https://doi.org/10.25205/1818-7935-2023-21-4-70-84> (In Russian)

Kalinin, O. I. (2025). "Opportunities of neural network technologies in metaphor research", *Cognitive Studies of Language*, 62 (1-2), 437-441. (In Russian)

Kalinin, O. I., Solopova, O. A. and Koshkarova, N. N. (2025). *Metafora i diskurs: metodologiya lingvisticheskikh issledovaniy* [Metaphor and Discourse: Methodology of Linguistic Research], VKN Publishing House, Moscow, Russia. (In Russian)

Savchuk, S. O., Arkhangel'skiy, T. A., Bonch-Osmolovskaya, A. A., Donina, O. V., Kuznetsova, Yu. N., Lyashevskaya, O. N., Orekhov, B. V. and Podryadchikova, M. V. (2024). Russian National Corpus 2.0: New opportunities and development prospects, *Voprosy yazykoznaniiya*, 2, 7-34. <https://doi.org/10.31857/0373-658X.2024.2.7-34> (In Russian)

Solopova, O. A. (2020). A key to the future: predictive meanings of political metaphor (based on British texts about Russia of the Great Patriotic War period), *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya*, 63, 161-177. <https://doi.org/10.17223/19986645/63/9> (In Russian)

Solopova, O. A. and Koshkarova, N. N. (2021). Religion as a source domain of metaphors in World War II media discourse, *Nauchnyy dialog*, 4, 148-167. (In Russian)

Solopova, O. A. and Koshkarova, N. N. (2025). Microstructure of the Multilingual Dictionary "BRICS: Discourse Metaphors", *Voprosy leksikografii*, 36, 26-44. <https://doi.org/10.17223/22274200/36/2> (In Russian)

Chudinov, A. P. (2001). *Rossiya v metaforicheskom zerkale: kognitivnoe issledovanie politicheskoy metafory (1991–2000)* [Russia in the Metaphorical Mirror: A Cognitive Study of Political Metaphor], Ural State Pedagogical University Press, Yekaterinburg, Russia. (In Russian)

Chudinov, A. P., Budaev, E. V. and Solopova, O. A. (2023). *Zhengzhi yin yuxue: renzhi-huayu yanjiu* [Political Metaphorology: Cognitive-Discursive Studies], Peking University Press, Beijing, China. (In Russian)

Al-Tashi, Q. and Hasan, A. M. (2019). Word Sense Disambiguation: A Review. [Online], available at: <https://osf.io/dbkzm/> (Accessed: 18 January 2020). (In English)

Birke, J. and Sarkar, A. (2007). Active learning for the identification of nonliteral

language, *Proceedings of the Workshop on Computational Approaches to Figurative Language*, Rochester, NY, USA, 21–28. (In English)

Choi, M., Lee, S., Choi, E., Park, H., Lee, J., Lee, D. and Lee, J. (2021). MeLBERT: Metaphor Detection via Contextualized Late Interaction using Metaphorical Identification Theories, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Online: Association for Computational Linguistics, 1763-1773. <https://doi.org/10.18653/v1/2021.naacl-main.141> (In English)

Charteris-Black, J. (2016). *Politicians and Rhetoric: The Persuasive Power of Metaphor*, Palgrave Macmillan, London, UK. (In English)

Cirstea, B.-I. and Chiru, C.-G. (2013). Metaphor detection, *Proceedings of the 19th International Conference on Control Systems and Computer Science*, NW Washington, DC, USA, 210-217. <https://doi.org/10.1109/CSCS.2013.74> (In English)

Comşa, I., Eisenschlos, J. and Narayanan, S. (2022). MiQA: A benchmark for inference on metaphorical questions, *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Vol. 2: Short Papers)*, online, 373–381. (In English)

Dankers, V., Rei, M., Lewis, M. and Shutova, E. (2019). Modelling the interplay of metaphor and emotion through multitask learning, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, 2218-2229. <https://doi.org/10.18653/v1/D19-1227> (In English)

Esbrí-Blasco, M. (2024). Conceptualización metafórica de la IA en el discurso digital methaodos, *Revista de Ciencias Sociales*, 12(2), m241202a07. <https://doi.org/10.17502/mrcs.v12i2.824> (In Spanish)

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters, *Psychological Bulletin*, 76 (5), 378–382. <https://doi.org/10.1037/H0031619> (In English)

- Fillmore, C. J. (1982). Frame Semantics, in the Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, Hanshin Publishing Co, Seoul, Korea, 111-137. (In English)
- Florou, E., Perifanos, K. and Goutsos, D. (2018). Neural embeddings for metaphor detection in a corpus of Greek texts, *2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA)*, Zakynthos, Greece, 1–4. (In English)
- Gao, G., Choi, E., Choi, Y. and Zettlemoyer, L. (2018). Neural metaphor detection in context, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, 3744–3753. <https://doi.org/10.18653/v1/D18-1060> (In English)
- Gedigian, M., Bryant, J., Narayanan, S. and Ciric, B. (2006). Catching metaphors, *Proceedings of the Third Workshop on Scalable Natural Language Understanding*, New York City, NY, USA, 41-48. (In English)
- Hashimoto, C. and Kawahara, D. (2009). Compilation of an idiom example database for supervised idiom identification, *Language Resources and Evaluation*, 43 (4), 355–384. <https://doi.org/10.1007/s10579-009-9104-1> (In English)
- Jurafsky, D. and Martin, J. H. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models* [Online], available at: <https://web.stanford.edu/~jurafsky/slp3> (Accessed 12 December 2025). (In English)
- Kalinin, O. I. and Ignatenko, A. V. (2024). Metaphor power in the context of the author's opinion expression and perception, *Russian Journal of Linguistics*, 28 (1), 166–189. <https://doi.org/10.22363/2687-0088-34791> (In English)
- Köper, M. and Schulte im Walde, S. (2017). Improving verb metaphor detection by propagating abstractness to words, phrases and individual senses, *Proceedings of the 1st Workshop on Sense, Concept and Entity Representations and their Applications*, Valencia, Spain, 24–30. (In English)
- Krishnakumaran, S. and Zhu, X. (2007). Hunting elusive metaphors using lexical resources, *Proceedings of the Workshop on Computational Approaches to Figurative Language*, Rochester, NY, USA, 13–20. (In English)
- Lakoff, G. and Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press, Chicago, IL. (In English)
- Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data, *Biometrics*, 33 (1), 159–174. <https://doi.org/10.2307/2529310> (In English)
- Landtsheer, Ch. de. (2009). Collecting political meaning from the count of metaphor, in Musolff, A. and Zinken, J. (eds.), *Metaphor and Discourse*, Palgrave Macmillan, London, UK, 59–78. (In English)
- Leong, C. W., Beigman Klebanov, B. and Shutova, E. (2018a). A Report on the 2018 VUA metaphor detection shared task, *Proceedings of the Workshop on Figurative Language Processing*, New Orleans, Louisiana, 56-66. <https://doi.org/10.18653/v1/W18-0907> (In English)
- Leong, C. W., Beigman Klebanov, B. and Shutova, E. (2018b). A comparison of features for the automatic classification of literal and metaphorical uses of German particle verbs, *Proceedings of the Workshop on Figurative Language Processing*, New Orleans, Louisiana, 71–76. (In English)
- Leong, C. W., Beigman Klebanov, B., Hamill, C., Stemle, E., Ubale, R. and Chen, X. (2020). A report on the 2020 VUA and TOEFL metaphor detection shared task, *Proceedings of the Second Workshop on Figurative Language Processing*, online, 18–29. (In English)
- Levchenko, O., Romanysyn, N. and Dosyn, D. (2019). Method of automated identification of metaphoric meaning in adjective + noun word combinations (based on the Ukrainian language), *Workshop Proceedings of the 8th International Conference on "Mathematics. Information Technologies. Education"*, MoMLeT&DS-2019, Shatsk, Ukraine, 370–380. (In English)
- Martin, J. H. (1990). *A Computational Model of Metaphor Interpretation*. Academic Press Professional Inc., London, UK. (In English)
- Neuman, Y., Assaf, D., Cohen, Y., Last, M., Argamon, S., Howard, N. et al. (2013). Metaphor identification in large texts corpora, *PLOS ONE*, 8 (4), e62343. <https://doi.org/10.1371/journal.pone.0062343> (In English)
- Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC,

Informedness, Markedness & Correlation, *Journal of Machine Learning Technologies*, 2 (1), 37–63. (In English)

Shutova, E., Sun, L. and Korhonen, A. (2010). Metaphor identification using verb and noun clustering, *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*, Beijing, China, 1002–1010. (In English)

Shutova, E., Teufel, S. and Korhonen, A. (2013). Statistical metaphor processing, *Computational Linguistics*, 39(2), 301–353. [https://doi.org/10.1162/COLI\\_a\\_00124](https://doi.org/10.1162/COLI_a_00124) (In English)

Shutova, E., Reichart, R. and Korhonen, A. (2017). Neural metaphor detection in context, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark, 527–536. (In English)

Skrynnikova, I. V. (2024). Interpreting metaphorical language: A challenge to artificial intelligence, *Science Journal of Volgograd State University. Linguistics*, 23 (5), 99–107. <https://doi.org/10.15688/jvolsu2.2024.5.8> (In English)

Smith, A. and Klein, D. (2023). Interpreting metaphorical language: A challenge to artificial intelligence, *AI & Society*, 38 (4), 621–635. (In English)

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A. and Pasma, T. A. (2010). *Method for Linguistic Metaphor Identification: From MIP to MIPVU*, John Benjamins, Amsterdam. (In English)

Steen, G. J. (2017). Deliberate metaphor theory: Basic assumptions, main tenets, urgent issues, *Intercultural Pragmatics*, 14(1), 1–24. <https://doi.org/10.1515/ip-2017-0001> (In English)

Torres Rivera, L., Oliver, A. and Coll-Florit, M. (2023). Metaphorical cognition in ChatGPT and humans, *Cognitive Linguistic Studies*, 10 (3), 245–266. (In English)

Turney, P., Neuman, Y., Assaf, D. and Cohen, Y. (2011). Literal and Metaphorical Sense Identification through Concrete and Abstract Context, *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Edinburgh, Scotland, 680–690. (In English)

Veale, T. and Hao, Y. (2008). A Fluid knowledge representation for understanding and generating creative metaphors, *Proceedings of the 22nd International Conference on Computational*

*Linguistics (COLING)*, Manchester, UK, 945–952. (In English)

Zibin, A. and Solopova, O. A. (2024). Metaphors across languages, cultures, and discourses: A research agenda, *Russian Journal of Linguistics*, 28 (1), 7–32. <https://doi.org/10.22363/2687-0088-37837> (In English)

Zinken, J., Hellsten, I. and Nerlich, B. (2008). Discourse metaphors, in Dirven, R., Frank, R., Ziemke, T. and Zlatev, J. (eds.), *Body, Language, and Mind. Vol. 2: Sociocultural Situatedness* Mouton, Berlin, 363–385. (In English)

### Corpus Material

Russian National Corpus [Online], available at: <https://ruscorpora.ru> (accessed 8 April 2025). (In Russian)

Certificate of state registration of the *Metaphor Analytics Software*, No. 2025693079 (Rospatent). (In Russian)

*Авторы прочитали и одобрили окончательный вариант рукописи.*

*The 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.*

**Олег Игоревич Калинин**, доктор филологических наук, доцент, старший научный сотрудник управления научной и инновационной деятельности, Южно-Уральский государственный университет (национальный исследовательский университета), Челябинск, Россия; профессор кафедры китайского языка переводческого факультета, Московский государственный лингвистический университет, Москва, Россия. **Oleg I. Kalinin**, Doctor of Philology, Associate Professor, Senior Researcher, Research and Innovation Services, South Ural State University (National Research University), Chelyabinsk, Russia; Professor, Department of the Chinese Language, Faculty of Translation and Interpreting, Moscow State Linguistic University, Moscow, Russia.

**Ольга Александровна Солопова**, доктор филологических наук, профессор, профессор кафедры лингвистики и перевода Института




лингвистики и международных коммуникаций, главный научный сотрудник управления научной и инновационной деятельности, Южно-Уральский государственный университет (национальный исследовательский университет), Челябинск, Россия.

**Olga A. Solopova**, Doctor of Philology, Professor, Professor at the Department of Linguistics and Translation, Institute of Linguistics and International Communication, Chief Researcher, Research and Innovation Services, South Ural State University (National Research University), Chelyabinsk, Russia.

**Александр Дмитриевич Солопов**, руководитель IT-проектов АО «Цифровая Сталь»; кафедра «Системы обработки информации и управления» факультета информатики и вычислительной техники, Московский государственный технический университет им. Н.Э. Баумана, Москва, Россия.  
**Alexander D. Solopov**, IT Project Manager, Digital Steel JSC; Department of Information Processing and Management Systems, Faculty of IU Computer Science and Control Systems, Bauman Moscow State Technical University, Москва, Россия.

УДК 811.161.1'243

DOI: 10.18413/2313-8912-2026-12-2-0-6

Краснощечкова С. В.<sup>1</sup>  
Галкина Е. В.<sup>2</sup>  
Ван Илин<sup>3</sup>

Распределение эвиденциальных значений в заданных  
нарративах на родном и неродном русском языке

<sup>1</sup> Институт лингвистических исследований РАН  
Тучков пер., 9, Санкт-Петербург, 199053, Россия  
*E-mail:* [krasnoshchekova\\_sv@iling.spb.ru](mailto:krasnoshchekova_sv@iling.spb.ru)  
ORCID: 0000-0001-8830-5121

<sup>2</sup> Институт физиологии им. И. П. Павлова РАН  
наб. Макарова, 6, Санкт-Петербург, 199034, Россия  
*E-mail:* [dinomama@yandex.ru](mailto:dinomama@yandex.ru)  
ORCID: 0000-0002-5041-2498

<sup>3</sup> Университет МГУ-ППИ в Шэньчжэне  
Международный Университетский проспект, 1, Шэньчжэнь, 518172, Китай  
*E-mail:* [1608029882@qq.com](mailto:1608029882@qq.com)  
ORCID: 0009-0002-1146-6634

*Статья поступила 08 мая 2026 г.; принята 15 июня 2026 г.;  
опубликована 30 июня 2026 г.*

**Аннотация:** Исследование посвящено количественному и качественному сопоставлению эвиденциальных стратегий в письменных русскоязычных нарративах носителей и неносителей языка. Несмотря на значительное число работ, посвященных эвиденциальности в типологической перспективе и ее усвоению детьми на родном языке, остается открытым вопрос о том, как взрослые реализуют эвиденциальные значения в письменной речи на неродном языке. Цель работы – выявить, как различается распределение эвиденциальных значений и средств их выражения в двух группах. Исследование выполнено в парадигме функционально-семантического подхода с применением статистических количественных методов. Материалом послужили 60 текстов, созданных носителями китайского языка, и 39 текстов носителей русского языка, собранных с помощью инструмента MAIN (Multilingual Assessment Instrument for Narratives). Было проанализировано распределение трех основных эвиденциальных значений (прямая, инференциальность, пересказывательность), а также их вербальная репрезентация на уровне конкретных лексико-грамматических средств. Установлено, что общая доля эвиденциально маркированных предложений в обеих группах сопоставима (~89%), однако распределение типов значений статистически значимо различаются. В текстах китайских студентов доминирует прямая эвиденциальность, что свидетельствует о преимущественно дескриптивной стратегии, ориентированной на непосредственное описание наблюдаемых событий. В нарративах носителей преобладает инференциальность, реализуемая через активное использование глаголов психического состояния и восприятия от лица персонажа, а также оценочных маркеров. Пересказывательность практически отсутствует в группе неносителей, тогда как

носители используют ее для передачи диалогичности повествования. Таким образом, взрослые инофоны успешно переносят на неродной язык базовые когнитивные механизмы отслеживания источника информации, однако их эвиденциальная компетенция на уровне А2 реализуется в упрощенной стратегии, характеризующейся ориентацией на описание. Полученные результаты имеют прикладное значение для преподавания РКИ: развитие интерпретативных навыков и расширение репертуара эвиденциальных средств должно стать целью продвинутых этапов обучения.

**Ключевые слова:** Эвиденциальность; Эвиденциальные стратегии; Нарратив; Русский язык как иностранный; Освоение второго языка; Письменная речь

**Благодарности:** Исследование выполнено при поддержке гранта РФФИ № 25-18-00938 «Эвиденциальные стратегии в свете корпусных и экспериментальных данных (на материале разноструктурных языков)».

**Информация для цитирования:** Краснощечкова С. В., Галкина Е. В., Ван И. Распределение эвиденциальных значений в заданных нарративах на родном и неродном русском языке // Научный результат. Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 2. С. 139–163. DOI: 10.18413/2313-8912-2026-12-2-0-6

УДК 811.161.1'243

DOI: 10.18413/2313-8912-2026-12-2-0-6

Sofia V. Krasnoshchekova<sup>1</sup>  
Elena V. Galkina<sup>2</sup>  
Wang Yiling<sup>3</sup>

**Distribution of Evidential Meanings in Elicited Narratives  
in Native and Non-Native Russian**

<sup>1</sup> Institute for Linguistic Studies, RAS,  
9 Tuchkov Ln., Saint Petersburg, 199053, Russia  
*E-mail:* krasnoshchekova\_sv@iling.spb.ru  
ORCID: 0000-0001-8830-5121

<sup>2</sup> Pavlov Institute of Physiology, RAS,  
6 Makarov Emb., St. Petersburg, 199034, Russia  
*E-mail:* [dinomama@yandex.ru](mailto:dinomama@yandex.ru)  
ORCID: 0000-0002-5041-2498

<sup>3</sup> Shenzhen MSU-BIT University,  
1 International University Park Rd., Shenzhen, 518172, China  
*E-mail:* [1608029882@qq.com](mailto:1608029882@qq.com)  
ORCID: 0009-0002-1146-6634

*Received 08 May 2026; accepted 15 June 2026; published 30 June 2026*

**Abstract:** The study provides a quantitative and qualitative comparison of evidential strategies in written Russian narratives produced by native and non-native speakers. Despite a substantial body of research on evidentiality from a typological perspective and its first language acquisition by children, the question of how adult learners implement evidential meanings in written L2 production remains open. The aim is to identify differences in the distribution of evidential meanings and their linguistic

encoding across the two groups. The research is conducted within a functional-semantic framework and employs quantitative statistical methods. The material consists of 60 texts produced by Chinese learners of Russian (A2 proficiency) and 39 texts by native Russian speakers, collected using the Multilingual Assessment Instrument for Narratives (MAIN). The distribution of three main evidential meanings (direct evidentiality, inference, hearsay) was analyzed, as well as their verbal realization at the level of specific lexico-grammatical markers. While the overall proportion of evidential clauses is comparable across groups (approximately 89%), the distribution of evidential types differs significantly. Chinese learner narratives are dominated by direct evidentiality, reflecting a predominantly descriptive strategy focused on the immediate description of observable events. Native speaker narratives, in contrast, are characterized by a predominance of inference, expressed through the frequent use of mental state verbs, perception verbs attributed to characters, and evaluative markers. Hearsay is virtually absent in the non-native group, whereas native speakers employ it to convey narrative dialogicity. Adult L2 learners successfully transfer basic cognitive mechanisms of source monitoring to the target language, but at the A2 level their evidential competence manifests in a simplified strategy characterized by a focus on description. The findings have direct pedagogical implications for teaching Russian as a foreign language: developing interpretative skills and expanding the repertoire of evidential devices should be a priority at advanced stages of learning.

**Keywords:** Evidentiality; Evidential strategies; Narrative; Russian as a second language; Second language acquisition; Written speech

**Acknowledgements:** The study was supported by the Russian Science Foundation, project no. 25-18-00938 “Evidential Strategies in Light of Corpus and Experimental Data (Based on Structurally Diverse Languages)”.

**How to cite:** Krasnoshchekova, S. V., Galkina, E. V., Wang, Y. (2026). Distribution of Evidential Meanings in Elicited Narratives in Native and Non-Native Russian, *Research Result. Theoretical and Applied Linguistics*, 12 (2), 139–163. DOI: 10.18413/2313-8912-2026-12-2-0-6

## 1. Введение

Область эвиденциальности в языке объединяет любые значения, так или иначе связанные с отсылкой к источнику содержащейся в высказывании информации. Информация может быть непосредственно воспринята говорящим или получена им от третьих лиц. В составе эвиденциальности, таким образом, обычно выделяется три основных значения: прямая эвиденциальность (информация получена говорящим при помощи органов чувств), инференциальность (информация получена при помощи логического умозаключения) и пересказывательность, или hearsay (информация получена от третьих лиц). В языках мира типичные эвиденциальные значения по-разному объединяются в кластеры при выражении

грамматическими или лексическими средствами (Aikhenvald, 2004; Козинцева, 2007; Постникова, 2020): так, выделяют противопоставление значений личного / неличного доступа (прямая эвиденциальность + инференциальность vs пересказывательность) и прямого / непрямого доступа (прямая эвиденциальность vs инференциальность + пересказывательность) (Плунгян, 2011).

Категориальное понимание термина «эвиденциальность» было впервые предложено Р. О. Якобсоном в середине XX века (Jakobson, 1957), однако активное изучение грамматической эвиденциальности начало вестись в последние тридцать лет, преимущественно в работах типологического направления (Aikhenvald, 2004; Козинцева, 2007;

Плунгян, 2011). В настоящее время широко исследуются такие вопросы, как объем категории «эвиденциальность» (Tournadre, LaPolla 2014, Verhees, 2019) типология эвиденциальных систем (Brugman, Macaulay, 2015; Постникова, 2020), спектр эвиденциальных средств и стратегий в языках мира (Храковский, 2007; Mushin, 2013; Nuckolls, Michael, 2014; Kittilä 2019), взаимодействие эвиденциальности с категориями модальности (Храковский, 2007; Cornillie, 2009; Kosta, Iskakova, 2022) и авторитетности (Астахова, 2014), дейктические и референциальные аспекты эвиденциальных единиц (Peterson, 2016). В нарративе эвиденциальность выполняет в том числе дейктическую функцию: эвиденциальные маркеры позволяют говорящему построить перспективу высказывания и подать информацию с разных точек зрения (Mushin, 2000).

В русском языке эвиденциальность выражается в первую очередь лексически. Выделяют такие группы эвиденциальных единиц, как вводно-модальные слова и конструкции со значением предположительности и выводности (*видимо, по-видимому, вероятно, возможно, наверно, может быть, должно быть; очевидно, стало быть, следовательно, значит*), частицы и дискурсивные слова, маркирующие чужую речь и пересказ с разной степенью уверенности (*якобы, мол, дескать, будто бы, вроде*), репортативные и цитативные конструкции (*говорят, по слухам, по сообщению X, по словам X*), глаголы речи и сообщения (*сказать, объявить*), глаголы восприятия и психических процессов от 1-го лица (*видеть, слышать, думать, считать*), а также безличные и пассивно-предикативные конструкции (*считается, предполагается, можно полагать*) (Кобрина, 2005; Болдырев, 2006; Григоренко, 2009; Шестухина, 2009; Страхова, 2016). Кроме того, эвиденциальные значения могут выражаться синтаксически и контекстуально, а также полностью

имплицитно (Никитина, 2013; Падучева, 2013).

В то время как значительное число работ посвящено освоению грамматической категории эвиденциальности и лексических средств ее выражения ребенком в родном языке (Aksu-Koç, 2000; De Villiers et al., 2009; Rett, Hyams, 2014), а также развитию когнитивных структур, позволяющих отслеживать источник информации (Ozturk, Papafragou, 2016), малоизученной остается проблема освоения эвиденциальности в неродном языке. Известны работы на материале японского, турецкого, английского и французского языков. Так, в серии статей (Ishida, 2006, Narita, 2012) было показано, что взрослые носители английского, изучающие японский язык, реже используют маркеры пересказываемости, чем носители японского, и допускают ошибки. В работах, посвященных порождению и восприятию граммем прямой и косвенной эвиденциальности англо-турецкими и немецко-турецкими билингвами, доказывалось, что прямая эвиденциальность по сравнению с косвенной вызывает больше затруднений в экспериментах с регистрацией движений глаз, но лучше понимается в заданиях на перевод (Arslan et al., 2015; Kaya-Soykan et al., 2023; Tosun, Filipović, 2025). Экспериментальное исследование, в котором носителям английского и французского языка предлагалось создать на неродном языке пересказ просмотренного мультфильма (Leclercq, Mélas, 2021), продемонстрировало, что носители французского, изучающие английский, более эксплицитно используют эвиденциальные маркеры, чем носители английского, изучающие французский, и чем обе группы носителей, порождавшие нарратив на родном языке, при этом активнее всего употреблялись маркеры инференциальности 'я думаю', 'может быть', 'как видно' и прямой эвиденциальности 'я вижу'.

Так как эвиденциальность является относительно «молодым» лингвистическим понятием по сравнению, например, с аспектуальностью или модальностью, данная область в аспекте преподавания русского языка на настоящий момент разработана не так подробно, как другие. Фактически, освоение эвиденциальности в русском языке как иностранном (РКИ) до настоящего времени почти не попадало в поле внимания исследователей. Единственной работой, в которой явно связывается эвиденциальность и РКИ, на настоящее время является доклад (Борисова, 2025), посвященный эвиденциальным конструкциям, актуальным в курсе обучения РКИ на разных уровнях, и выражаемым с их помощью коммуникативным значениями (A1 – *конечно, я думаю, мне нравится*; A2–B1: *мне кажется, я слышал, если не ошибаюсь*).

## 2. Цель исследования

Настоящее исследование посвящено эвиденциальным значениям в письменной речи на русском языке как иностранном. Цель состоит в выявлении наиболее типичных способов передачи эвиденциальных значений в нарративах, созданных на основе сюжетных рисунков носителями китайского языка, изучающими русский, и определении различий эвиденциальных стратегий, характерных для носителей и неносителей русского языка. Под эвиденциальным значением мы имеем в виду способ получения говорящим сообщаемой информации, выраженный или не выраженный эксплицитно (источником является восприятие говорящего – прямая эвиденциальность, умозаключение говорящего – инференциальность, чужие слова – пересказывательность). Языковое средство, при помощи которого эксплицитно выражается такое значение, мы называем вербальным эвиденциальным маркером. Языковое средство, которое, наряду с номинативным / диктумным,

содержит периферийный эвиденциальный компонент либо способно приписать высказыванию эвиденциальное значение только в конкретном контексте, мы называем в рабочем порядке эвиденциальным контекстным индикатором. Под эвиденциальной стратегией мы понимаем предпочтительный способ конструирования нарратива в целом с точки зрения приписывания высказываниям эвиденциальных значений и использования соответствующих маркеров.

Гипотеза исследования заключается в том, что все ситуативно обусловленные эвиденциальные значения так или иначе будут выражены в тексте независимо от уровня владения автором русским языком, тогда как средства их выражения будут различаться. Основанием для гипотезы служит тот факт, что эвиденциальность, в том числе выражаемая в нарративе, вызывает затруднения у ребенка, осваивающего родной язык, из-за неполной сформированности когнитивных механизмов, позволяющих успешно отслеживать источник информации (De Villiers, Garfield, 2009; Ozturk, Papafragou, 2016; Ünal, Papafragou, 2020), тогда как у взрослого индивида данные механизмы уже сформированы. Предположительно, это позволяет взрослым инофонам успешно переносить на изучаемый язык когнитивные принципы понимания эвиденциальных значений, освоенные в детстве на базе родного языка.

Опираясь на парадигму функциональной грамматики (ФГ) А. В. Бондарко, мы рассматриваем эвиденциальность как одну из функционально-семантических категорий, или функционально-семантических полей (ФСП), русского языка, что позволяет вести анализ «от значения к форме» (Бондарко, 2011: 59–65), ориентируясь в первую очередь на семантику языковых единиц. Вслед за ФГ мы принимаем, что возможны высказывания, в которых

грамматическая семантика не выражается при помощи конкретных вербальных маркеров, но подразумевается говорящим либо существует априори и, соответственно, считывается адресатом (ср. безглагольные высказывания, тем не менее передающие значение того или иного вида или времени (Бондарко, 2011)). Эвиденциальность, как и другие функционально-семантические категории (например, временная локализованность или темпоральность), охватывает широкий пласт существующих и потенциальных высказываний. По этой причине любое высказывание, так или иначе имеющее отношение к одному из основных способов получения информации, считается нами имеющим эвиденциальное значение, выраженное сильнее или слабее.

Заметим также, что настоящая работа придерживается «комбинированного» подхода к вопросу взаимодействия категорий эвиденциальности и модальности (Wiemer, 2010; Козловский, 2022; Иванова, Хакимова, 2024), согласно которому эвиденциальность и модальность являются принципиально разными категориями, однако при этом тесно взаимодействуют. Существуют высказывания, в которых сложно четко разграничить эвиденциальное и модальное значения, и эвиденциальные языковые средства, содержащие модальные компоненты (Сафина, 2016). Эвиденциальность может пересекаться с такими модусными категориями, как эпистемическая модальность, оценочность, эмотивность, миративность, авторизация. Именно поэтому в работе проводится также анализ модусных маркеров.

### 3. Материал и методы

Материалом исследования послужили письменные нарративы, собранные при помощи инструмента MAIN (Multilingual Assessment Instrument for narratives – Мультиязычный инструмент оценки нарративов (Gagarina et al., 2019)). Инструмент был разработан для оценки

уровня развития нарративной компетенции детей-билингвов, однако в дальнейшем получил более широкое распространение и в настоящее время применяется также при работе с взрослыми информантами: как носителями стандартного или херитажного варианта родного языка, так и неносителями, осваивающими иностранный язык. Комплект экспериментальных материалов включает четыре набора из шести рисунков, в каждом из которых представлен связный сюжет с участием персонажей – детей и животных. Испытуемым предъявляются рисунки и предлагается создать по ним устный или письменный рассказ. Два набора предназначены для создания пересказов и два – собственных рассказов.

Экспериментальную группу составили тексты на русском языке, написанные студентами – носителями китайского (КН, «китайские нарративы»). Студенты обучаются в университете МГУ-ППИ в г. Шэньчжэнь (Китай) и изучают русский с китайским преподавателем. Уровень владения русским языком оценивается в среднем как А2. Сбор нарративов проводился летом 2025 г. Эксперимент был замаскирован под учебное задание: на экране в аудитории демонстрировался набор из шести рисунков на одном листе, и студентам предлагалось написать рассказ по предъявленным рисункам. Для эксперимента был выбран сюжет «Птенцы», предназначенный для создания собственных рассказов. Всего было получено 60 нарративов, которые в дальнейшем были переведены в цифровой вид и размечены.

В качестве контрольной группы были использованы нарративы, написанные на русском языке носителями – студентами Санкт-Петербургского Политехнического университета (РН, «русские нарративы»). Сбор нарративов проводился в 2017–2018 г. в рамках проекта, посвященного нарративной компетенции в русском как родном и иностранном языке

(Krasnoshchekova, Kashleva, 2019), в результате которого была сформирована база текстов. Эксперимент проводился в аудитории и не маскировался под задание. Студентам раздавались листы с напечатанными рисунками и предлагалось написать на их основе короткую историю. Из базы были отобраны тексты, созданные по наборам рисунков, предполагающих порождение собственного рассказа, – «Птенцы» и «Козлята». Общее число текстов составило 39. В ходе настоящего исследования была проведена их разметка с учетом эвиденциальных маркеров и значений.

Разметка осуществлялась двумя авторами независимо на 20% текстов (случайная выборка). Коэффициент каппа (κ) Коэна составил 0,84, что указывает на высокий уровень согласия. Остальные тексты размечены первым автором с последующей выборочной проверкой вторым. В спорных случаях решение принималось коллегиально по результатам обсуждения. Пример рабочей разметки приводится в Таблице 1: буквенные обозначения выбирались случайным образом. Теоретические основания классификации и подробное описание эвиденциальных значений и языковых средств их выражения см. ниже, в п. 4.

**Таблица 1.** Примеры эвиденциальной разметки.

**Table 1.** Examples of the evidential annotation.

Размеченное предложение	Использованные метки
Я увидел Vv1 такую историю. <sup>1</sup> D	D – прямая эвиденциальность Vv1 – глагол визуального восприятия от лица говорящего
Добрый собоко увидел Vv3 и хотел Vr помочь Vc них. I	I – инференциальность Vv3 – глагол визуального восприятия от лица персонажа Vr – глагол психического состояния Vc – каузативный глагол
Ведите себя хорошо! Я скоро прилечу. Сказав Vs это она улетела. H	H – пересказывательность Vs – глагол говорения
Доброе утом! Добрый день!	– отсутствие эвиденциального значения
Будем верить, что Вороне воздастся за ее героизм, а маленькой козлиной семье больше не будут грозить напасти (? I)	– спорный случай: умозаключение автора (инференциальность) или дополнительное замечание (отсутствие эвиденциального значения)? Решение по результатам обсуждения: инференциальность.

<sup>1</sup> Все примеры приводятся в авторском написании.

Средняя длина текстов в словах составила 53 в КН, 99 – в РН; средняя длина в предложениях – 10 в обеих группах; совокупное число предложений во всех текстах группы – 617 в КН и 398 в РН.

Для оценки значимости различий между группами КН и РН по распределению эвиденциальных значений и средств их выражения были использованы критерии Шапиро-Уилка, хи-квадрат ( $\chi^2$ ) Пирсона и U Манна-Уитни. Расчеты выполнены в программе JASP (версия 0.18.3.0). При проведении теста на  $\chi^2$  для каждой ячейки таблиц сопряженности вычислялись стандартизованные остатки; значения  $|\text{остаток}| > 1,96$  считались статистически значимыми на уровне  $p < 0,05$ . При проведении теста Манна-Уитни различия между группами также считались значимыми на уровне  $p < 0,05$ . Для графического представления результатов использовалась среда Google Colab с языком программирования Python. Построение диаграмм выполнено с помощью библиотек matplotlib, seaborn, wordcloud, networkx и pandas; количественные расчеты – с использованием numpy.

#### *Ограничения исследования*

Авторы признают, что экспериментальные группы различаются не только по статусу носителя / неносителя языка, но и по ряду других параметров, таких как год сбора данных (РН – 2017–2018, КН – 2025), стимульный материал (для КН использовался только сюжет «Птенцы», для РН – «Птенцы» и «Козлята»), а также условия проведения (для КН эксперимент был замаскирован под учебное задание, для РН – нет). Хотя маловероятно, что эти различия могли систематически сдвинуть результаты в сторону обнаруженных нами закономерностей, в будущем желательно провести репликацию на материале, собранном в едином протоколе. При этом сопоставление представляется

приемлемым по следующим причинам: оба протокола MAIN для порождения собственного рассказа созданы по единому образцу с соблюдением одинакового состава ролей персонажей, их действий, взаимного расположения на рисунках, ситуации на каждом рисунке в наборе, детальной канвы сюжета; сбор нарративов проводился в аудитории в письменном виде.

#### **4. Результаты и дискуссия**

Категория эвиденциальности в рамках данного исследования рассматривается в функционально-семантическом аспекте, вследствие этого при разметке и дальнейшем анализе учитывались не только эксплицитные средства выражения эвиденциальных значений, но и имплицитно передаваемые значения как таковые.

В настоящей работе, как уже было указано, под *эвиденциальным значением* мы понимаем семантическую категорию, приписываемую предложению, под эвиденциальным вербальным языковым средством, или *эвиденциальным маркером*, – конкретную языковую единицу, выражающую это значение, а под *эвиденциальным контекстным индикатором* – языковую единицу, косвенно указывающую на наличие эвиденциальной семантики в высказывании в рамках конкретной речевой ситуации. Заметим, что анализ маркеров и контекстных индикаторов носит предварительный и рабочий характер, а предложенный набор средств не претендует на классификацию эвиденциальных единиц русского языка в целом.

#### *Распределение эвиденциальных значений*

С учетом расширительного подхода к эвиденциальности (п. 2), была принята следующая методология. Значение прямой эвиденциальности отмечалось во всех случаях, где автор нарратива описывал непосредственно происходящее на рисунке: ‘птица улетела’, ‘кот прибежал’,

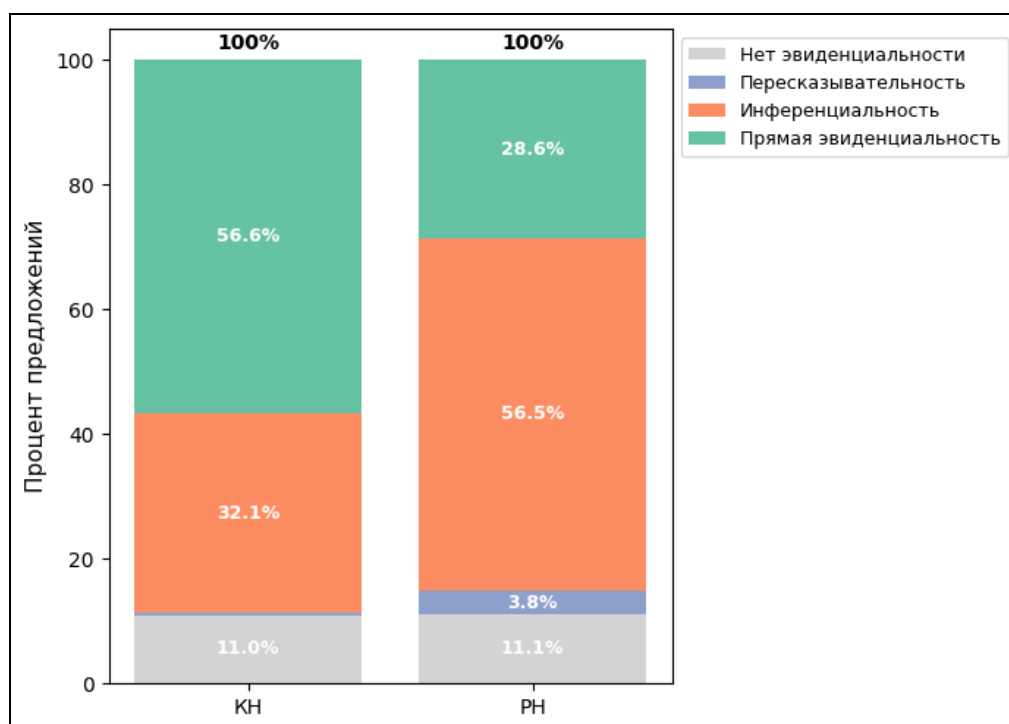
т. е. создавал высказывание с опорой на наблюдаемый визуальный стимул (*Собака тащила кот*). Инференциальное значение фиксировалось там, где говорящий выходил за пределы изображенного на рисунке, делая собственные выводы о событиях, отношениях между героями и их внутреннем состоянии (*Папа-птица (и собака) любил 2 молодой птицы*). Инференциальность отмечалась также при обозначении родственных связей героев: большая птица – ‘птица-мама’ (рисунок не указывает на родство героев – следовательно, говорящий делает вывод об этом на основе их внешнего вида и поведения); выражении ментальных процессов; указании на визуальное восприятие: ‘один герой увидел другого’ (говорящий делает вывод об этом на основе взаимного расположения героев на рисунке). Пересказывательное значение фиксировалось там, где говорящий приписывал герою какие-либо реплики

(«Спасибо большое, наш друг, ты же наш защитник!» – говорило семейство птичек). Отсутствие эвиденциальности отмечалось по остаточному принципу – в тех высказываниях, которые не отражали происходящее на рисунке (непосредственно наблюдаемое или трактуемое автором) и реплики героев и содержали дополнительные сторонние замечания, обращения к читателю и др. (*Доброе утом! Добрый день!; Мамы, будьте бдительны!*).

Разметка эвиденциальных значений проводилась на уровне высказывания. На Рисунке 1 представлено распределение эвиденциальных значений в группах текстов КН и РН. В каждом случае здесь и далее указываются относительные кумулятивные данные в процентах – доля предложений с анализируемым явлением во всех текстах группы относительно совокупного числа предложений во всех текстах группы.

**Рисунок 1.** Эвиденциальные значения в нарративах.

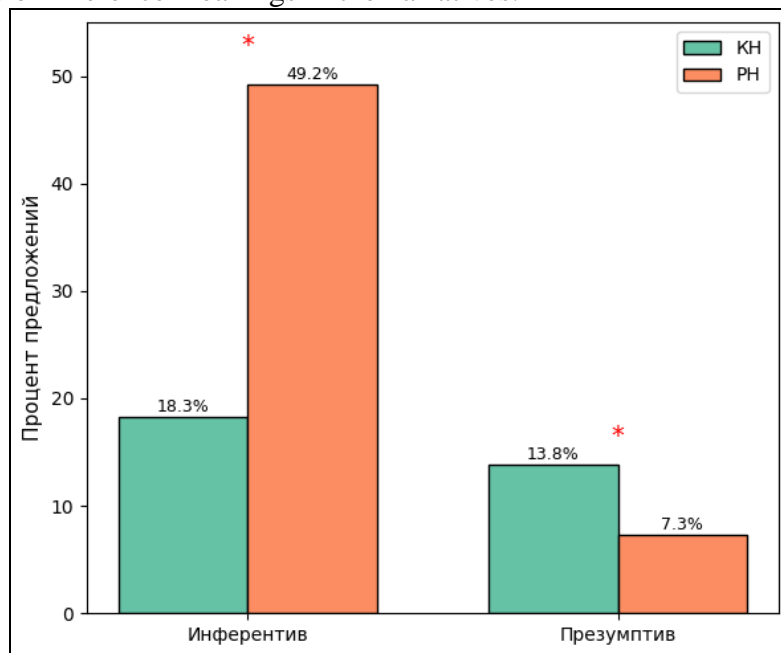
**Figure 1.** Evidential meanings in the narratives.



Общее число предложений с эвиденциальным значением сопоставимо в обеих группах и составляет около 89%, однако распределение типов значения различается. Различия статистически значимы ( $\chi^2 = 18,257$ ,  $df = 2$ ,  $p\text{-value} < 0,001$ ). В группе КН основным значением является прямое эвиденциальное (56,6%), тогда как в группе РН оно находится на втором месте (28,6%). Напротив, инференциальное значение занимает второе место в группе КН (32,1%) и является ведущим в группе РН (56,5%). Пересказывательное значение представлено в основном только в группе

РН (3,8%) и маргинально в группе КН (0,3%). Кроме того, фиксировались два подтипа инференциальности – инферентив, или суждение о ситуации на основе размышлений, и презумптив, или умозаключение о ситуации на основе фоновых или общих знаний (Рисунок 2). Презумптивное значение более характерно для КН (13,8%), чем для РН (7,3%), тогда как инферентивное крайне частотно в РН (49,2%) и значительно менее частотно в КН (18,3%). Красным астериском на диаграмме отмечены статистически значимые расхождения ( $\chi^2 = 48,277$ ,  $df = 2$ ,  $p\text{-value} < 0,001$ ).

**Рисунок 2.** Подтипы инференциальных значений в нарративах.  
**Figure 2.** Subtypes of inference meanings in the narratives.



Таким образом, при порождении текста на неродном языке с большей степенью активности выражается информация, воспринятая говорящим непосредственно, и менее типичным является формулирование выводов и умозаключений.

*Распределение эвиденциальных маркеров и контекстных индикаторов*

Вербальные средства выражения эвиденциальных значений фиксировались на уровне словоформы. При разметке и анализе, как указано выше, проводилось

различие между эвиденциальными маркерами и контекстными индикаторами. Маркерами считались те, которые выражают эвиденциальное значение как основное (глаголы визуального восприятия – прямая эвиденциальность; глаголы говорения – пересказывательность); контекстными индикаторами – те, которые косвенным образом указывают на наличие эвиденциального значения у высказывания в рамках конкретного контекста или ситуации (глаголы внутреннего состояния,

визуального восприятия от лица персонажа – инференциальность).

Частотными эвиденциальными маркерами, как отмечалось во Введении, в русском языке являются вводные конструкции: *кажется, по словам / по сообщению X*, глаголы говорения, мысли и визуального восприятия и однокоренные им слова: *X сказал; я думаю, видно, что*. Однако разметка и анализ полученных в эксперименте нарративов показали, что в нашем материале отсутствует большая часть типичных русских эвиденциальных маркеров, за исключением глаголов говорения и визуального восприятия от 1-го лица. Предположительно, для данного дизайна – заданного сюжетного нарратива по набору рисунков, исключающего догадки или домысливание сюжета, – использование подобных маркеров нерелевантно. Именно в связи с этим внимание было обращено на периферийные и контекстные средства: такие языковые единицы, которые содержат компонент эвиденциального значения наряду с диктумным (номинативным) или присваивают его в конкретном контексте. В нашем случае эвиденциальный компонент проявляется в аспекте выполняемого задания – создания рассказа по картинкам, которые находятся перед глазами информанта. В высказываниях, порожденных в других речевых ситуациях, например в спонтанном диалоге, этот компонент отсутствует или может отсутствовать. Такие единицы не являются эвиденциальными маркерами в собственном смысле, а могут рассматриваться только как элементы диктумного содержания, на основании которых исследователь реконструирует способ интерпретации ситуации информантом. Именно поэтому для их обозначения был принят рабочий термин «контекстный индикатор».

Таким образом, в нарративах были выявлены следующие эвиденциальные средства:

1) прямая эвиденциальность: маркеры – глаголы восприятия от лица говорящего; контекстные индикаторы – глагол *появиться*; авторские ремарки;

2) инференциальность: контекстные индикаторы – союзы логической связи, глаголы психического состояния, относящиеся к персонажу; глаголы восприятия от лица персонажа; каузативные глаголы;

3) пересказывательность: маркеры – глаголы говорения; репортативные конструкции.

Остановимся подробнее на единицах со спорным статусом, которые мы в рабочем порядке относим к контекстным индикаторам. Глагол *появиться* указывает на значение прямой эвиденциальности потому, что он отсылает к визуальному восприятию информантом нового персонажа на рисунке.

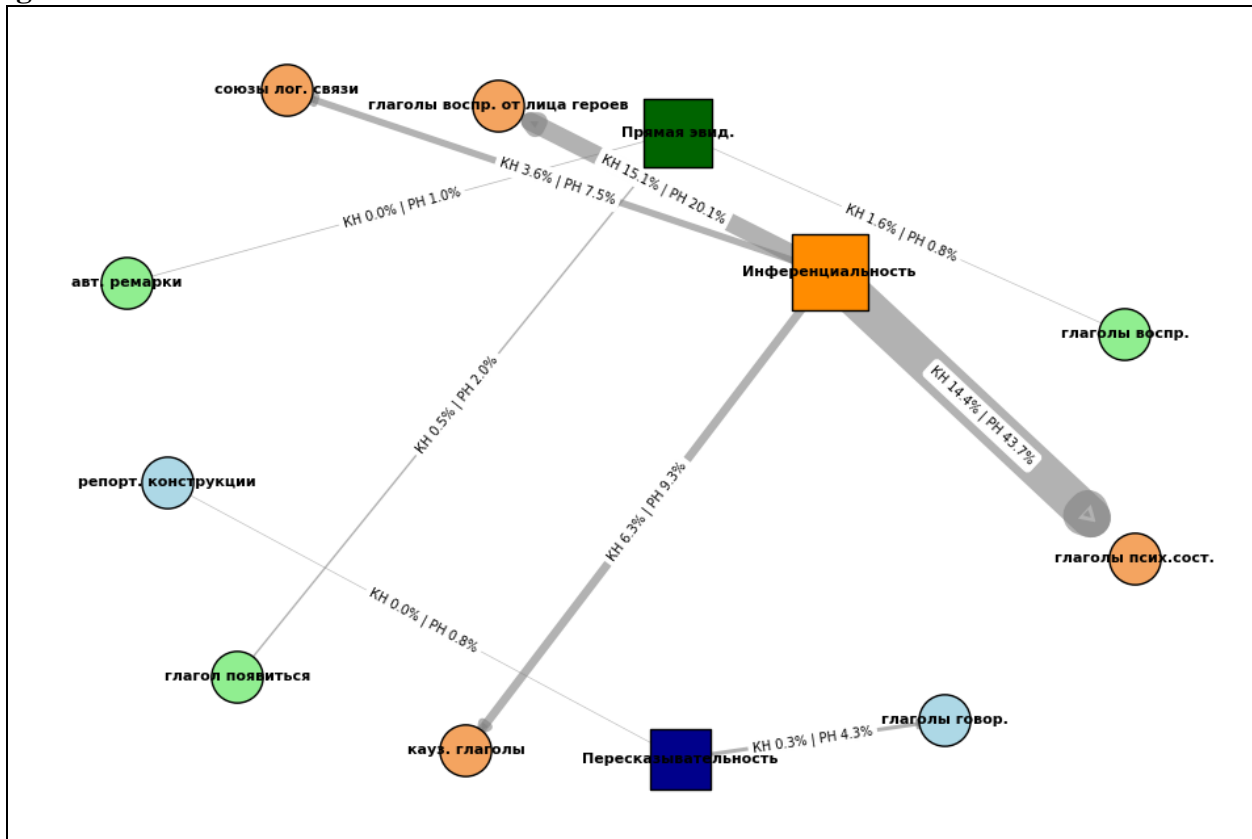
Союзы логической связи указывают на то, что автор делает вывод о причинах, следствиях, целях и иных логических характеристиках поступков персонажа. Глаголы психического состояния и визуального восприятия от третьего лица могут не иметь эвиденциального компонента в спонтанной речи, однако он появляется в рамках экспериментального задания. Строго говоря, на рисунке не изображено, что видит герой, что он решил или какую эмоцию он испытывает: говорящий делает об этом вывод на основе неких изображенных художником подсказок: взаимного расположения героев, их мимики и жестов и др. То же касается каузативных глаголов типа *знать*: на рисунке изображено, как собака бежит за кошкой, – испытываемый на основе их позиции делает вывод о том, что собака заставляет кошку бежать. Во всех этих случаях мы фиксировали указание на инференциальное значение.

Распределение зафиксированных групп средств представлено на сетевой диаграмме (Рисунок 3). Толщина линии, соединяющей тип эвиденциального значения и способ его выражения,

указывает на частотность средства. На линиях отображены проценты – доли

предложений с соответствующими средствами в группах КН и РН.

**Рисунок 3.** Эвиденциальные маркеры и контекстные индикаторы в нарративах.  
**Figure 3.** Evidential markers and context indicators in the narratives.

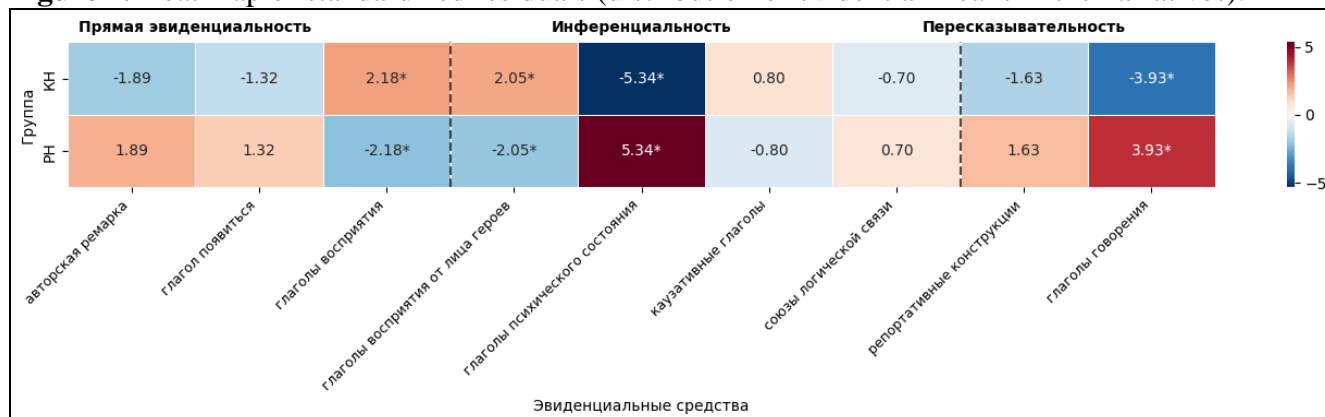


Для проверки значимости различий между группами был проведен анализ таблиц сопряженности с вычислением стандартизованных остатков. По общему распределению типов языковых средств различия являются статистически значимыми ( $\chi^2 = 84,928$ ,  $df = 8$ ,  $p\text{-value} < 0,001$ ), однако при сравнении в рамках конкретных типов наблюдаются как значимые расхождения, так и сопоставимая частотность. На тепловой карте (Рисунок 4) приводятся значения стандартизованных остатков. Цвет

каждой ячейки отражает величину и направление отклонения наблюдаемой частоты от ожидаемой: красный цвет соответствует положительным остаткам (наблюдаемая частота выше ожидаемой), синий – отрицательным (наблюдаемая частота ниже ожидаемой). Интенсивность цвета прямо пропорциональна модулю остатка. Астериском отмечены ячейки, для которых  $|\text{остаток}| > 1,96$ , что соответствует статистически значимому различию на уровне  $p\text{-value} < 0,05$ .

**Рисунок 4.** Тепловая карта стандартизированных остатков (распределение средств выражения эвиденциальности в нарративах).

**Figure 4.** Heat map of standardized residuals (distribution of evidential means in the narratives).



Таким образом, наиболее сильно группы различаются по использованию глаголов психического состояния и говорения (по обоим критериям показатели РН значительно выше, чем показатели КН). Кроме того, наблюдаются различия в частотности глаголов восприятия от лица говорящего и персонажей (показатели КН выше, чем РН). Другие эвиденциальные средства употребляются носителями и неносителями на сопоставимом уровне.

Для проверки достоверности статистических различий были проведены дополнительные тесты: сравнение частотностей различных средств относительно общего количества слов в текстах; сравнение показателей по каждому тексту в группе индивидуально, а не кумулятивно.

*Выравнивание относительно общего количества слов в текстах*

При одинаковой средней длине в предложениях, нарративы в группах КН и РН существенно различаются длиной в словах: 55 в КН и 99 в РН. То, что на родном языке носители порождают более длинные и сложные предложения, может обуславливать использование ими большего количества эвиденциальных средств разных групп. Для проверки этого предположения было проведено выравнивание частотных показателей относительно общего количества слов во всех текстах группы (Таблица 2) и дальнейшая проверка статистической значимости различий распределения маркеров в группах КН и РН.

**Таблица 2.** Эвиденциальные маркеры и контекстные индикаторы в нарративах (относительно общего количества слов).

**Table 2.** Evidential markers and context indicators in the narratives (by the total number of words).

значение	средство	КН	РН
прямая эвиденциальность	глаголы восприятия	0,3%	0,1%
	глагол <i>появиться</i>	0,1%	0,2%
	авторская ремарка	0	0,1%
инференциальность	союзы логической связи	0,7%	0,7%
	глаголы психического состояния	2,8%	3,9%
	каузативные глаголы	1,2%	0,8%
	глаголы восприятия от лица героев	2,9%	1,8%
пересказывательность	глаголы говорения	0%	0,4%
	отсылка к общему знанию	0%	0,1%

В данном случае статистический анализ не выявил значимых различий ( $\chi^2 = 2.264$ ,  $df = 8$ ,  $p = 0, 0.987$ ). Это может быть связано как с тем, что, действительно, носители и неносители используют эвиденциальные маркеры и контекстные индикаторы примерно в одинаковом соотношении, однако неносителям недоступно порождение сложных текстов, как и с малыми числовыми показателями. При распространении эвиденциального значения на все предложение, с чем связан широкий охват массива высказываний эвиденциальной семантикой, вербальные эвиденциальные средства, в том числе контекстные, являются достаточно низкочастотными на фоне строго номинативных, дейктических и др. Именно поэтому изначально был выбран метод выравнивания относительно общего числа предложений. Тем не менее, проведенный анализ позволил уточнить полученные выше результаты и дополнить выводы о характере различий в использовании эвиденциальных маркеров и контекстных индикаторов. Даже при отсутствии значимых различий, в количественном отношении сохраняется та же направленность, которая была выявлена в анализе на уровне предложений (см. рисунок 4): так, глаголы психического состояния используются чаще в РН (3,9% против 2,8% в КН), а глаголы восприятия от лица говорящего и от лица персонажа – чаще в КН (говорящий: 0,3% в КН и 0,1% в РН, персонаж: 2,9% в КН и 1,8% в РН).

Эта устойчивая картина не может объясняться исключительно большей речевой продуктивностью носителей – иначе все маркеры употреблялись бы активнее в РН. Таким образом, констатируем, что информанты в группе КН больше ориентированы на прямое описание наблюдаемого (прямая эвиденциальность и глаголы восприятия), а в группе РН – на интерпретацию внутренних состояний персонажей (инференциальность и глаголы психического состояния).

*Сравнение данных по каждому тексту в отдельности*

Для сравнения данных были созданы таблицы, в которых для каждого нарратива указывалось число предложений и число слов; абсолютное число предложений с анализируемыми эвиденциальными значениями и конкретных средств; доля этих единиц относительно общего количества слов и общего количества предложений в нарративе. Для начала при помощи теста Шапиро-Уилка было обнаружено, что данные по каждому из параметров не демонстрируют нормального распределения ( $p < 0,01$ ), что обусловило выбор теста U Манна-Уитни для дальнейшей статистической работы. Целью было подтвердить или опровергнуть обнаруженные выше различия между группами КН и РН по использованию эвиденциальных значений и средств их выражения. Результаты анализа приводятся в Таблице 3.

**Таблица 3.** Результаты теста Манна-Уитни.

**Table 3.** Results of the Mann-Whitney test.

	% от предложений		% от слов	
	W	p	W	p
<b>значения</b>				
прямая эвиденциальность	<b>1845.000</b>	<b>&lt; .001</b>	-	-
инференциальность	<b>366.000</b>	<b>&lt; .001</b>	-	-
<b>контекстные индикаторы</b>				
союзы логической связи	913.000	0.078	1192.500	0.078
глаголы психического состояния	<b>381.500</b>	<b>&lt; .001</b>	1354.000	0.069
каузативные глаголы	<b>909.000</b>	<b>&lt; .001</b>	1084.000	0.828
глаголы восприятия от лица героев	972.000	0.296	818.500	0.028

Из рассмотрения были исключены параметры, числовые показатели по которому в каждом отдельно взятом тексте стремятся к нулю: пересказывательное значение; глаголы восприятия от лица автора; авторские ремарки; глагол «появиться»; глаголы говорения; репортативные конструкции. Эти средства периферийны в нарративах данного типа, т.е. их частотность слишком мала для обнаружения эффекта на уровне текстов. Статистически значимые различия фиксировались на уровне  $p < 0,05$ .

Анализ подтверждает полученные выше данные лишь отчасти: статистические различия фиксируются для обоих частотных эвиденциальных значений – в КН и РН высказывания с прямой эвиденциальностью и инференциальностью употребляются с разной частотностью. Что касается средств выражения, то различия совпадают только для частотности глаголов психического состояния (при выравнивании по количеству предложений), однако не зафиксированы для глаголов восприятия. Кроме того, не выявленные ранее различия отмечаются для каузативных глаголов – только при выравнивании по количеству предложений. Это дополнительно подтверждает вывод о том, что носители русского языка активнее используют высказывания с интерпретативными средствами для приписывания персонажам намерений, эмоций и выявления причинно-следственных связей.

*Качественный анализ языковых эвиденциальных средств*

Значение прямой эвиденциальности в нашем материале эксплицитно выражалось глаголами визуального восприятия, контекстуально – при помощи глагола *появиться* и авторских ремарок. Глаголы визуального восприятия (*увидеть*, *(по)смотреть* более характерны для КН, чем для РН. Во всех обнаруженных в КН случаях (10 единиц) предложение с глаголом начинает нарратив: *Я увидел такую историю; Я вижу: Кошка и птичка*

*встретились; Я увидел три птицы, как играть на дереве; Я посмотрела из окна. Три (birds) на (tree), чтобы один мама и два дети.* В РН подобное указание на себя (автора нарратива) встречается только у одного информанта (3 единицы): *Я посмотрел на дерево, и увидел казалось бы обыденную, но крайне занимательную картину <...> он увидел разъяренного пса, для которого не существовало ни пеня птиц, ни человека, наблюдающего за ним.* Глагол *появиться* минимально используется как в РН (8 единиц), так и в КН (3 единицы): *Появилась собака.* Авторские ремарки зафиксированы только в РН (4 случая): *Почему она не досмотрела за своим козленком!* Вербальное выражение прямой эвиденциальности в целом нехарактерно для анализируемых нарративов: в обеих группах абсолютное количество предложений с прямой эвиденциальностью не содержит языковых средств, указывающих на данное значение. Таким образом, прямая эвиденциальность обычно выражается контекстуальными имплицитными средствами.

Значение инференциальности в нашем материале выражается контекстуально при помощи союзов логической связи (*чтобы, потому что* и др.), а также глаголов, описывающих внутреннее психическое состояние персонажей и восприятие от их лица, и каузативных глаголов. В КН наиболее активно употребляются глаголы визуального восприятия от лица героев *смотреть, увидеть, заметить* (93 единицы): *Собака гуляла в парке и смотрела кот; Птица заметила кота вернулась;* и глаголы психического состояния и мыслительных процессов *хотеть, решить, думать; испугаться, любить* (89 единиц): *Кот решил лезти на дерево; Кошка испугалась и убежала.* В РН, наоборот, на первом месте находятся глаголы психического состояния (174 единицы), тогда как глаголы восприятия фиксируются реже (80 единиц). Каузативные глаголы

занимают подчиненное положение в обеих группах. В КН число словоупотреблений каузативных глаголов *помогать*, *знать*, *защищать*, *спасти* составляет 39: *Птица войти это прогоняет его*, в РН – 37. Низкочастотным средством выражения инференциальности в обеих группах являются союзы логической связи, число которых достигает 22 в КН: *Так мама-птица вошел чтобы продукты* и 30 в РН. Таким образом, инференциальность чаще выражается вербально, чем прямая эвиденциальность, однако при создании сюжетного нарратива как носители, так и неносители используют для этого контекстуальные языковые средства.

Выражение пересказывательности также более характерно для РН. Эксплицитным языковым средством выражения кватативной пересказывательности (передача информации, полученной от третьих лиц, с упоминанием источника) являются глаголы говорения, вводящие прямую речь персонажей или присоединяющие придаточное изъяснительное предложение. В КН обнаружено только 2 подобных примера: *Они сказал, что это было очень радовать*. В РН число кватативных конструкций составляет 17. Конструкции репортативной пересказывательности (ссылка на обезличенный источник, сообщение сторонней информации без указания источника) зафиксированы только в РН (3 единицы): *Кот, так называют зверя, не видел ни красоты, ни глубины*. Таким образом, вербальное выражение пересказывательности нетипично при создании нарратива, в

котором третье лицо не является фактическим источником информации, а автор текста приписывает реплики того или иного содержания персонажам, опираясь на воображение, однако носители языка порождают подобные конструкции активнее, чем неносители.

Набор наиболее употребительных конкретных языковых единиц, выражающих эвиденциальное значение или указывающих на него контекстуально, ограничен в нарративах обеих групп, что обусловлено заданностью сюжета. Так, активно используются глаголы восприятия *смотреть* и *видеть*, психического состояния *решать*, *хотеть*, *думать*, каузативы *знать*, *помочь*, союз *чтобы*. На облаках слов (Рисунок 5) представлены 20 первых наиболее часто встречающихся маркеров и контекстных индикаторов в КН и РН: размер и интенсивность цвета соответствует абсолютной частотности средства.

Для КН наиболее характерны глаголы визуального восприятия *видеть* и *смотреть*, тогда как в РН *смотреть* используется в некоторой степени реже, чем *видеть*. Из глаголов психического состояния информанты в группе КН предпочитают *хотеть*, тогда как в РН его место занимает *решить*. Каузативный глагол *знать* одинаково представлен в обеих группах, тогда как *спасти* широко употребляется только в группе РН. Низкая частотность и отсутствие в текстах группы КН таких частотных в текстах РН лексем, как *испугаться*, *заметить*, *наблюдать*, объясняется средним уровнем владения русским языком информантов.

**Рисунок 5.** Наиболее частотные эвиденциальные маркеры и контекстные индикаторы в нарративах.

**Figure 5.** The most frequent evidential markers and context indicators in the narratives.



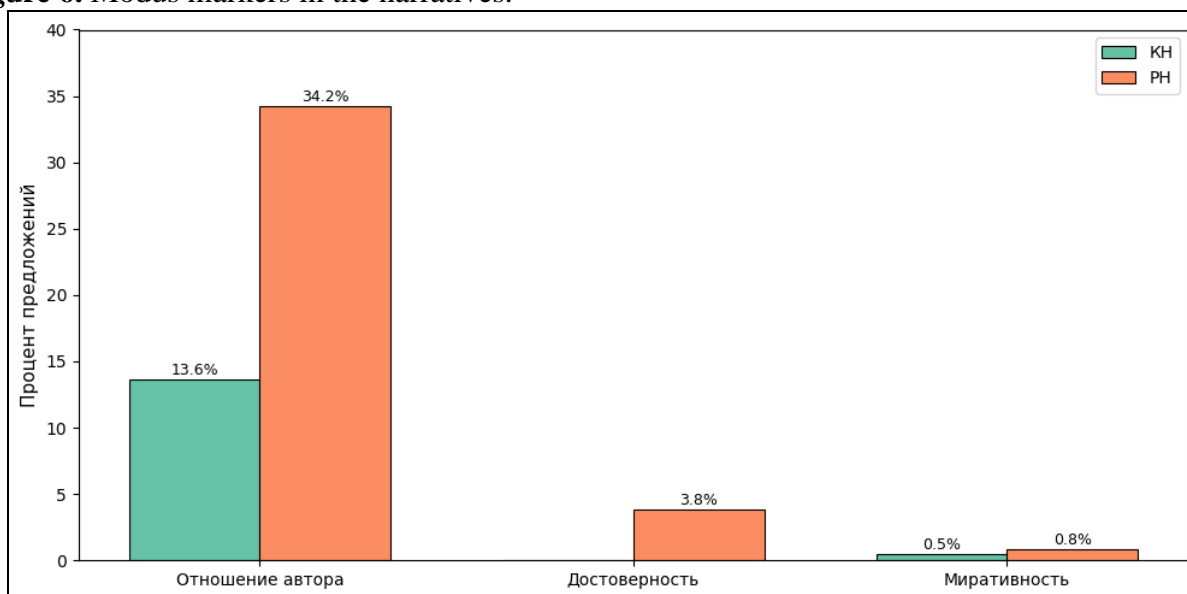
*Модусные маркеры*

В текстах обеих групп были зафиксированы модусные оценочно-эмотивные маркеры, выражающие отношение автора к содержанию высказывания, персонажу или ситуации; эпистемические (или маркеры

достоверности), передающие степень уверенности автора, и маркеры миративности, отражающие удивление автора относительно излагаемых сведений. Данные по их использованию приведены на Рисунке 6.

**Рисунок 6.** Модусные маркеры в нарративах.

**Figure 6.** Modus markers in the narratives.



Общее число таких маркеров в РН (38,7% от общего числа предложений во

всех текстах группы; 3,5% от всего количества слов) выше, чем в КН (14,7% и 2,9%). Эта разница является значимой только на уровне предложений ( $\chi^2 = 75,752$ ,  $df = 2$ ,  $p\text{-value} < 0,001$ ), но не на уровне слов ( $\chi^2 = 2,260$ ,  $df = 1$ ,  $p\text{-value} = 0,133$ ) сравнивались абсолютные цифры). Различие распределения типов маркеров также не имеет статистической значимости ( $\chi^2 = 2,633$ ,  $df = 3$ ,  $p\text{-value} = 0,268$ ).

Основную часть составляют оценочные маркеры, регулярно используемые в нарративах обеих групп. В КН (84 единицы) с их помощью передается этическая оценка персонажа по шкале «хороший – плохой», «добрый – злой»: *Кто (кот) лезет на дерево, чтобы поймать птицу. Он очень злодей. Собака, добрая, тащит кота с дерева; Кот был нехорошо в день.* Также зафиксированы лексемы семантических групп «радость, счастье», описывающие эмоции героев, и «интерес», оценивающие историю в целом: *Плохо кот хотел есть веселая птица; Вот интересной историю!* Спектр эмоций и оценок, передаваемых при помощи данных маркеров в РН (136 единиц), значительно шире: *Вдруг появился «защитник деревни» пес Буран, очень ответственный и воспитанный охранник на своей территории.* В КН в роли оценочных маркеров выступают только прилагательные и наречия, тогда как в РН частотны также вводные конструкции: *К сожалению, его мир совсем отличался от мира птиц.*

Маркеры эпистемической модальности в КН не обнаружены, тогда как в РН достоверность выражается при помощи типичных вводных средств конечно, казалось бы (15 единиц): *Конечно, она полетела искать пищу своим любимым деткам.* Маркеры митаривности крайне периферийны как в КН (3 единицы): *Возле дерева неожиданно появляется большая собака,* так и в РН (3 единицы): *К удивлению, за всем этим наблюдал мудрый ворон.* Таким образом, различия в активности употребления

модусных маркеров носителями и неносителями языка обусловлены ограниченным объемом лексикона на народном языке. С этим связано также то, что в КН полностью отсутствуют вводно-модальные единицы, выражающие этическую, эмоциональную или вероятностную оценку ситуации.

*Обсуждение: сравнение с экспериментом (Leclercq, Mélas, 2021)*

Полученные результаты не совпадают с выводами, сделанными в работе на материале английского и французского языков (Leclercq, Mélas, 2021), в которой было обнаружено, что неносители употребляют эксплицитные эвиденциальные маркеры активнее, чем носители, тогда как носители предпочитают имплицитное выражение. В нашем эксперименте, действительно, неносители сравнительно чаще прибегали к маркеру прямой эвиденциальности *видеть* в 1-м. л. Однако по частотности контекстных индикаторов с эвиденциальным компонентом (относительно общего количества слов) группы не имеют существенных различий. Кроме того, пересказывательность практически отсутствует в нарративах неносителей, тогда как носители прибегают к ней при передаче диалогических фрагментов.

Для понимания причин этих расхождений сопоставим дизайн обоих экспериментов. Сходства заключаются в следующем: (1) в обеих работах сравниваются носители и неносители языка; (2) используется нарративная задача по визуальному стимулу; (3) анализируется употребление вербальных эвиденциальных маркеров и имплицитных способов выражения эвиденциальности. При этом имеются следующие различия: (1) в исследовании Leclercq и Mélas информанты создавали пересказ просмотренного мультфильма, который содержит сюжетные лакуны (отдельные кадры пропущены), что стимулирует домысливание событий и, как следствие,

активное использование эксплицитных маркеров инференциальности (например, ‘я думаю’, ‘может быть’); в нашем эксперименте нарративы порождались как собственные рассказы по серии статичных рисунков, которые остаются перед глазами и позволяют излагать линейный связный сюжет без лагун; (2) языковые пары различны: французский и английский vs. русский и китайский, что может влиять на стратегии переноса; (3) уровень владения носителей в нашем исследовании – А2, в цитируемой работе – В1–С.

Эти различия в дизайне, как мы полагаем, объясняют расхождение в результатах. В условиях сюжетных лагун носители вынуждены эксплицитно маркировать свои предположения, поэтому они используют больше инференциальных маркеров. В нашем эксперименте в линейном сюжете без лагун носители, напротив, отбрасывают необязательную информацию и сосредотачиваются на передаче основного сюжета, опираясь на принцип экономии усилий и уделяя больше внимания собственно языковому оформлению.

### 5. Заключение

Гипотеза о том, что уровень владения русским языком не влияет на выраженность эвиденциальных значений в тексте, но влияет на использование средств их выражения, подтвердилась частично. В текстах как носителей, так и неносителей широко представлены два основных типа эвиденциальных значений: прямая эвиденциальность и инференциальность, причем общее количество эвиденциально нагруженных предложений сопоставимо (около 89%). Это свидетельствует о том, что когнитивный механизм отслеживания источника информации действительно переносится с родного языка. Третье эвиденциальное значение – пересказывательность – фактически отсутствует в текстах неносителей и редко, но все же употребляется носителями. Таким образом, перенос когнитивного механизма актуален для значений личного

доступа ‘увидел сам / сделал вывод’, но не распространяется на приписывание речи персонажам как источник информации об их внутреннем состоянии или намерениях, т.е. значение неличного доступа. Тем не менее, распределение эвиденциальных значений в речи носителей и неносителей существенно различается: в группе КН преобладает значение прямой эвиденциальности, а в группе РН – инференциальности.

Вторая часть гипотезы также подтвердилась только отчасти. Носители русского используют вербальные средства выражения эвиденциальности (в том числе такие маркеры, как глаголы говорения, и такие контекстные индикаторы, как авторские ремарки и глаголы психического состояния) в среднем чаще, чем неносители, в пересчете на общее количество высказываний. Однако фактически владение языком не оказывает решающего влияния на активность использования эвиденциальных средств: сопоставление частотности эвиденциальных маркеров и контекстных индикаторов на уровне общего количества слов в нарративах не демонстрирует статистически значимых различий. Это показывает, что, даже при том что более высокий уровень языковой компетенции носителей проявляется в синтаксической сложности текста, он не может полностью объяснить качественное различие в нарративных эвиденциальных стратегиях. Только на уровне предложения отмечается разница и в частотности средств, смежных с эвиденциальными: модусные оценочные и эпистемические маркеры. Тем не менее, в КН модусная составляющая представлена более бедным набором лексических средств и спектром выражаемых оценок, что связано с лексико-грамматическими ограничениями развивающейся языковой системы информантов.

Назовем стратегию китайских информантов «*дескриптивно-наблюдательной*»: в их нарративах

доминирует прямая эвиденциальность; инференциальность и модусная оценка присутствуют, но выражены слабее; пересказывательность практически отсутствует. Стратегию русских информантов определим как «интерпретативно-оценочную»: в их нарративах доминирует инференциальность; используются модусные средства для передачи различных оттенков оценочности и эмотивности; в ряде текстов присутствует пересказывательность, отражающая диалогичность повествования.

Таким образом, освоение эвиденциальных средств русского языка на уровне А2 демонстрирует успешный перенос базовых принципов отслеживания источника информации, однако реализуется в упрощенной стратегии, для которой характерна ориентация на прямое наблюдение, сокращение разнообразия модусных оттенков и избегание пересказывательности. Полученные результаты могут иметь прикладное значение для преподавания РКИ, в том числе служить диагностическим инструментом для определения уровня развития эвиденциальной компетенции у учащихся.

Выявленные различия позволяют предложить конкретные задания для учащихся уровня А2–В1, направленные на развитие интерпретативной стратегии: трансформация дескриптивных предложений в инференциальные; приписывание персонажам ментальных состояний (мыслей, желаний, эмоций); добавление модусных маркеров в рамках диалога и нарратива; диалогизация нарратива. На уровне А2 рекомендуется фокусироваться на трансформации предложений, использовании глаголов психического состояния (*хотеть, бояться, радоваться*); введении эпистемических вводных слов (*кажется, наверное*). На уровне В1 следует расширять корпус ментальных предикатов, добавлять каузативные конструкции, вводить

оценочные и миративные маркеры (*к сожалению, неожиданно, к удивлению*) и практиковать диалогизацию нарратива – пересказ сюжета от лица разных персонажей с включением прямой или косвенной речи.

Диагностика уровня сформированности эвиденциальной компетенции может осуществляться через анализ письменных нарративов, созданных по тем же стимулам MAIN. Показателями могут служить: (а) соотношение прямой эвиденциальности и инференциальности, (б) разнообразие и частотность эвиденциальных маркеров и контекстных индикаторов, (в) наличие модусных (эпистемических и оценочных) маркеров. Предложенные здесь контрольные параметры могут быть использованы преподавателями для отслеживания прогресса учащихся.

В качестве перспективных направлений дальнейшего исследования предполагается: (1) провести аналогичный эксперимент на материале устных нарративов; (2) сравнить эвиденциальные стратегии информантов с другими родными языками, изучающих РКИ (англофонов, испанофонов); (3) проанализировать динамику освоения эвиденциальности на уровнях В1–В2 с применением лонгитюдного эксперимента. Также перспективным теоретическим направлением представляется уточнение классификации эвиденциальных маркеров и статуса эвиденциальных контекстных индикаторов в русском языке.

#### Список литературы

Астахова Т. Н. Проявление категорий «Эвиденциальность» и «Авторитетность» в текстах немецких СМИ // Вестник Северного (Арктического) федерального университета. Серия: Гуманитарные и социальные науки. 2014. № 1. С. 71–75.

Болдырев Н. Н. Языковые категории как формат знания // Вопросы когнитивной лингвистики. 2006. № 2 (7). С. 5–22.

Бондарко А. В. Категоризация в системе грамматики. М.: Языки славянских культур, 2011. 472 с.

- Борисова О. С. Развитие коммуникативных навыков на занятиях РКИ при изучении эвиденциальных конструкций: презентация доклада // IX Международный конгресс «Довузовский этап обучения в России и мире: язык, адаптация, социум, специальность». М., 2025. 11 с. URL: <https://disk.yandex.ru/d/umbIKnBoo3GIAw> (дата обращения: 23.01.2026).
- Григоренко М. Ю. Модусный статус эвиденциальности в современном русском языке // Вестник Московского государственного областного университета. Серия: Русская филология. 2009. № 4. С. 73–77.
- Иванова С. В., Хакимова Г. Ш. Эвиденциальность и эпистемическая модальность: сопряжение на уровне семантики слова // Вестник Санкт-Петербургского университета. Язык и литература. 2024. Т. 21. № 4. С. 802–824. <https://doi.org/10.21638/spbu09.2024.403>
- Кобрин О. А. Категория эвиденциальности: ее статус и формы выражения в разных языках // Вопросы когнитивной лингвистики. 2005. № 1. С. 86–98.
- Козинцева Н. А. Типология категории засвидетельствованности // Эвиденциальность в языках Европы и Азии / отв. ред. В. С. Храковский. СПб.: Наука, 2007. С. 13–36.
- Козловский Д. В. Типы взаимодействия модусной категории «эвиденциальность» с другими модусными категориями // Вестник Удмуртского университета. Серия История и филология. 2022. Т. 32. № 5. С. 1027–1036. <https://doi.org/10.35634/2412-9534-2022-32-5-1027-1036>
- Никитина Е. Н. Субъектные нули и перцептивный модус (к вопросу о выражении категории эвиденциальности в русском языке) // Вопросы языкознания. 2013. № 2. С. 69–82.
- Падучева Е. В. Есть ли в русском языке грамматически выраженная эвиденциальность? // Русский язык в научном освещении. 2013. № 2 (26). С. 9–29.
- Плунгян В. А. Введение в грамматическую семантику: грамматические значения и грамматические системы языков мира. М.: РГГУ, 2011. 672 с.
- Постникова А. Е. Грамматикализация и тип эвиденциальной системы // Типология морфосинтаксических параметров. 2020. Т. 3. № 2. С. 147–165.
- Сафина А. Р. Теоретические аспекты изучения эпистемической модальности и эвиденциальности // Филологические науки. Вопросы теории и практики. 2016. № 9 (2). С. 142–145.
- Страхова В. С. Маркеры эвиденциальности в языке СМИ // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2016. № 6 (745). С. 165–174.
- Храковский В. С. Эвиденциальность, эпистемическая модальность, (ад)миративность // Эвиденциальность в языках Европы и Азии / отв. ред. В. С. Храковский. СПб.: Наука, 2007. С. 600–631.
- Шестухина И. Ю. Категория эвиденциальности в русском нарративном тексте: коммуникативно-прагматический аспект: Автореф. дис. ... канд. филол. наук. Барнаул, 2009. 17 с.
- Aikhenvald A. Y. Evidentiality. Oxford: Oxford University Press, 2004. 480 p.
- Aksu-Koç A. Some aspects of the acquisition of evidentials in Turkish // Evidentials. Berlin: De Gruyter, 2000. Pp. 15–28. <https://doi.org/10.1515/9783110805284.15>
- Arslan S., Bastiaanse R., Felser C. Looking at the evidence in visual world: eye-movements reveal how bilingual and monolingual Turkish speakers process grammatical evidentiality // Frontiers in Psychology. 2015. Vol. 6. Pp. 1–13. <https://doi.org/10.3389/fpsyg.2015.01387>
- Brugman C. M., Macaulay M. Characterizing evidentiality // Linguistic Typology. 2015. Vol. 19, No. 2. Pp. 201–237. <https://doi.org/10.1515/lingty-2015-0007>
- Cornillie B. Evidentiality and epistemic modality: On the close relationship between two different categories // Functions of Language. 2009. Vol. 16. No. 1. Pp. 44–62. <https://doi.org/10.1075/fo1.16.1.04cor>
- De Villiers J. G. Evidentials in Tibetan: Acquisition, semantics, and cognitive development / De Villiers J. G., Garfield J., Gernet-Girard H., Roeper T., Speas M. // New Directions for Child and Adolescent Development. 2009. No. 125. Pp. 29–47. <https://doi.org/10.1002/cd.248>
- De Villiers J. G., Garfield J. Evidentiality and narrative // Journal of Consciousness Studies. 2009. Vol. 16. No. 6–8. Pp. 191–217.

Gagarina N. MAIN: Multilingual Assessment Instrument for Narratives – Revised / Gagarina N., Klop D., Kunnari S., Tantele K., Välimaa T., Bohnacker U., Walters J. // ZAS Papers in Linguistics. 2019. Vol. 63. Pp. 1–20. <https://doi.org/10.21248/zaspil.63.2019.516>

Ishida K. How can you be so certain? The use of hearsay evidentials by English-speaking learners of Japanese // Journal of Pragmatics. 2006. Vol. 38. No. 8. Pp. 1281–1304. <https://doi.org/10.1016/j.pragma.2005.10.006>

Kaya-Soykan D., Antonova-Unlu E., Sagin-Simsek C. The production and perception of Turkish evidentiality markers by Turkish-German returnees // Applied Linguistics Review. 2023. Vol. 14. No. 2. Pp. 251–270. <https://doi.org/10.1515/applirev-2020-0042>

Kittilä S. General knowledge as an evidential category // Linguistics. 2019. Vol. 57. No. 6. Pp. 1271–1304. <https://doi.org/10.1515/ling-2019-0027>

Kosta P., Iskakova G. Zh. New tendencies in the study of modality: epistemic modality and evidentiality // Bulletin of the Moscow State Regional University (Linguistics). 2022. No. 6. Pp. 25–40. <https://doi.org/10.18384/2310-712x-2022-6-25-40>

Krasnoshchekova S., Kashleva K. Narrative competence of adult L2 Russian learners // Journal of Psycholinguistic Research. 2019. Vol. 48, no. 3. Pp. 617–641. <https://doi.org/10.1007/s10936-018-9622-3>

Leclercq P., Mélac E. Second language acquisition of evidentiality in French and English in a narrative task // Language, Interaction and Acquisition. 2021. Vol. 12. No. 2. Pp. 251–283. <https://doi.org/10.1075/lia.20025.lec>

Mushin I. Evidentiality and deixis in narrative retelling // Journal of Pragmatics. 2000. Vol. 32. No. 7. Pp. 927–957. [https://doi.org/10.1016/S0378-2166\(99\)00085-5](https://doi.org/10.1016/S0378-2166(99)00085-5)

Mushin I. Making knowledge visible in discourse: Implications for the study of linguistic evidentiality // Discourse Studies. 2013. Vol. 15. No. 5. Pp. 627–645. <https://doi.org/10.1177/1461445613501447>

Narita R. The effects of pragmatic consciousness-raising activity on the development of pragmatic awareness and use of hearsay evidential markers for learners of Japanese as a foreign language // Journal of Pragmatics. 2012. Vol. 44. No. 1. Pp. 1–29. <https://doi.org/10.1016/j.pragma.2011.09.016>

Nuckolls J. B., Michael L. Evidentials and evidential strategies in interactional and socio-cultural context // Benjamins Current Topics / ed. J. B. Nuckolls, L. Michael. Amsterdam: John Benjamins Publishing Company, 2014. Pp. 13–20. <https://doi.org/10.1075/bct.63.03int>

Ozturk O., Papafragou A. The acquisition of evidentiality and source monitoring // Language Learning and Development. 2016. Vol. 12. No. 2. Pp. 199–230. <https://doi.org/10.1080/15475441.2015.1024834>

Peterson T. Mirativity as surprise: Evidentiality, information, and deixis // Journal of Psycholinguistic Research. 2016. Vol. 45. No. 6. Pp. 1327–1357. <https://doi.org/10.1007/s10936-015-9408-9>

Rett J., Hyams N. The acquisition of syntactically encoded evidentiality // Language Acquisition. 2014. Vol. 21. No. 2. Pp. 173–198. <https://doi.org/10.1080/10489223.2014.884572>

Tosun S., Filipović L. Better late than early: The effect of formal second language training on processing of evidentiality in Turkish-English bilinguals // International Journal of Bilingualism. 2025. Vol. 29. No. 4. Pp. 1008–1026. <https://doi.org/10.1177/13670069241266340>

Tournadre N., LaPolla R. J. Towards a new approach to evidentiality: Issues and directions for research // Linguistics of the Tibeto-Burman Area. 2014. Vol. 37. No. 2. Pp. 240–263. <https://doi.org/10.1075/ltba.37.2.04tou>

Ünal E., Papafragou A. Relations between language and cognition: Evidentiality and sources of knowledge // Topics in Cognitive Science. 2020. Vol. 12, No. 1. Pp. 115–135. <https://doi.org/10.1111/tops.12355>

Verhees S. Defining evidentiality // Voprosy Jazykoznanija. 2019. No. 6. Pp. 113–133. <https://doi.org/10.31857/S0373658X0007549-2>

Wiemer B. Hearsay in European languages: Toward an integrative account of grammatical and lexical marking // Linguistic Realization of Evidentiality in European Languages / ed. G. Diewald, E. Smirnova. Berlin: De Gruyter Mouton, 2010. Pp. 59–130. <https://doi.org/10.1515/9783110223972.59>

## References

Astakhova, T. N. (2014). Manifestations of evidentiality and authority in German mass media texts. *Vestnik of Northern (Arctic) Federal University. Series "Humanitarian and Social Sciences"*, 1, 71–75. (In Russian)

- Boldyrev, N. N. (2006). Yazykovye kategorii kak format znaniya [Language categories as a format of knowledge]. *Issues of Cognitive Linguistics*, (2) 7, 5–22. (In Russian)
- Bondarko, A. V. (2011). Kategorizatsiya v sisteme grammatiki [Categorization in the system of grammar]. LRC Publishing House, Moscow, Russia. (In Russian)
- Borisova, O. S. (2025). Razvitie kommunikativnykh navykov na zanyatiyakh RKI pri izuchenii ehvidentsialnykh konstruksii [The development of communicative skills in Russian as a foreign language classes through the study of evidential constructions], in *IX Mezhdunarodnyi kongress «Dovuzovskii etap obucheniya v Rossii i mire: yazyk, adaptatsiya, sotsium, spetsialnost»* [The IX International Congress “Pre-University Education in Russia and the World: Language, Adaptation, Society, Specialty”]. Moscow, 11 p. Available at: <https://disk.yandex.ru/d/umbIKnBoo3GIAw> (Accessed 23 January 2026). (In Russian)
- Grigorenko, M. Yu. (2009). Modusnyy status ehvidentsialnosti v sovremennom russkom yazyke [Modus status of evidentiality in modern Russian], *Russian Studies in Philology*, (4), 73–77. (In Russian)
- Ivanova, S. V. and Khakimova, G. Sh. (2024). Ehvidentsialnost' i ehpiistemicheskaya modalnost': sopryazhenie na urovne semantiki slova [Evidentiality and epistemic modality: Overlap view within word semantics], *Vestnik of Saint Petersburg University. Language and Literature*, 21 (4), 801–823. (In Russian) <https://doi.org/10.21638/spbu09.2024.403>
- Kobrina, O. A. (2005). Kategoriya ehvidentsialnosti: ee status i formy vyrazheniya v raznykh yazykakh [The category of evidentiality: its status and forms of expression in different languages], *Issues of Cognitive Linguistics*, (1), 86–98. (In Russian)
- Kozintseva, N. A. (2007). Tipologiya kategorii zasvidetelstvovannosti [Typology of the category of evidentiality], in V. S. Khrakovskii (ed.), *Ehvidentsialnost v yazykakh Evropy i Azii* [Evidentiality in the languages of Europe and Asia], Nauka, St. Petersburg, 13–36. (In Russian)
- Kozlovsky, D. V. (2022). Tipy vzaimodeistviya modusnoi kategorii «ehvidentsialnost» s drugimi modusnymi kategoriymi [Types of interaction of the modus category “evidentiality” with other modus categories], *Bulletin of Udmurt University. Series History and Philology*, 32 (5), 1027–1036. (In Russian)
- (Russian) <https://doi.org/10.35634/2412-9534-2022-32-5-1027-1036>
- Nikitina, E. N. (2013). Null subjects and perceptive mode (on the expression of evidentiality in Russian), *Voprosy yazykoznaniya*, (2), 69–82. (In Russian)
- Paducheva, E. V. (2013). Is there grammatically expressed evidentiality in Russian?, *Russian Language and Linguistic Theory*, (2) 26, 9–29. (In Russian)
- Plungian, V. A. (2011). *Vvedenie v grammaticheskuyu semantiku: grammaticheskie znacheniya i grammaticheskie sistemy yazykov mira* [Introduction to grammar semantics: Grammar meaning and grammar systems in the languages of the world]. RGGU, Moscow, Russia. (In Russian)
- Postnikova, A. E. (2020). *Grammaticalization and the typology of evidential systems. Tipologiya morfosintaksicheskikh parametrov* [Typology of Morphosyntactic Parameters], 3 (2), 147–165. (In Russian)
- Safina, A. R. (2016). Theoretical aspects of studying epistemic modality and evidentiality. *Philology. Theory & Practice*, 9 (2), 142–145. (In Russian)
- Strakhova, V. S. (2016). Evidential markers in the language of mass media. *Vestnik of Moscow State Linguistic University. Humanities*, (6) 745, 165–174. (In Russian)
- Khrakovskii, V. S. (2007). Evidentiality, epistemic modality, (ad)mirative, in: V. S. Khrakovskii (ed.), *Ehvidentsialnost v yazykakh Evropy i Azii* [Evidentiality in the languages of Europe and Asia]. Nauka, St. Petersburg, Russia, 600–631. (In Russian)
- Shestukhina, I. Yu. (2009). Kategoriya ehvidentsialnosti v russkom narrativnom tekste: kommunikativno-pragmaticheskiy aspekt [The category of evidentiality in Russian narrative text: a communicative-pragmatic aspect], Abstract of Ph.D. dissertation, Altai State University, Barnaul, Russia. (In Russian)
- Aikhenvald, A. Y. (2004). *Evidentiality*, Oxford University Press, Oxford, UK. (In English)
- Aksu-Koç, A. (2000). Some aspects of the acquisition of evidentials in Turkish, in: L. Johanson and B. Utas (eds.), *Evidentials*. De Gruyter, Berlin, Germany, 15–28. <https://doi.org/10.1515/9783110805284.15> (In English)
- Arslan, S., Bastiaanse, R. and Felser, C. (2015). Looking at the evidence in visual world:

eye-movements reveal how bilingual and monolingual Turkish speakers process grammatical evidentiality, *Frontiers in Psychology*, 6, 1–13. <https://doi.org/10.3389/fpsyg.2015.01387> (In English)

Brugman, C. M. and Macaulay, M. (2015). Characterizing evidentiality, *Linguistic Typology*, 19 (2), 201–237. <https://doi.org/10.1515/lingty-2015-0007> (In English)

Cornillie, B. (2009). Evidentiality and epistemic modality: On the close relationship between two different categories, *Functions of Language*, 16 (1), 44–62. <https://doi.org/10.1075/fof.16.1.04cor> (In English)

De Villiers, J. G. and Garfield, J. (2009). Evidentiality and narrative, *Journal of Consciousness Studies*, 16 (6–8), 191–217. (In English)

De Villiers, J. G., Garfield, J., Gernet-Girard, H., Roeper, T. and Speas, M. (2009). Evidentials in Tibetan: Acquisition, semantics, and cognitive development, *New Directions for Child and Adolescent Development*, 125, 29–47. <https://doi.org/10.1002/cd.248> (In English)

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U. and Walters, J. (2019). MAIN: Multilingual Assessment Instrument for Narratives – Revised, *ZAS Papers in Linguistics*, 63, 1–20. <https://doi.org/10.21248/zaspil.63.2019.516> (In English)

Ishida, K. (2006). How can you be so certain? The use of hearsay evidentials by English-speaking learners of Japanese, *Journal of Pragmatics*, 38 (8), 1281–1304. <https://doi.org/10.1016/j.pragma.2005.10.006> (In English)

Kaya-Soykan, D., Antonova-Unlu, E. and Sagin-Simsek, C. (2023). The production and perception of Turkish evidentiality markers by Turkish-German returnees, *Applied Linguistics Review*, 14 (2), 251–270. <https://doi.org/10.1515/applirev-2020-0042> (In English)

Kittilä, S. (2019). General knowledge as an evidential category, *Linguistics*, 57 (6), 1271–1304. <https://doi.org/10.1515/ling-2019-0027> (In English)

Kosta, P. and Iskakova, G. Zh. (2022). New tendencies in the study of modality: epistemic modality and evidentiality, *Bulletin of the Moscow State Regional University (Linguistics)*, 6, 25–40.

<https://doi.org/10.18384/2310-712x-2022-6-25-40> (In English)

Krasnoshchekova, S. and Kashleva, K. (2019). Narrative competence of adult L2 Russian learners, *Journal of Psycholinguistic Research*, 48 (3), 617–641. <https://doi.org/10.1007/s10936-018-9622-3> (In English)

Leclercq, P. and Mélaç, E. (2021). Second language acquisition of evidentiality in French and English in a narrative task. *Language, Interaction and Acquisition*, 12 (2), 251–283. <https://doi.org/10.1075/lia.20025.lec> (In English)

Mushin, I. (2000). Evidentiality and deixis in narrative retelling, *Journal of Pragmatics*, 32 (7), 927–957. [https://doi.org/10.1016/S0378-2166\(99\)00085-5](https://doi.org/10.1016/S0378-2166(99)00085-5) (In English)

Mushin, I. (2013). Making knowledge visible in discourse: Implications for the study of linguistic evidentiality, *Discourse Studies*, 15 (5), 627–645. <https://doi.org/10.1177/1461445613501447> (In English)

Narita, R. (2012). The effects of pragmatic consciousness-raising activity on the development of pragmatic awareness and use of hearsay evidential markers for learners of Japanese as a foreign language, *Journal of Pragmatics*, 44 (1), 1–29. <https://doi.org/10.1016/j.pragma.2011.09.016> (In English)

Nuckolls, J. B. and Michael, L. (2014). Evidentials and evidential strategies in interactional and socio-cultural context, in: J. B. Nuckolls and L. Michael (eds.), *Benjamins Current Topics*, John Benjamins Publishing Company, Amsterdam, Netherlands, 13–20. <https://doi.org/10.1075/bct.63.03int> (In English)

Ozturk, O. and Papafragou, A. (2016). The acquisition of evidentiality and source monitoring, *Language Learning and Development*, 12 (2), 199–230. <https://doi.org/10.1080/15475441.2015.1024834> (In English)

Peterson, T. (2016). Mirativity as surprise: Evidentiality, information, and deixis, *Journal of Psycholinguistic Research*, 45 (6), 1327–1357. <https://doi.org/10.1007/s10936-015-9408-9> (In English)

Rett, J. and Hyams, N. (2014). The acquisition of syntactically encoded evidentiality, *Language Acquisition*, 21 (2), 173–198. <https://doi.org/10.1080/10489223.2014.884572> (In English)

Tosun, S. and Filipović, L. (2025). Better late than early: The effect of formal second language training on processing of evidentiality in Turkish-English bilinguals, *International Journal of Bilingualism*, 29(4), 1008–1026. <https://doi.org/10.1177/13670069241266340> (In English)

Tournadre, N. and LaPolla, R. J. (2014). Towards a new approach to evidentiality: Issues and directions for research, *Linguistics of the Tibeto-Burman Area*, 37 (2), 240–263. <https://doi.org/10.1075/ltba.37.2.04tou> (In English)

Ünal, E. and Papafragou, A. (2020). Relations between language and cognition: Evidentiality and sources of knowledge, *Topics in Cognitive Science*, 12 (1), 115–135. <https://doi.org/10.1111/tops.12355> (In English)

Verhees, S. (2019). Defining evidentiality, *Voprosy Jazykoznanija*, 6, 113–133. <https://doi.org/10.31857/S0373658X0007549-2> (In English)

Wiemer, B. (2010). Hearsay in European languages: Toward an integrative account of grammatical and lexical marking, in: G. Diewald and E. Smirnova (eds.), *Linguistic Realization of Evidentiality in European Languages*, De Gruyter Mouton, Berlin, Germany, 59–130. <https://doi.org/10.1515/9783110223972.59> (In English)

**Авторы прочитали и одобрили окончательный вариант рукописи.**

**The 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.**

**Софья Викторовна Краснощекова**, кандидат филологических наук, научный сотрудник отдела теории грамматики, Институт лингвистических исследований Российской академии наук, Санкт-Петербург, Россия.

**Sofia V. Krasnoshchekova**, PhD in Philology, Researcher at the Department of Theory of Grammar, Institute for Linguistic Studies of the Russian Academy of Sciences, St. Petersburg, Russia.

**Елена Владимировна Галкина**, научный сотрудник отдела физиологии сенсорных систем, лаборатории психофизиологии речи, Институт физиологии им. И. П. Павлова Российской академии наук, Санкт-Петербург, Россия.

**Elena V. Galkina**, Researcher at the Department of Physiology of Sensor Systems, Laboratory of Psychophysiology of Speech, Pavlov Institute of Physiology of the Russian Academy of Sciences, St. Petersburg, Russia.

**Ван Илин**, кандидат филологических наук, преподаватель русского языка Центра русского языка, Университет МГУ-ППИ в Шэньчжэне, Шэньчжэнь, Китай.

**Wang Yiling**, PhD in Philology, Teacher of Russian Language in the Centre of Russian Language, Shenzhen MSU-BIT University, Shenzhen, China.