

ISSN 2313-8912

НАУЧНЫЙ РЕЗУЛЬТАТ

ВОПРОСЫ ТЕОРЕТИЧЕСКОЙ И ПРИКЛАДНОЙ ЛИНГВИСТИКИ

RESEARCH RESULT. THEORETICAL AND APPLIED LINGUISTICS

12(3) 2026

16+

Сайт журнала:
rlinguistics.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, № 3. 2026
СЕТЕВОЙ НАУЧНО-ПРАКТИЧЕСКИЙ
ЖУРНАЛ
Издается с 2014 г.
ISSN 2313-8912



Volume 12 (3). 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
Карасик 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-Софии 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. Ополе, 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 "Volgograd 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. Panasenکو, 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

РЕДАКТОРСКАЯ СТАТЬЯ

EDITORIAL

Литвинова Т. А., Дехнич О. В.

Когда естественный текст
выглядит искусственным:
воспринимаемое ИИ-
происхождение как свойство текста 3

Litvinova T. A., Dekhnich O. V.

When human writing looks
artificial: perceived
AI-likeness as a property
of text 3

РАЗДЕЛ I. ТЕОРИЯ ЯЗЫКА

SECTION I. THEORY OF LANGUAGE

Раренко М. Б.

Языковые индикаторы
персуазивности в социальной
рекламе 32

Rarenko M. B.

Linguistic indicators
of invasiveness in social
advertising 32

РАЗДЕЛ II. ПРИКЛАДНАЯ ЛИНГВИСТИКА

SECTION II. APPLIED LINGUISTICS

**Бочкарев В. В., Афанасьев А. С.,
Бакиров Р. А.**

Ранговое расстояние Спирмена как
инструмент стилометрической
атрибуции: сравнительный анализ
на разнородных корпусах 52

**Bochkarev V. V., Afanasev A. S.,
Bakirov R. A.**

Spearman's rank distance as
a tool for stylometric attribution:
a comparative analysis on
heterogeneous corpora 52

Can L., Akbaş E.

Discourse-Organizing Metadiscourse in
ELF Academic Speech: A Corpus Study
Across Speech Events 80

Can L., Akbaş E.

Discourse-Organizing Metadiscourse
in ELF Academic Speech: A Corpus
Study Across Speech Events 80

Жемилева М. М.,

Керо Хервилья Э. Ф.

Когнитивно-семантические
механизмы концептуализации
коллективности в английских
номинациях сообществ животных:
корпусный анализ 113

Zhemileva M. M.,

Quero Gervilla E. F.

Cognitive-semantic mechanisms of
collective conceptualization in
English animal community
nominations: a corpus-based
analysis 113

Исламов Р. С., Савельева И. В.

Кооперативность ответов
голосового ассистента и психолога:
соблюдение принципов Герберта
Пола Грайса и Джеффри Лича 145

Islamov R. S., Savelyeva I. V.

Cooperative responses of a voice
assistant and a psychologist:
adherence to the principles of Herbert
Paul Grice and Geoffrey Leech 145

Колмогорова А. В.

AI как писатель глазами читателя:
результаты психолингвистического
эксперимента по оценке
сгенерированных художественных
текстов любительской прозы 170

Kolmogorova A. V.

AI as a literary author
from the reader's perspective:
findings from a psycholinguistic
experiment assessing synthetic
amateur prose 170

EDITORIAL
РЕДАКТОРСКАЯ СТАТЬЯ

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

DOI: 10.18413/2313-8912-2026-12-3-0-1

Tatiana A. Litvinova¹ 
Olga V. Dekhnich² When human writing looks artificial:
perceived AI-likeness as a property of text

¹ Belgorod State National Research University,
85 Pobedy St., Belgorod, 308015, Russia
E-mail: centr_rus_yaz@mail.ru
ORCID: 0000-0002-6019-3700

² Belgorod State National Research University,
85 Pobedy St., Belgorod, 308015, Russia
E-mail: dekhnich@bsuedu.ru
ORCID: 0000-0001-6088-2656

Received 24 August 2026; accepted 15 September 2026; published 30 September 2026

Abstract: A piece of writing can look as if it were produced by AI even when it wasn't, and conversely, text generated by AI can still be mistaken for human work. We investigated this mismatch between the true source of 28 Russian texts (14 written by humans and 14 generated by AI) and their perceived AI-likeness. The texts were assessed by professional linguists under three different evaluation setups. For each text, we extracted 46 embedding-based semantic alignment features and 12 stylometric measures. In the third evaluation regime (W3), we also used a balanced subset of 10 texts and collected ratings along 13 perceptual dimensions before participants made their source judgments.

Across both semantic and stylometric space, human and AI texts showed a clear separation. Unsupervised components related to origin produced large Human-AI differences (Hedges' $|g| \approx 1.4$), and nested leave-one-text-out ridge models achieved AUCs between .832 and .878. However, the objective dimensions that distinguished the actual origin were not sufficient on their own to explain AI attribution once the origin was controlled for. The importance of cues also varied across evaluation regimes. In particular, seven features showed FDR-significant regime $\times$ feature interactions between W2 and W3, all with reversed slopes; mean word length also shifted direction between W1 and W2. In W3, the perceptual ratings were largely driven by a dimension that contrasted authorial-experiential presence with schematic generality. Shifting toward the authorial-experiential end substantially lowered AI-oriented attribution on the original five-category scale (OR = .228, 95% CI [.145, .359], $p < .001$).

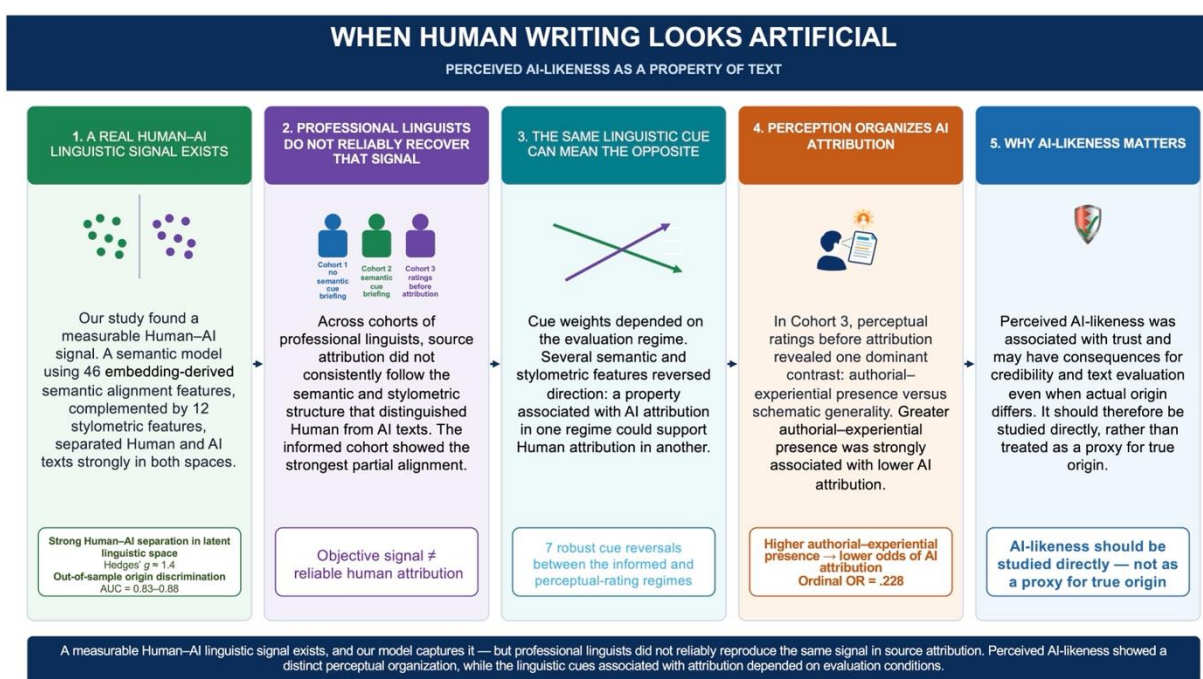
So, perceived AI-likeness cannot be viewed as a straightforward reflection of either actual authorship or the linguistic markers tied to origin. In these results, it functioned more like a perceptual property of the text: it depended on whether readers felt there was a situated author behind what they read. Meanwhile, the linguistic cues that supported AI attribution varied depending on the evaluation regime. This distinction is

important because perceived AI authorship can affect trust, credibility, and later assessments even when the source judgment turns out to be incorrect.

Keywords: AI-Generated Text; Perceived AI-Likeness; AI Attribution; Human Text Perception; Authorial Presence; Semantic Features; Stylometry; Source Attribution

Acknowledgements: This editorial is dedicated to the 150th anniversary of Belgorod State National Research University and its enduring tradition of scholarly inquiry into language, human communication, and emerging forms of human – AI interaction. Tatiana A. Litvinova acknowledges the support of Russian Science Foundation, grant N 25-28-03166, <https://rscf.ru/project/25-28-03166/>. Olga V. Dekhnich received no financial support for the research, authorship, and publication of this article.

How to cite: Litvinova, T. A., Dekhnich, O. V. (2026). When human writing looks artificial: perceived AI-likeness as a property of text, *Research Result. Theoretical and Applied Linguistics*, 12 (3), 3–31. DOI: 10.18413/2313-8912-2026-12-3-0-1



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

DOI: 10.18413/2313-8912-2026-12-3-0-1

Литвинова Т. А.¹ 
Дехнич О. В.² 

Когда естественный текст выглядит искусственным:
воспринимаемое ИИ-происхождение как свойство
текста

¹ Воронежский государственный педагогический университет
ул. Ленина, 86, Воронеж, 394043, Россия

E-mail: centr_rus_yaz@mail.ru
ORCID: 0000-0002-6019-3700

² Белгородский государственный национальный исследовательский университет
ул. Победы, 85, Белгород, 308015, Россия

E-mail: dekhnich@bsuedu.ru
ORCID: 0000-0001-6088-2656

Статья поступила 24 августа 2026 г.; принята 15 сентября 2026 г.;
опубликована 30 сентября 2026 г.

Аннотация: Текст может восприниматься как созданный искусственным интеллектом, даже если на самом деле его написал человек; и наоборот, текст, сгенерированный ИИ, может казаться «человеческим». Мы исследовали разрыв между реальным происхождением текста и его воспринимаемым ИИ-происхождением на материале 28 текстов на русском языке (14 написанных людьми и 14 ИИ-сгенерированных), которые оценивали профессиональные лингвисты в рамках трёх разных режимов оценки. Каждый текст был представлен в признаковом пространстве 46 семантических показателей, полученных на основе эмбедингов, и 12 стилометрических индексов. В третьем эксперименте тексты дополнительно оценивались по 13 перцептивным шкалам до того, как респонденты выносили суждение о происхождении текста.

Тексты, написанные людьми и сгенерированные ИИ, хорошо разделялись как в семантическом, так и в стилометрическом пространстве построенной нами модели. Многомерный анализ показал значительные различия между человеческими и ИИ-текстами ($|g|$ Хеджеса $\approx 1,4$), а вложенные ridge-модели с проверкой по схеме leave-one-text-out достигали AUC от 0,832 до 0,878. Однако объективные лингвистические показатели, различавшие тексты по реальному происхождению, не объясняли независимо вероятность отнесения лингвистом текста к ИИ. Кроме того, вес отдельных признаков менялся в зависимости от эксперимента. Для семи признаков были обнаружены значимые после FDR-коррекции взаимодействия «режим $\times$ признак» между экспериментами 2 и 3, причём во всех случаях направление связи менялось на противоположное. Средняя длина слова дополнительно меняла направление связи между первым и вторым экспериментом. В третьем эксперименте структура перцептивных оценок в основном определялась одной осью, противопоставляющей авторское присутствие и схематическую обобщённость. Смещение к полюсу авторского присутствия резко снижало вероятность отнесения текста к ИИ по пятибалльной шкале ($OR = 0,228$; 95% ДИ $[0,145; 0,359]$; $p < 0,001$).

Таким образом, воспринимаемое ИИ-происхождение текстов нельзя рассматривать ни как отражение реального авторства, ни как отражение объективного лингвистического сигнала происхождения текста. В наших экспериментах оно проявлялось как перцептивное свойство и было связано с тем, возникает ли у читателя ощущение конкретного, стоящего за текстом автора, тогда как лингвистические признаки, объективно связанные с происхождением текста, зависели от условий эксперимента. Данное различие важно, поскольку предполагаемое ИИ-авторство может влиять на степень доверия читателя к тексту, его воспринимаемую достоверность и последующую оценку текста даже в тех случаях, когда суждение об источнике происхождения текста ошибочно.

Ключевые слова: ИИ-сгенерированный текст; Воспринимаемое ИИ-происхождение текста; Атрибуция текста; Восприятие текста; Авторское присутствие; Семантические признаки; Стилометрия; Атрибуция источника

Информация об источниках финансирования или грантах: Данная редакционная статья посвящена 150-летию Белгородского государственного национального исследовательского университета и его многолетней традиции научного исследования языка, человеческой коммуникации и новых форм

взаимодействия человека и искусственного интеллекта. Т.А. Литвинова благодарит за финансовую поддержку Российский научный фонд, грант № 25-28-03166, <https://rscf.ru/project/25-28-03166/>. О.В. Дехнич не получала финансовой поддержки за выполнение исследований, написание и публикацию статьи.

Информация для цитирования: Литвинова Т. А., Дехнич О. В. Когда естественный текст выглядит искусственным: воспринимаемое ИИ-происхождение как свойство текста // Научный результат. Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 3. С. 3–31. DOI: 10.18413/2313-8912-2026-12-3-0-1



1. INTRODUCTION

Whether a text was written by AI and whether readers believe it was are two separate questions. Writing done by humans can still seem artificial, while AI-generated content can often blend in without raising suspicion. This distinction is important because it can affect how a text is judged and received. Research experiments show that adding an AI-authorship label can reduce perceptions of message and source credibility, as well as judgments about quality and author competence. It can also make people less willing to share or use the text, even when the wording is kept the same – or even if the text is actually human-written (Altay and Gilardi, 2024; Lermann Henestrosa and Kimmerle, 2024; Proksch et al., 2024). However, when information on text origin was missing,

humans favored machine-generated text in over half of the cases, particularly when they could not recognize which text was written by a human (Wang et al., 2026). In the context of news communication, the perceived level of AI involvement – rather than an explicit label – has also been linked to lower message and source credibility (Jia et al., 2024). The evidence in education points the same way: when texts were presented as ChatGPT-generated, they received lower credibility ratings (Latini et al., 2026). Thus, across the studies cited above, a recurring anthropocentric bias emerged: texts believed to be human-authored were evaluated more favorably than those perceived as AI-generated.

A lot of the existing research frames AI-generated writing mainly as a detection issue:

if someone is given a text, can they – or a computational system – figure out its real origin? Human success varies a great deal. For example, in a study with 63 university lecturers, AI-written academic passages were correctly recognized only 57% of the time, and texts produced at a professional level were particularly hard to spot (Fiedler and Döpke, 2025). Hungarian readers achieved 66% accuracy (irrespective of self-reported AI-skills) on narrative texts, and human-written passages were incorrectly attributed more often than AI ones (Szabó et al., 2026), while Czech readers identified whether AI or human wrote poetry at about chance level (Marklová et al., 2026). A recent study on register perception also found that it can be difficult to detect texts generated by ChatGPT (Demir, 2026). On the opposite end, a large multilingual project that looked across 16 datasets, nine languages, and nine domains reported an average human (19 annotators – LLM researchers or practitioners) detection accuracy of 87.6%, using test conditions meant to push toward the upper limit of human performance (Wang et al., 2026). In other words, people's ability to detect AI writing isn't a single fixed skill that stays the same regardless of language, genre, context, or method.

Calling every mismatch with the ground truth a detection error can obscure a potentially useful effect. A human-authored text that keeps being labeled as AI-generated may have characteristics that make a machine-authorship explanation plausible. On the other hand, an AI-generated text that consistently appears human may miss the cues readers typically associate with machine writing. These judgments are not purely random: readers rely on signals such as writing style, grammatical form, social and personal content, concreteness, cultural specificity, warmth, and impersonality. However, these heuristics can mislead and can also shift depending on context (Hau et al., 2026; Jakesch et al., 2023; Liu et al., 2026; Wang et al., 2026). As a result,

attribution can be systematic without being correct.

For this, we look at perceived AI-likeness at the text level. In other words, we mean how likely a text – under a specific evaluation setup – is to be judged as AI-authored. Practically speaking, it's the share of AI attributions that a text gets within that setting. This measure isn't the same as actual authorship, and it's not the same as accuracy. By framing the goal this way, both false positives and false negatives become genuinely useful findings rather than just leftover errors. A text that's human-written but repeatedly comes across as AI-like – or an AI-written text that repeatedly seems human – becomes especially interesting, because actual authorship and what people perceive don't match.

At the same time, human and AI writing can differ in clear, objective ways at the linguistic level. Corpus and stylometric research have reported differences in lexical choices, syntax, discourse organization, affect, and overall structure of AI-generated and human texts. For instance, AI-generated academic writing has been linked to distinctive word choices, less variation in syntax, and only partial overlap with the formulaic patterns humans tend to use (Tudino and Qin, 2024). Studies of EFL writing have also found multivariate linguistic fingerprints that separate human essays from ChatGPT essays (Mizumoto et al., 2024). Similar patterns show up in analyses of hotel reviews, which report systematic differences in content, style, and structure (Markowitz et al., 2024). In creative writing, human texts tend to show more stylistic variety than outputs from large language models (O'Sullivan, 2025). A recent study of 4,346 paired essays across 110 subject areas found the same kind of pattern at scale: AI-generated writing tends to be stylistically more uniform than the variability seen in human writing (Chen et al., 2026). In picture-description tasks, AI-generated texts were more homogeneous than human-authored texts and differed markedly in lexical

diversity, morphological density and entropy, syntactic roles and complexity, and indices of modality engagement (Litvinova et al., 2025). So, if readers misidentify a text, that doesn't necessarily mean there are no measurable linguistic differences.

The main unresolved issue sits between these two strands of work. Linguistic-fingerprint studies ask what, objectively, distinguishes human writing from AI writing. Attribution studies ask what makes a reader decide that a text is human or AI. But those questions are rarely tested using the same texts, and even less is known about whether the explanatory power of a linguistic feature stays the same when the evaluation context changes. Because of this, we separate three concepts that are often mixed together: 1) the objective structure that reflects the true source; 2) the linguistic and perceptual factors that go along with AI attribution; 3) attribution accuracy itself. We also don't assume that the same cue means the same thing across different evaluation regimes.

That leads to a direct question: are the properties that distinguish AI-generated writing from human-written text also the properties that make a text look AI-generated? If the answer is yes, then perceived AI-likeness may be picking up on real distributional differences between human and machine language. If the answer is no, then attribution has to involve more than just a reader's perceptual response to a source signal. There are good reasons to expect the latter. Jakesch et al. (2023) showed that some authorship heuristics don't line up well with actual source distributions, while Hau et al. (2026) found that how useful attribution strategies are depends heavily on genre and contextual knowledge. What's still largely untested is whether objective origin structure and perceived AI-likeness can be separated within the same set of stimuli, whether cue weights shift across evaluation regimes, and what readers rely on when objective origin cues aren't enough.

We tackle these questions using 28 Russian-language texts with known human or

AI origins and three evaluation regimes involving professional linguists. The design places actual source and perceived source in the same analytical frame. Misclassification isn't treated only as an error; instead, we analyze text-level AI-attribution propensity as a way of measuring perceived AI-likeness. The same texts are also represented using 46 embedding-derived semantic alignment measures and a predefined 12-feature stylometric space. This lets us test whether the dimensions that separate true origins also explain attribution. We then use matched stimuli to examine whether text-level AI-likeness and cue weighting stay consistent across regimes. Finally, in one of the three evaluation regimes, 13 perceptual ratings collected before source attribution allow us to study how readers experienced the text - and the implied author - before deciding who, or what, wrote it. The Russian-language setting also adds to a literature where substantial variation across languages and domains has already been documented (Marklová et al., 2026; Szabó et al., 2026; Wang et al., 2026).

These choices boil down to four questions:

1. Do human- and AI-authored texts occupy different objective semantic and stylometric spaces?
2. Do those origin-related dimensions also explain AI attribution?
3. Does text-level AI-likeness, along with linguistic cue weighting, remain stable across evaluation regimes?
4. When detailed reader judgments are available, what perceptual dimensions organize AI attribution?

The larger point is that perceived AI-likeness may deserve to be studied on its own. As AI-assisted writing becomes more common, a simple Human/AI label captures only part of what matters communicatively. A human-authored text can lose credibility when it's presented as AI-generated (Altay and Gilardi, 2024), and perceived AI involvement has been linked to lower message and source credibility, partly through reduced perceptions of humanness (Jia et al., 2024).

Understanding why some texts attract AI attribution is therefore important not just for AI detection, but also for reader trust, editorial decisions, textual authenticity, and how AI-mediated communication is ultimately received.

2. MATERIALS AND METHODS

2.1. Study design and evaluation regimes

We tested how AI-like Russian texts seemed to professional linguists under three different evaluation setups. The study design separated the true origin of each text from what participants perceived. The main result we observed was AI attribution, while "perceived AI-likeness" was defined at the text level as how likely a text was to be judged as AI-generated within a given evaluation regime.

Our materials included 28 Russian-language texts with known provenance - 14 written by humans and 14 generated by AI. Since the three experiments we conducted differed in participant makeup, procedure, prior information, and which texts were included, we treated each wave as its own evaluation regime rather than combining them into one uniform condition. To keep things clear, we refer to the participant groups as Cohort 1, Cohort 2, and Cohort 3, which correspond to evaluation waves 1 (W1), 2 (W2), and 3 (W3), respectively. We use cohort labels when talking about who took part, and W1–W3 when referring to the evaluation regimes themselves or to comparisons between them.

In W1 (Cohort 1), experienced linguists assessed texts together in small groups outside a controlled laboratory environment, and they could use any analytical tools they wanted. They received no background information about semantic features that might distinguish AI-generated texts. Different groups reviewed different parts of the dataset, resulting in 75 group-level judgments of source across 27 texts. From the correctness of those judgments and the known

origin of each text, we reconstructed how often AI attribution occurred at the group level.

In W2 (Cohort 2), six experienced linguists evaluated all 28 texts individually on site. Before starting, they listened to an informational presentation that described semantic differences that had already been identified between human-authored and AI-generated texts in the corpus. They then gave a binary source attribution along with a confidence rating for each text. W2 was handled as an informed evaluation regime rather than a training study.

In W3 (Cohort 3), 16 linguists assessed a fixed subset of 10 texts individually. They were not given any prior information about semantic cues related to AI. For each text, they first completed 13 perceptual ratings, and then made a source judgment using a five-point Human–AI scale, followed by a confidence rating.

Thus, W1 and W2 overlapped on 27 texts, W2 and W3 overlapped on 10 texts, and W1 and W3 overlapped on nine texts.

In all three evaluation regimes, participants were instructed to assess each text independently, without comparing it with the other stimuli.

2.2. Text stimuli

The stimulus set consisted of 28 short texts in Russian: 14 written by humans and 14 generated by AI. The human-authored pieces were taken from a number of earlier text-production studies involving participants who did not participate in any of the evaluation procedures described here. The topics were chosen to reflect short, believable pieces of everyday communication that could reasonably occur in applied text-analysis settings. Accordingly, the corpus was deliberately varied in both topic and communicative context, rather than being limited to a single genre.

The AI-generated texts were produced using several modern LLM and straightforward prompts based on specific topics. Besides the topic, the prompts added

only a limited number of surface-level requirements. The texts had to include between 5 and 10 sentences, be written in a conversational style, avoid numbered lists, include emojis, and allow for typographical and spelling errors. The AI outputs were used exactly as generated, with no post-generation editing, correction, or normalization.

For W3, a balanced subset of 10 stimuli was selected from the full set of 28 texts, comprising five human-authored and five AI-generated texts. The subset was chosen to preserve the range of topics, communicative contexts, and text types represented in the full stimulus set.

2.3. Source Attribution and perceptual measures

For each judgment, we recorded whether the participant classified the text as Human or AI-generated. We then calculated, for each text, the proportion of judgments in which it was classified as AI-generated. This proportion served as our measure of perceived AI-likeness: the more often a text was judged to be AI-generated, the more AI-like it was considered to be. Importantly, this measure is independent of whether the judgment was correct. A high AI-attribution rate means correct identification for an AI-generated text, but false AI attribution for a human-authored text.

For W1, AI-attribution rates were derived from group-level accuracy. W2 included binary Human/AI judgments that were provided directly. In W3, participants made source judgments using a five-category ordered scale ranging from definitely Human to definitely AI. For the binary analyses, the two Human and two AI categories were combined, and the three responses in the undecided category were omitted; the original five-category response was kept for the ordinal sensitivity analysis. Confidence was not included in the primary analyses.

Before performing source attribution, Cohort 3 participants rated each text along 13 five-point perceptual dimensions related to experiential, imagistic, authorial, schematic,

affective, and trust-related characteristics. These ratings were treated as perceptual responses from readers to texts, rather than as objective linguistic features. The complete W3 questionnaire is available in the Supplementary Materials, both in the original Russian and in an English translation.

2.4. Linguistic representation of the texts

Each text was represented in two predefined linguistic spaces.

The semantic representation was obtained using the Semantic Domain Profiling framework (Litvinova et al., 2024). The framework represents each text by its degree of semantic alignment with 46 interpretable domains covering perceptual, sensorimotor, emotional, cognitive, spatial, and social aspects of meaning. The domain system was developed from psycholinguistic and neurocognitive semantic-feature resources.

Semantic alignment was estimated in a pretrained Russian embedding space (ruwikiruscorpora_upos_cbow_300_10_2021). Each domain was represented by a prototype derived from semantically characteristic lexical items; for some features, opposing lexical sets were used to define a semantic direction. Text-to-domain alignment was quantified with Concept Mover's Distance (CMD) using the text2map implementation (Taylor and Stoltz, 2021; Stoltz and Taylor, 2022). CMD measures the transport distance between the semantic distribution of a document and a target concept representation in embedding space. Scores were oriented so that higher values indicated stronger alignment with the corresponding semantic domain or direction. The resulting 46 values constituted the semantic profile of each text. These measures therefore describe a text's position relative to semantic prototypes in embedding space and should not be interpreted as frequencies or counts of particular lexical categories.

The stylometric representation included 12 predefined surface and structural measures covering text length, sentence structure,

lexical complexity and diversity, the use of function words and pronouns, noun–verb balance, and punctuation (Litvinova et al., 2025). The full set of features is listed in the Supplementary Materials.

Three coarse content-orientation ratings – Experience, Opinion, and Emotion – were manually assigned to each text by one researcher. Each dimension was scored on a three-point scale: 0 = not expressed, 1 = weakly expressed, and 2 = clearly expressed. Experience captured the presence of experiential content, Opinion the expression of personal views or evaluative stance, and Emotion the presence of explicitly expressed emotional content. These ratings were not treated as primary measures and were used only as covariates in supplementary robustness analyses examining whether the origin-related latent dimensions could be accounted for by broad differences in content orientation.

2.5. Statistical analysis

2.5.1. Attribution performance and cross-regime stability

Attribution performance was reported separately for W1, W2, and W3. To evaluate whether perceived AI-likeness was stable across different regimes, we looked at text-level AI-attribution rates from matched stimuli and used Spearman correlations, bootstrap confidence intervals, and permutation-based inference. Any comparisons within the same origin were considered exploratory.

2.5.2. Objective linguistic signal of actual origin

Individual semantic and stylometric features were linked to the actual Human/AI origin using bias-reduced logistic regression, with Benjamini-Hochberg correction applied separately within each of the 46 semantic and 12 stylometric feature groups.

To study the underlying structure, we used unsupervised PCA separately in semantic and stylometric space. We then quantified Human-AI separation along the main components using Hedges' g and rank-

based associations with the true origin. To check that the findings were reliable, we used permutation tests and leave-one-text-out analyses. We also assessed multivariate separation using centroid-based tests.

Next, we tested how much predictive information was available about the actual origin by running nested leave-one-text-out ridge logistic regression on semantic, stylometric, and combined feature sets. Importantly, all preprocessing and parameter tuning were confined to the training data. Predictive discrimination was summarized using ROC AUC, while the Brier score and accuracy were reported as complementary measures of performance. Finally, secondary models were used to adjust the dominant origin-related components for Experience, Opinion, and Emotion.

2.5.3. Objective origin signatures versus AI attribution

To check whether the key linguistic dimensions tied to actual origin also account for perceived AI-likeness, we examined relationships between semantic and stylometric PC1 scores and AI attribution separately within each evaluation regime.

For the text-level analyses, we relied on Spearman correlations and origin-adjusted grouped-binomial models. We also looked at W2 at the level of individual responses, using mixed-effects logistic regression with random intercepts for both linguists and texts. For the W2 correlations within each origin group, we tested whether the observed associations remained present among texts that shared the same true source.

2.5.4. Regime-dependent linguistic cue weighting

We estimated regime-specific links between each of the 46 semantic features and 12 stylometric features and AI attribution, while accounting for actual origin. The features were standardized once across the entire corpus so that the coefficients could be compared across evaluation regimes. We then compared the coefficient patterns from matched stimuli descriptively, using rank

correlations and leave-one-text-out sensitivity analyses.

We tested whether there were formal differences across regimes by running evaluation-regime $\times$ feature interaction models on the matched stimuli. For the W2-W3 comparisons, we used individual judgments for the 10 texts that were shared and fitted mixed-effects logistic models with linguist and text as random intercepts. For the W1-W2 and W1-W3 comparisons, we relied on grouped decision counts from their respective matched stimulus sets. We applied the Benjamini-Hochberg correction separately to the semantic and stylometric interaction families, and we used leave-one-text-out analyses to check how robust the interactions that remained significant under FDR were.

We interpreted all cross-regime effects as being associated with the regime rather than causal, since participant makeup, procedures, and informational conditions differed at the same time across regimes.

2.5.5. Perceptual structure in Cohort 3 and AI attribution

For W3, each perceptual rating was initially analyzed in its own mixed-effects logistic regression model. The model predicted binary AI attribution while controlling for the actual origin. It included crossed random intercepts for both the linguist and the text, and applied the Benjamini-Hochberg correction across the 13 predictors.

To separate stable differences across texts from perceptions that vary by reader, the ratings were split into between-text means and within-text deviations. In addition, to capture the shared structure across the 13 ratings, we ran a PCA using within-linguist centered responses. Component retention was determined through permutation-based parallel analysis. Importantly, neither actual origin nor attribution was included in the PCA.

We then tested the retained perceptual component in relation to AI attribution using crossed mixed-effects logistic models, again using the between-text and within-text

decomposition. As a sensitivity check, we also fit a cumulative-link mixed model to the original five-category attribution responses, instead of collapsing them into a binary outcome.

All statistical analyses were conducted in R (version 4.6.0).

3. RESULTS

3.1. Actual text origin was encoded in distributed semantic and stylometric structure

We first checked whether the two source classes differed in their objective linguistic structure, without considering source-attribution judgments.

3.1.1. Semantic structure

At the level of individual semantic features, none of the 46 predefined semantic variables survived Benjamini-Hochberg correction. Still, eight effects clustered just above the threshold ($q = .0555$). AI origin was positively linked with *SomatNociception*, *AudLIWC*, *VisBody*, *TemporalLIWC*, *SocialLIWC*, and *EmoFear*, while *EmoAngry* and *VisMotion* leaned in the opposite direction. Overall, this looked more like a distributed origin difference than a contrast driven by specific markers (Supplementary Table S1).

The multivariate results were clearer. In the unsupervised semantic PCA, the first four components explained 81.3% of the variance, with PC1 alone accounting for 35.4%. AI-generated texts scored higher on PC1 than human-authored texts (mean difference = 4.79, Hedges' $g = 1.42$, Spearman's $\rho = .628$, permutation $p = .00095$, $q = .004$). By contrast, PC2 through PC4 showed no origin effect of a similar magnitude. Importantly, the PC1 split stayed strong in leave-one-text-out analyses ($g = 1.35$ – 1.75 ; Table 1).

At the AI-associated end of semantic PC1, the strongest pattern was alignment with several auditory, visual, bodily, motor, spatial, and temporal dimensions. The opposite end was more strongly associated with motion, taste, and several affective dimensions. These variables reflect

embedding-based semantic alignments rather than simple lexical counts. Human and AI groups were also separated within the four-dimensional semantic PCA space (squared distance = 27.4, permutation $p = .00050$; Supplementary Table S2).

3.1.2. Stylometric structure

The 12 stylometric measures showed a closely related multivariate pattern. Again, no single stylometric feature survived Benjamini-Hochberg correction (Supplementary Table S1). However, stylometric PC1 – explaining 29.2% of the variance – did strongly distinguish the actual source ($|Hedges' g| = 1.43$, $|\rho| = .628$, permutation $p = .00075$, $q = .004$). The effect remained substantial in the leave-one-text-out analysis ($|g| = 1.35$ – 1.63 ; Table 1), while PC2 through PC5 showed no comparable association with origin (Supplementary Table S2).

For interpretation, stylometric PC1 was oriented so that positive scores represented the AI-associated pole. This side of the dimension was mainly characterized by higher rates of pronouns, first-person terms, and function words. The human-associated pole showed greater lexical density, longer words, higher MATTR, and a higher noun-to-verb ratio. Sentence length and punctuation-

related variables contributed more weakly. Figure 1 presents both latent structures.

The two origin-based dimensions were only moderately connected (Spearman's $\rho = .338$, $p = .079$), indicating that the semantic and stylometric spaces reflected somewhat different facets of the Human-AI contrast. On a descriptive level, 17 out of 28 texts fell on the origin-consistent side of both PC1 axes. The remaining 11 texts were origin-consistent on one axis but not the other, and none of the texts were inconsistent with their true origin on both axes.

We then tested whether the original semantic and stylometric feature sets could discriminate actual text origin out of sample. Nested leave-one-text-out ridge models were fitted directly to the 46 semantic features, the 12 stylometric features, and their combined 58-feature representation, rather than to PCA scores. AUC values were .832 for semantics, .878 for stylometry, and .862 for the combined representation. The corresponding Brier scores were .182, .135, and .172, while accuracies were .786, .750, and .786.

These findings are summarized in Table 1, and further leave-one-text-out model information is provided in Supplementary Table S3.

Figure 1. Distributed semantic and stylistic structure of actual text origin: A) Scores of the 28 texts on the first two components of the unsupervised semantic PCA; B) The six strongest loadings at each pole of semantic PC1, oriented so that negative values represent the human-associated pole and positive values the AI-associated pole; C) Scores on the first two components of the unsupervised stylistic PCA; (D) Loadings of all 12 stylistic variables on PC1, using the same orientation

Рисунок 1. Распределённая семантическая и стилометрическая структура реального происхождения текста: А) Положение 28 текстов на первых двух компонентах семантического PCA, построенного без учёта информации об источнике текста; В) Шесть признаков с наибольшими нагрузками по модулю на каждом полюсе первой семантической компоненты (PC1); отрицательные значения соответствуют полюсу, связанному с человеческими текстами, положительные – связанному с ИИ-текстами; С) Положение текстов на первых двух компонентах стилометрического PCA; D) Нагрузки всех 12 стилометрических признаков на PC1 в той же ориентации

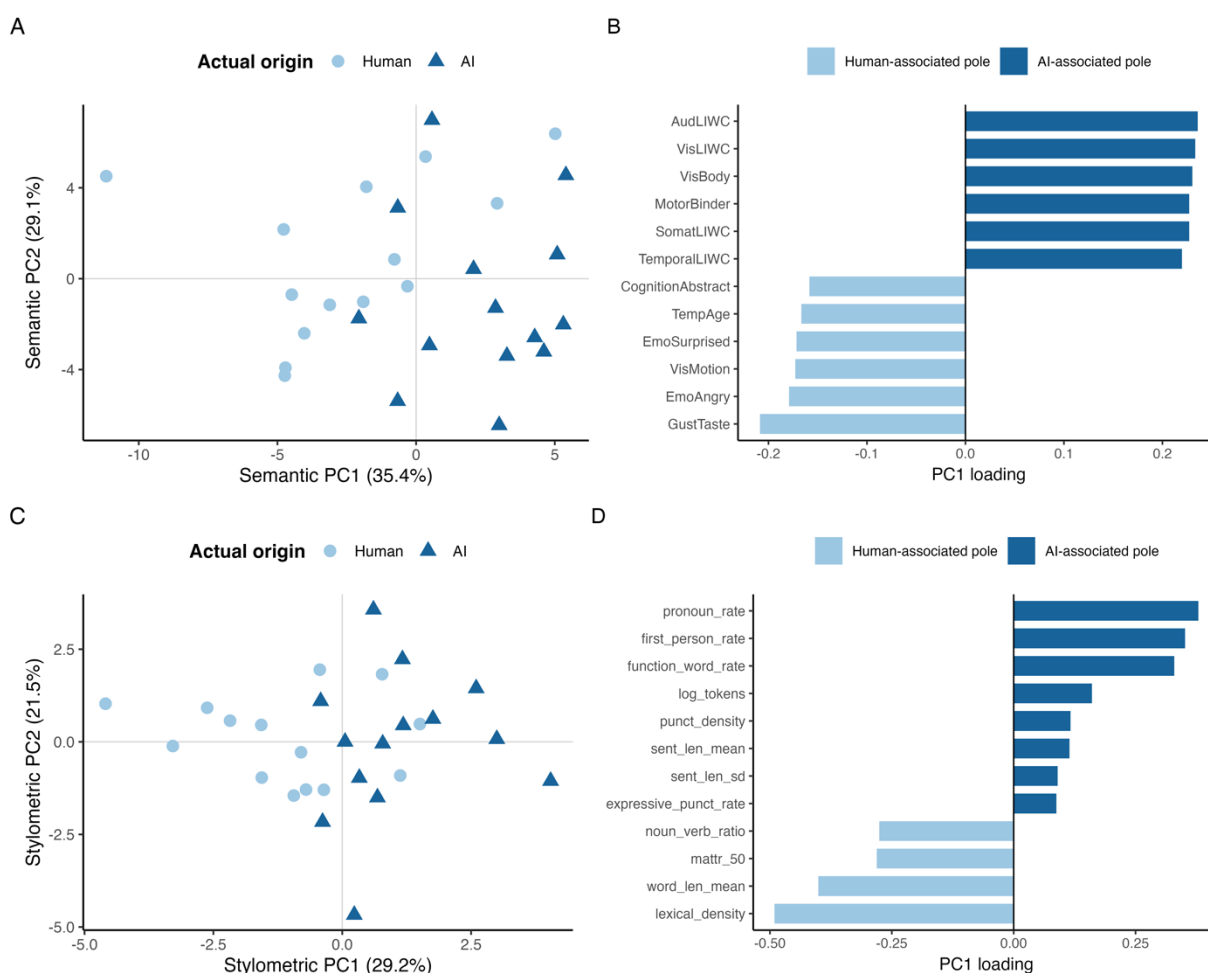


Table 1. Objective linguistic structure and out-of-sample discrimination of actual text origin

Таблица 1. Объективная лингвистическая структура и качество классификации реального происхождения текста на данных, не использованных при обучении

Feature space	PC1 variance (%)	[Hedges' g]	PC1 permutation p	LOTO g range	Ridge AUC	Brier score	Accuracy
Semantic	35.4	1.42	9.5×10^{-4}	1.35–1.75	.832	.182	.786
Stylometric	29.2	1.43	7.5×10^{-4}	1.35–1.63	.878	.135	.750
Combined	-	-	-	-	.862	.172	.786

Even after adjusting for the researcher-coded Experience, Opinion, and Emotion dimensions, the latent structure tied to origin was not removed. Semantic PC1 continued to track actual origin after we made all adjustments at once (OR = 9.49 per SD, $p = .028$), and the same was true for stylometric PC1 (OR = 5.92, $p = .021$). Semantic PC1 overlapped only partially with the broader content orientation, especially with Experience ($\rho = .600$, $q = .004$) and Opinion ($\rho = -.492$, $q = .023$), but the Human–AI difference could not be reduced to these coarse content dimensions (Supplementary Table S4).

Overall, actual origin was reflected in two strong distributed configurations – one semantic and one stylometric – rather than in isolated diagnostic markers. The PCA results revealed these latent structures without using source labels, while the ridge models showed that the original feature spaces also contained enough information to discriminate held-out texts.

3.2. AI attribution varied across evaluation regimes and perceived AI-likeness was not stable

The source-attribution results varied markedly across the three evaluation settings (Fig. 2). In W1, 41 out of 75 collective

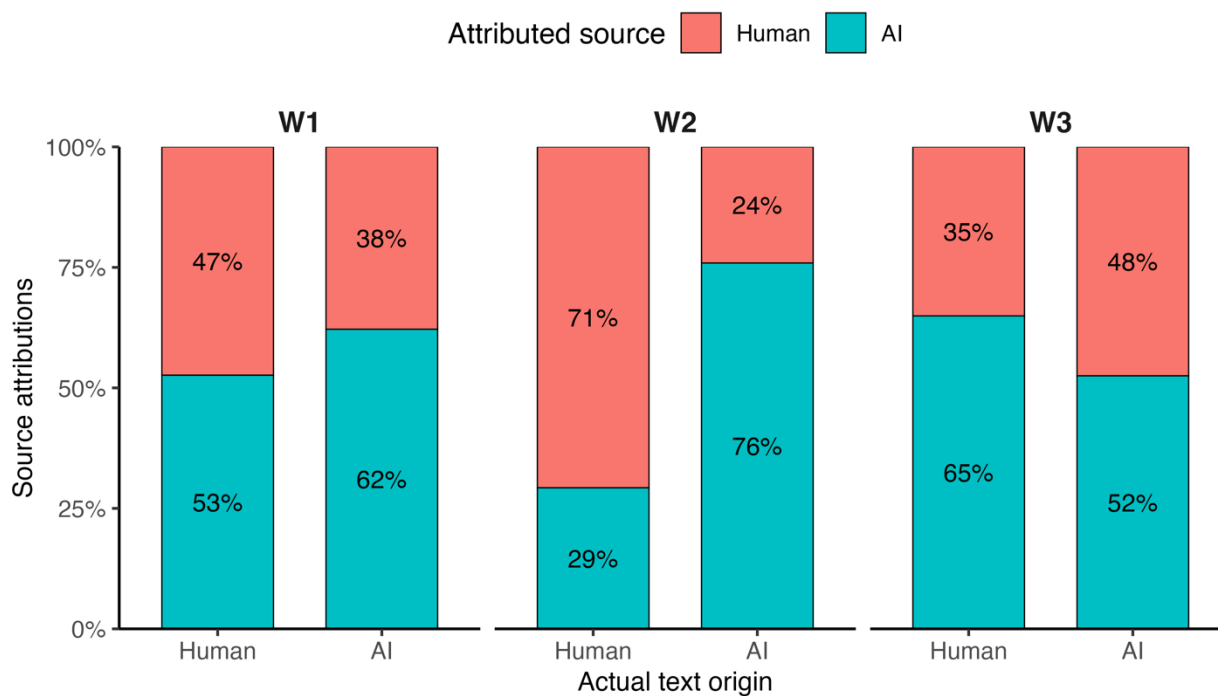
judgments were correct, for an accuracy of 54.7%. Accuracy was noticeably better for AI-generated texts: 23 out of 37 (62.2%) correct, compared with human-authored texts, where 18 out of 38 (47.4%) were correct. Overall, 43 of the 75 W1 judgments (57.3%) identified the text as AI, suggesting only a mild bias toward attributing to AI.

W2 produced the best descriptive accuracy. Here, 121 of 165 valid individual judgments were correct (73.3%). Performance was fairly even across the two true origins: 63 of 83 judgments for AI-generated texts were correct (75.9%), and 58 of 82 judgments for human-authored texts were correct (70.7%). In total, 52.7% of W2 judgments attributed the text to AI.

In contrast, W3 showed much lower accuracy and a stronger inclination toward attributing to AI. Only 69 of 157 binary judgments were correct (43.9%). Accuracy was 52.5% for AI-generated texts (42/80), but dropped to 35.1% for human-authored texts (27/77). Moreover, 58.6% of all valid W3 judgments attributed the text to AI. As a result, the poor accuracy in W3 appears to be driven mainly by false-AI attribution of human-authored texts, rather than by balanced uncertainty between the two source categories.

Figure 2. Source-attribution patterns across the three evaluation regimes. Proportions of Human and AI attributions are shown separately for human-authored and AI-generated texts.

Рисунок 2. Распределение атрибуций источника текста в трёх режимах оценки. Доли ответов «человек» и «ИИ» показаны отдельно для текстов, написанных людьми, и текстов, сгенерированных ИИ.



These asymmetries are the reason the analysis now moves away from overall accuracy and toward perceived AI-likeness at the text level. Instead of asking only how often attribution breaks down, the key issue is whether the same texts are consistently judged as AI-like under different evaluation conditions.

Perceived AI-likeness was not stable across evaluation regimes (Supplementary Table S5). Of the 27 texts shared by W1 and W2, text-level AI-attribution rates were basically unrelated (Spearman's $\rho = -.07$, bootstrap 95% CI $[-.45, .34]$, permutation $p = .72$). The W1-W3 correlation was also weak across nine shared texts ($\rho = .33$, 95% CI $[-.46, .83]$, $p = .38$), and the W2-W3 correlation across 10 texts was similarly weak ($\rho = -.25$, 95% CI $[-.70, .38]$, $p = .48$). Overall, there was little evidence that texts kept the same ordering from more human-like

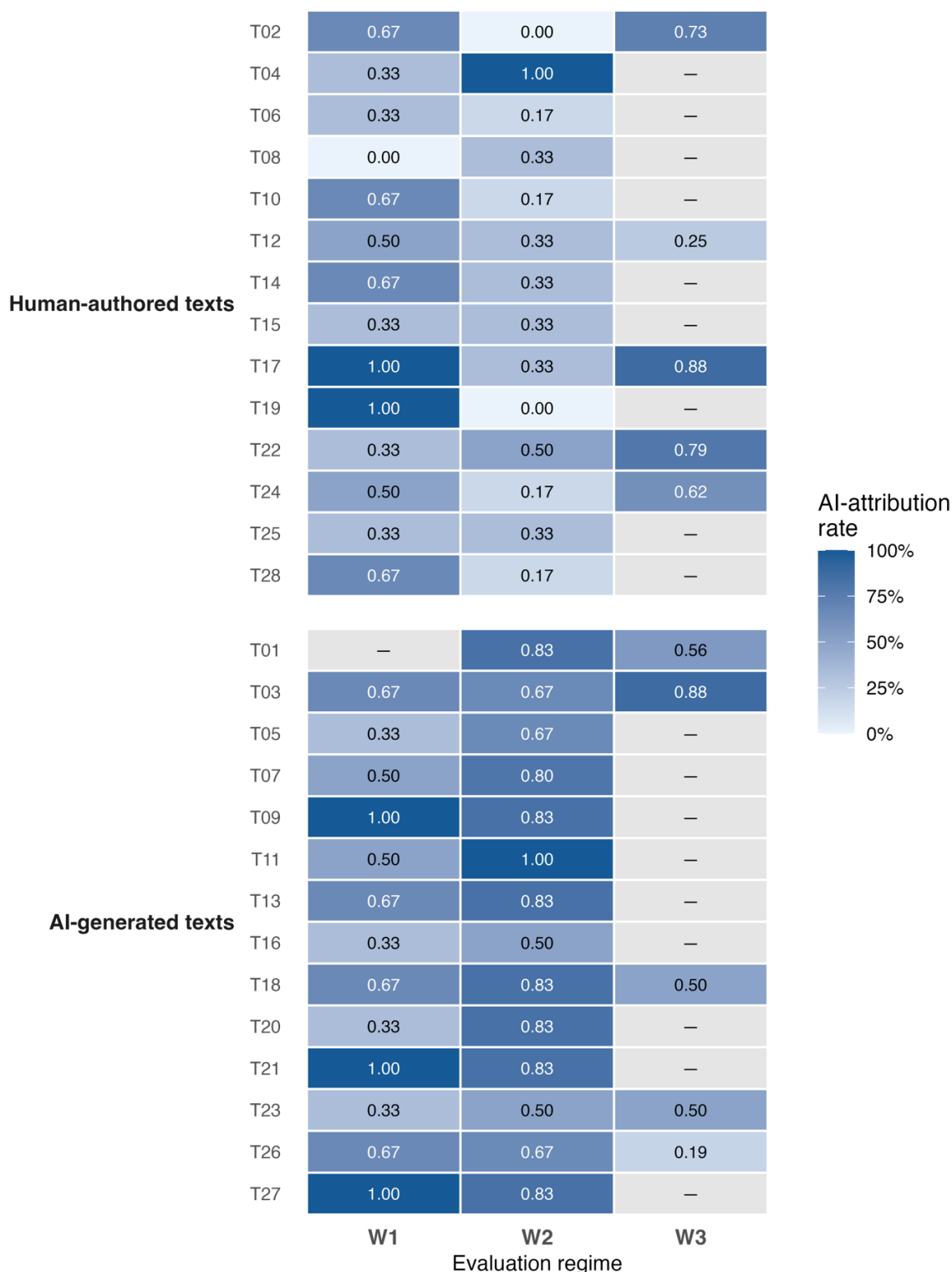
to more AI-like across different evaluation regimes.

This instability was especially clear in the W3 subset. For the same 10 texts, the mean AI-attribution rate in W2 was .700 for AI-generated texts and .267 for human-authored texts. In W3, however, the pattern reversed: the corresponding rates were .525 and .654, respectively. Among the nine stimuli that were also evaluated in W1, the mean rates were .585 for AI-generated texts and .600 for human-authored texts.

Individual texts highlight this dependence even more (Figure 3). The AI-generated T03 stayed consistently AI-like across W1, W2, and W3 (.67, .67, and .88). By contrast, the human-authored T02 shifted from .67 in W1 to 0 in W2 and then to .73 in W3, and T17 moved from 1.00 to .33 and then to .88. Meanwhile, the AI-generated T26 moved in the opposite direction, going from .67 in both W1 and W2 down to .19 in W3.

Figure 3. Text-level perceived AI-likeness across evaluation regimes. AI-attribution rates are shown for all 28 stimuli, separately for human-authored and AI-generated texts and ordered by text identifier within each origin group.

Рисунок 3. Воспринимаемая ИИ-похожесть отдельных текстов в разных режимах оценки. Доли атрибуций ИИ показаны для всех 28 текстов; внутри каждой группы тексты упорядочены по их идентификаторам.



Perceived AI-likeness, however, wasn't consistent across different evaluation setups. While some texts repeatedly appeared more AI-like or more human-like, others shifted their classification considerably depending on the particular evaluation setting.

3.3. Objective linguistic structure did not map consistently onto AI attribution

We next checked whether the linguistic features that distinguished the two true origins were also the same features that made the texts seem AI-generated. For the most part, the answer was no.

Looking at the text level, neither the semantic dimension nor the stylometric dimension of PC1 was linked to the AI-attribution rate in W1 (semantic: $\rho = .190$, $p = .342$; stylometric: $\rho = .068$, $p = .736$) or in W3 (semantic: $\rho = -.311$, $p = .382$; stylometric: $\rho = -.317$, $p = .372$). W2 seemed different at first: AI-attribution rates correlated with both the AI-leaning semantic PC1 ($\rho = .425$, $p = .024$) and the stylometric PC1 ($\rho = .518$, $p = .0048$). But once we adjusted for the actual origin, those links disappeared. In the grouped, bias-reduced models, neither semantic PC1 (OR = .935, $p = .760$) nor stylometric PC1 (OR = 1.16, $p = .516$) independently predicted W2 AI attribution.

The same conclusion showed up at the individual-response level. In a mixed-effects logistic model that included both objective dimensions, actual AI origin strongly increased the odds of an AI attribution (OR = 6.97, 95% CI [2.55, 19.0], $p = .00015$). By contrast, semantic PC1 (OR = .928, 95% CI [.601, 1.43], $p = .737$) and stylometric PC1 (OR = 1.17, 95% CI [.750, 1.84], $p = .484$) did not add independent predictive power. Within-origin correlations in W2 were also weak, and once again, origin-adjusted effects stayed non-significant in W1 and W3 (Supplementary Table S6).

Overall, the structure of actual origin and the structure of AI attribution were clearly separable. The latent dimensions that most strongly distinguished human-written texts from AI-generated ones did not reliably

explain differences in perceived AI-likeness once actual origin was accounted for.

We can see this dissociation in individual examples. Human-authored T17 landed on the AI-associated side of the semantic axis and was often attributed to AI. Meanwhile, T02 attracted high false-AI attribution even though it sat on the human-associated side of that same axis. Conversely, AI-generated T26 fell on the human-associated side of the stylometric axis and was mostly judged as Human in W3. So, at the level of single texts, an objective position in the linguistic space tied to origin did not consistently map onto perceived AI-likeness.

AI generation did leave a measurable semantic and stylometric signature in this dataset, but readers did not end up treating that signature as a stable perceptual cue for source attribution.

That finding points to a more specific possibility: maybe particular objective features still carry similar attributional meaning across different regimes, even if the main origin dimensions do not. To examine this, we estimated regime-specific associations for all 46 semantic and 12 stylometric features, controlling for actual origin, and then compared matched stimulus sets using fixed full-corpus standardization.

At the level of complete feature profiles, cross-regime correspondence was limited. For semantic features, the matched-stimulus coefficient vectors correlated negatively between W1 and W2 (Spearman's $\rho = -.310$; 27 shared texts) and between W2 and W3 ($\rho = -.567$; 10 shared texts). In contrast, W1 and W3 showed essentially no correspondence ($\rho = .070$; nine shared texts). Stylometric profiles showed weaker and less consistent agreement ($\rho = .042$, .266, and .378, respectively). Leave-one-text-out analyses largely kept the negative semantic correspondence patterns for W1–W2 and W2–W3, though the W2–W3 result was somewhat sensitive to leaving out one stimulus (Supplementary Table S7). In short, the cue profile linked to AI attribution did not carry over across the three regimes.

The clearest differences appeared in the direct W2-W3 comparison, using the same 10 stimuli. Seven Wave $\times$ Feature interactions remained significant after FDR correction (four semantic and three stylometric) and all seven involved a reversal in the direction of association between W2 and W3 (Supplementary Table S8). For stylometric features, mean word length flipped from a negative association with AI attribution in W2 ($\beta = -.871$) to a positive association in W3 ($\beta = .631$; interaction OR = 4.49, $p = .00022$). Lexical density showed the same pattern of reversal ($\beta = -.987$ vs. $.261$; interaction OR = 3.48, $p = .00037$), whereas function-word rate reversed in the opposite direction ($\beta = .416$ vs. $-.483$; interaction OR = .407, $p = .0020$).

These stylometric reversals are especially striking when viewed alongside the actual-origin structure described in Section 3.1. In the full corpus, greater use of function words characterized the AI-associated stylometric pole, whereas longer words and higher lexical density characterized the human-associated pole. The W2 slopes for all three features followed the same direction as that objective contrast, but the W3 slopes went in the opposite direction. This alignment is descriptive and does not mean that W2 participants consciously adopted the objective origin signature.

Four semantic alignment dimensions displayed similar reversals. *EmoPleasant* moved from $\beta = -1.08$ in W2 to $\beta = 1.34$ in W3 (interaction OR = 11.2, $p = .00042$); *EmoSentiment* shifted from $\beta = 1.14$ to $-.781$ (OR = .147, $p = .00072$); *EmoSurprised* went from $\beta = -.675$ to $.440$ (OR = 3.05, $p = .0017$); and *SpatialNumber* changed from $\beta = 1.06$ to $-.641$ (OR = .183, $p = .0019$). Since these predictors are embedding-based

measures of semantic alignment, the reversals reflect changes in the attributional relevance of semantic positioning toward these domains, rather than straightforward effects of lexical frequency.

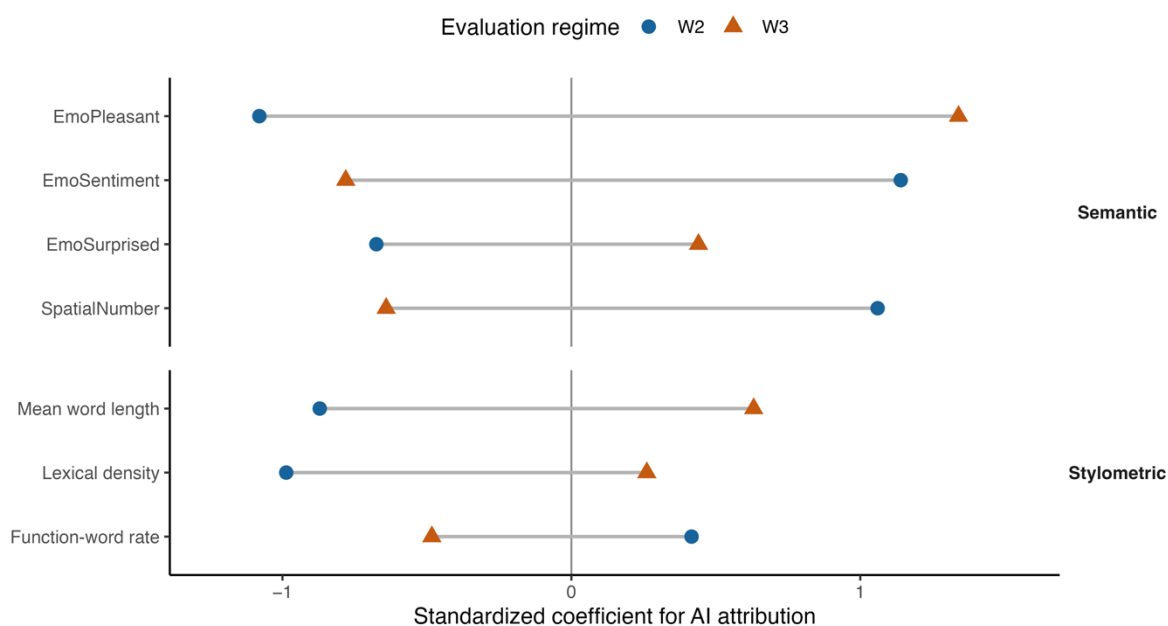
These interaction patterns also proved highly stable when stimuli were deleted. Across all seven W2-W3 features, the sign of the Wave $\times$ Feature coefficient stayed the same in all 10 leave-one-text-out analyses. In six of the seven features, the wave-specific slopes reversed in every iteration, and mean word length did so in 9 out of 10 cases (Supplementary Table S8).

The W1-W2 comparison was different in scope. Of the 27 shared texts, no semantic interactions remained after FDR correction, and only one stylometric variable emerged: mean word length. Longer words corresponded to higher AI attribution in W1 ($\beta = .533$) but lower AI attribution in W2 ($\beta = -.448$), producing an interaction OR of .375. This reversal held the same direction even after removing each individual stimulus. For W1-W3, no FDR-significant interactions were found, although this analysis involved only nine shared texts (Supplementary Table S9). The main regime-dependent reversals are summarized in Figure 4. The language signals linked to AI attribution thus differed depending on the evaluation setup. The biggest contrast showed up between W2 and W3, where seven semantic or stylometric cues shifted in the opposite direction; in comparison, W1 and W2 diverged in a more limited way. That means a feature like lexical density, average word length, or the rate of function words could be tied to AI attribution under one regime and to Human attribution under another. in Figure 4.

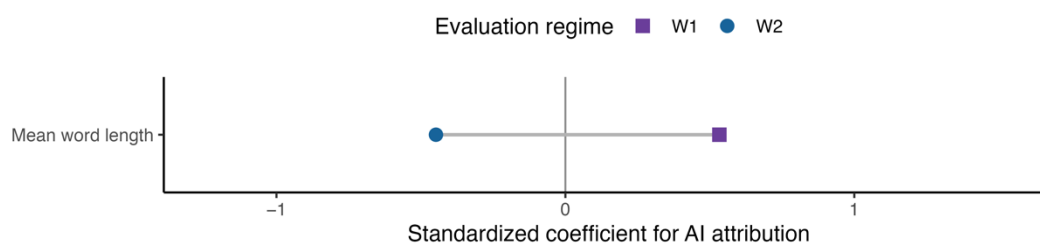
Figure 4. Regime-dependent differences in linguistic cue associations with AI attribution:
A) Regime-specific standardized coefficients for the seven semantic and stylometric features showing FDR-significant W2-W3 Wave $\times$ Feature interactions; all seven reversed direction between regimes; B) Mean word length, the single feature showing an FDR-significant W1-W2 interaction; C) No semantic or stylometric Wave $\times$ Feature interaction survived FDR correction in the W1-W3 comparison, which was based on nine shared texts.

Рисунок 4. Различия между режимами оценки в связи языковых признаков с атрибуцией ИИ: А) Стандартизованные коэффициенты для семи семантических и стилометрических признаков, у которых во сравнении W2-W3 взаимодействие *режим* $\times$ *признак* осталось значимым после FDR-коррекции; во всех семи случаях направление связи между режимами менялось на противоположное; Б) Средняя длина слова – единственный признак, для которого взаимодействие W1-W2 осталось значимым после FDR-коррекции; С) В сравнении W1-W3, основанном на девяти общих текстах, ни для одного семантического или стилометрического признака взаимодействие *режим* $\times$ *признак* не сохранило значимость после FDR-коррекции.

A



B



C

No FDR-significant Wave $\times$ Feature interactions
46 semantic + 12 stylometric features
9 shared texts
W1 vs W3

3.4. Perceived AI-likeness in Cohort 3 was organized by authorial-experiential presence versus schematic generality

The W3 ratings gave a different perspective on what readers were reacting to. Out of the 13 perceptual dimensions, 11 were linked to AI attribution after Benjamini-Hochberg correction, using separate crossed mixed-effects models that controlled for the texts' actual origin. AI attribution became more likely with greater generality, template-likeness, and obviousness. In contrast, stronger signals of authorial belief and emotion, along with personal and physical experience, author imaginability, vividness, pleasantness, and trust all lowered the odds of attributing the text to AI. Auditory and scene imagery, however, did not hold up after correction (see Supplementary Table S10).

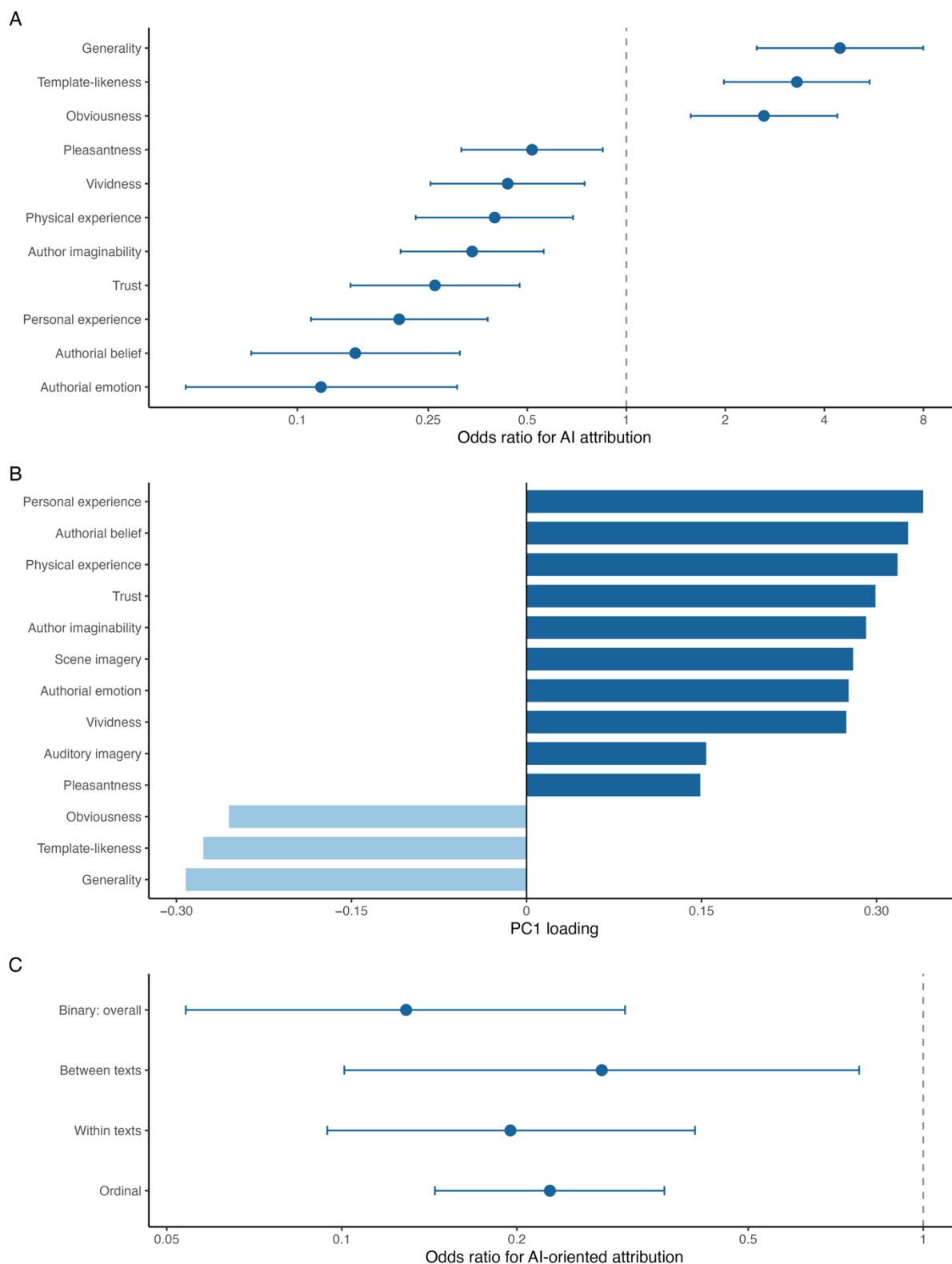
The pattern showed up at two different levels. At the between-text level, seven dimensions kept FDR-significant differences, while at the within-text level, 12 out of 13 displayed significant effects. On average, texts that readers perceived as more general or template-like drew more AI attributions. In contrast, texts that were seen as more

personally grounded, imaginable, emotionally situated, and trustworthy led to fewer AI attributions. Similar relationships also appeared across different readers evaluating the same stimulus, which suggests the finding wasn't just a side effect of consistent differences among the 10 texts (Supplementary Table S11).

A principal component analysis (PCA) on the within-linguist centered ratings captured this coordination using a single retained component. PC1 had an eigenvalue of 5.90, accounted for 45.4% of the variance, and was larger than the 95th-percentile null eigenvalue of 1.74 in permutation-based parallel analysis. Variables such as personal experience, authorial belief, physical experience, trust, author imaginability, scene imagery, authorial emotion, and vividness loaded positively. Meanwhile, generality, template-likeness, and obviousness loaded negatively. We interpret this component as reflecting a trade-off between authorial-experiential presence and schematic generality (Supplementary Table S12). Figure 5 summarizes the individual effects and the latent structure.

Figure 5. Perceptual organization of AI attribution in Cohort 3: A) Effect estimates for the 11 perceptual dimensions significantly associated with binary AI attribution after Benjamini-Hochberg correction. Odds ratios below 1 indicate lower odds of AI attribution and values above 1 higher odds; B) Loadings of all 13 perceptual ratings on the retained unsupervised component, oriented from schematic generality at the negative pole to authorial-experiential presence at the positive pole; C) Associations of the retained component with AI attribution in the primary binary model, its between-text and within-text decomposition, and the ordinal sensitivity analysis.

Рисунок 5. Перцептивная организация атрибуции ИИ в когорте 3: А) Оценки эффектов для 11 перцептивных шкал, которые сохранили значимую связь с бинарной атрибуцией ИИ после коррекции Бенджамини-Хохберга. Отношения шансов меньше 1 соответствуют снижению вероятности атрибуции ИИ, значения больше 1 – её увеличению; В) Нагрузки всех 13 перцептивных оценок на единственную сохранённую компоненту, полученную без учёта атрибуций; отрицательный полюс соответствует схематической обобщённости, положительный – авторскому присутствию; С) Связь сохранённой компоненты с атрибуцией ИИ в основной бинарной модели, при разложении эффекта на межтекстовую и внутритекстовую составляющие, а также в дополнительном анализе устойчивости с исходной порядковой шкалой.



Within W3, shifting toward the authorial-experiential end of the spectrum was closely linked to fewer AI attributions. In the main crossed mixed-effects model, a one-standard-deviation increase in PC1 lowered

the likelihood of being assigned an AI attribution (OR = .129, 95% CI [.054, .307], $p = 3.80 \times 10^{-6}$). By contrast, actual origin was not significant (OR = .849, $p = .863$). This relationship held both across the 10

different stimuli (OR = .280, 95% CI [.101, .777], $p = .0145$) and within individual texts across different readers (OR = .195, 95% CI [.094, .405], $p = 1.10 \times 10^{-5}$).

The same overall pattern emerged when the original five-category attribution response was kept. In a cumulative-link mixed model that used all 160 evaluations, higher PC1 again reduced the odds of making a more AI-oriented attribution (OR = .228, 95% CI [.145, .359], $p = 1.76 \times 10^{-10}$), while actual origin stayed non-significant (OR = .712, 95% CI [.262, 1.94], $p = .507$; Supplementary Table S12).

Importantly, the component was derived without using source labels or attribution outcomes, yet it was strongly associated with subsequent AI attribution both across texts and among readers evaluating the same text.

4. DISCUSSION

The study started with a basic inconsistency: a text can be measurably AI-like in its linguistic structure while still not appearing AI-generated to readers, and a human-written text can still draw AI judgments. The findings refine this mismatch. Human- and AI-produced texts differed clearly in how semantic information and stylistic features were distributed, but these origin-related aspects didn't translate straightforwardly into attribution. Moreover, the influence readers seemed to place on individual cues shifted across different conditions, and in some cases even reversed direction. In Cohort 3 specifically, however, AI attribution seemed to follow a perceptual dimension that ran from schematic generality toward a stronger sense of the author's presence and lived experience.

By placing objective indicators of origin alongside readers' perceived AI-likeness on the same texts, the difference becomes clear. Previous research has already shown that human and AI writing can diverge linguistically and that readers rely on systematic cues when forming judgments. Here, those two patterns could be compared directly. They did not simply align, and the

relationship between any given linguistic cue and AI attribution was not consistent across the different evaluation regimes.

4.1. Actual AI generation leaves a distributed linguistic signature, but readers do not simply perceive that signature

The objective-origin results align with a growing body of computational-linguistics research suggesting that human and LLM writing can still be told apart by how their language is organized at scale. Recent work finds differences in stylistic organization, information density, stance, functional appropriateness, and multivariate stylometric patterns (Mizumoto et al., 2024; Litvinova et al., 2025; O'Sullivan, 2025; Chen et al., 2026; Demir, 2026).

In our corpus, we saw this separation across two only partly overlapping spaces. On the semantic side, the main dimension reflected broad differences in alignment with bodily, sensory, spatial, temporal, and affective domains. On the stylometric side, the contrast was between function-word and pronominal orientation versus lexical density, word length, lexical diversity, and the noun-to-verb relationship. These semantic variables should be understood as coordinates in embedding space, not as simple counts of particular word types.

After correction for multiple testing, individual semantic and stylometric predictors were weak. At the multivariate level, however, the pattern was much stronger. Unsupervised latent dimensions showed large Human-AI differences that remained stable in leave-one-text-out analyses, and nested leave-one-text-out ridge models fitted to the original feature sets discriminated held-out texts with AUCs of .832 for semantics, .878 for stylometry, and .862 for the combined representation. Thus, in this corpus, information about source origin was distributed across combinations of features rather than concentrated in a single diagnostic marker.

The more revealing finding came when we compared these dimensions with human source judgments. Once actual origin was accounted for, neither the semantic nor the stylometric axis on its own could explain AI attribution. Even the apparent match in W2 vanished in origin-adjusted and within-origin analyses. In other words, a computationally detectable Human-AI signal exists, but readers do not seem to reproduce it directly in their judgments.

This gap fits with research showing that people use systematic but imperfect heuristics for authorship. Jakesch et al. (2023) showed that cues people treat as diagnostic of AI authorship may not reflect their true underlying distribution, and more recent studies suggest that judgments rely on combinations of writing style, personal experience, genre expectations, and communicative context rather than a fixed checklist of formal indicators (Hau et al., 2026; Liu et al., 2026).

The takeaway isn't just that people sometimes misclassify AI. This distinction was also visible at the level of individual texts. A human-written text could appear AI-like without necessarily falling on the AI-associated side of objective linguistic space, while an AI-generated text could occupy a more Human-associated region and be judged accordingly. Objective origin discrimination and perceived AI-likeness should therefore be treated as distinct analytical tasks.

4.2. Linguistic cue weighting depended on the evaluation regime

Cross-regime comparison revealed a second kind of instability: the same linguistic cue did not consistently map onto the same attributional value across all settings. In the matched W2-W3 group, seven objective features produced FDR-significant Wave $\times$ Feature interactions, and each one flipped direction. Six of those reversals held up under every leave-one-text-out deletion, while the seventh held up in 9 out of 10. Mean word length also reversed between W1 and W2 and then stayed reversed across all 27 stimulus

deletions. No W1-W3 interaction survived FDR correction, though with a nine-text overlap there's limited statistical power to treat that null result as evidence of similarity.

This goes beyond a simple difference in overall detection accuracy. The attributional meaning of a feature itself shifted. In W2, using more function words went with AI attribution, whereas higher lexical density and longer words went with Human attribution – matching the same directional pattern observed in the objective stylometric origin contrast. In W3, all three of these relationships reversed. The texts hadn't changed; what changed was the link between their properties and the judgments.

This pattern aligns with accounts of authorship attribution that treat it as contextually framed rather than driven by a fixed set of diagnostic markers. Hau et al. (2026) argue that readers integrate textual cues with expectations about the communicative domain, the sender, and the broader context, and that surface cues can become misleading when they're used outside the conditions where they're diagnostic. Wang et al. (2026) similarly found large variation in Human-AI discrimination across languages and domains, and showed that knowledge about relevant Human-machine differences can shift detection performance.

W2 also deserves caution because it followed an informational presentation about semantic differences in the corpus. One plausible interpretation is that the presentation "corrected" participants' perceptual template and adjusted cue weighting so it better matched the objective origin signature. However, the design cannot confirm that conclusion. W2 and W3 also used different participant cohorts and different procedures, and W3 included perceptual ratings before source attribution. What the data support is a narrower claim: cue weighting differed across regimes, with the most prominent feature-level divergence occurring between W2 and W3.

Earlier studies already show that source-attribution performance varies across

genre, language, and informational context. Our results extend that context dependence to individual features. A measurable property of the text can change not only the strength of its association with AI attribution, but also the direction of that association.

This matters for how we understand perceived AI-likeness. A detector-like account would predict that features contribute with fairly stable weights to an underlying AI score. Our findings don't match that idea. The same property took on different attributional value under different evaluation conditions, which fits better with context-sensitive use of cues than with a single regime-invariant profile.

4.3. Perceived AI-likeness may reflect the perceived presence of a situated author

The perceptual ratings from Cohort 3 point to what readers may have been reacting to instead. In this set of 10 texts, AI attribution went up when the writing felt more general, more like a template, and more obviously constructed. It went down when readers thought it reflected personal and physical experience, authorial belief and emotion, a believable imagined author, vividness, and trust. These ratings all seemed to align along a main dimension we labeled *Authorial-experiential presence versus schematic generality*.

Importantly, this component wasn't built directly from the source labels or the attribution outcomes. Still, it was able to predict later attribution across the 10 stimuli, including among different readers who evaluated the same text. And the relationship held even when the original five-category response was examined. So, the effect can't simply be explained by the actual Human/AI origin, or by stable differences between the texts by themselves.

This view also fits with a growing body of work on authorship heuristics. Jakesch et al. (2023) found that readers treat personal and social content, among other cues, as evidence of Human authorship – even when those heuristics aren't reliable. Liu et al.

(2026) similarly identify references to personal experience as a recurring cue in judgments about perceived AI authorship. And Hau et al. (2026) argue that source attribution reflects a combination of textual evidence and expectations about the sender, genre, and communicative context, rather than relying on a fixed checklist.

One way to interpret our W3 pattern is that Human-likeness depends in part on whether a text supports the sense that it comes from a particular, situated source of experience and evaluation. Generality and template-likeness don't have to be direct "AI markers". Instead, they may reduce the impression that the discourse is coming from a specific communicator who is experiencing, believing, and emotionally situated. By contrast, personal experience, belief, emotion, vividness, and how imaginable the author feels may strengthen that impression.

This is a relational interpretation rather than an essentialist one. The data don't show that AI-generated writing inherently lacks authorial presence, or that human writing necessarily includes it. In our corpus, there are AI-like human texts and Human-like AI texts, which runs against the idea that authorial presence is intrinsically determined. Here, authorial presence is something readers perceive. A text written by a human may fail to evoke a plausible situated author, while an AI-generated text may succeed.

This distinction also helps explain why human attribution and computational origin classification can sometimes diverge. A classifier may take advantage of distributed statistical differences between the source populations without actually representing an "author." Readers may instead be implicitly asking whether the discourse feels like it was produced by a situated Human communicator. In that sense, perceived AI-likeness may involve an inference about communicative authorship rather than a direct readout of the statistical signature of AI generation.

4.4. From AI detection to attribution risk: why perceived AI-likeness matters

Actual origin and perceived AI-likeness are important for different reasons. Detecting origin focuses on whether AI produced the text. Perceived AI-likeness, in contrast, focuses on why, for a given readership and under specific conditions, people are more likely to attribute the text to AI.

Blending these two questions is especially risky in high-stakes judgments of authorship. Even in professional or academic settings, human evaluators can make significant mistakes in how they attribute authorship (Fiedler and Döpke, 2025), and having high confidence doesn't necessarily prevent people from leaning on misleading signals (Hau et al., 2026). At the same time, stylometric differences observed at the level of whole corpora cannot, on their own, justify conclusions about the origin of any single text (O'Sullivan, 2025).

Our results point to an additional, complementary concern: false AI attribution can be systematic within a particular evaluation setup, not just an accidental error. Human-authored texts differed in how often they were judged to be AI-written, and some looked strongly AI-like under certain conditions. As a result, perceived AI-likeness can still matter even when the true source is known to be Human.

This asymmetry is not unique to our data. Recent studies have likewise found that human-authored texts can be misclassified as AI-generated more often than AI texts are mistaken for Human authorship (e.g., Szabó et al., 2026). Viewed from the perspective of perceived AI-likeness, such errors may be informative rather than merely residual: some human texts may systematically evoke an AI interpretation.

There are also real-world reasons to worry about this mismatch. When a text is labeled as AI-authored, it can reduce message and source credibility, the perceived competence of the author, the judged quality of the work, and people's willingness to use or share it (Lermann Henestrosa and

Kimmerle, 2024; Proksch et al., 2024). Perceived AI contribution has likewise been linked to lower message and source credibility (Jia et al., 2024), and readers tend to evaluate narratives and poetry less favorably when they believe the texts were generated by AI (Marklová et al., 2026; Szabó et al., 2026).

Altay and Gilardi (2024) offer a particularly clear example of the downside of false AI attribution: human-written headlines were seen as less accurate and people were less willing to share them when the headlines were labeled as AI-generated. So the cost of an AI attribution doesn't require the text to be actually produced by AI. A human author may be penalized simply because the text is believed to be AI-produced.

This means AI-attribution risk is a different practical issue than detector accuracy. The real question isn't only whether a source classifier gives the correct answer, but whether a text is vulnerable to AI attribution within a particular evaluation context – and what results from that attribution judgment.

The link to trust in W3 is suggestive, but it should not be taken too far. Trust, which loaded on the authorial-experiential end of the scale, was negatively associated with later AI attribution. Since trust was measured before source attribution and perceived source wasn't experimentally manipulated, we can't claim that perceiving AI authorship actually lowered trust. What we can support is narrower: lower trust and stronger AI attribution appeared together in the same perceptual pattern in W3.

A logical next step is not simply to establish whether perceptual impressions precede source attribution: W3 already showed that the perceptual ratings were made before the source judgment and were strongly associated with it. What remains unresolved is what happens after an AI attribution has been formed. An experiment separating actual origin, presented origin, and spontaneously emerging perceived AI-likeness could test whether false-AI attribution itself changes

credibility or trust. By holding the text constant while manipulating the presented source – and including a blind condition – it would be possible to distinguish effects of linguistic form from the downstream effects of source belief.

If it does, then the practical challenge goes beyond detection alone. It would become useful to identify unwanted AI-likeness in cases where source perception affects credibility, trust, or communicative effectiveness. Our findings across different regimes also argue against relying on a single AI-likeness score that works independently of context. Any practical assessment would need to be calibrated for a specific audience and a specific evaluative setting.

4.5. Limitations and future directions

The study has clear limitations. The objective linguistic analyses were based on 28 texts, while the detailed W3 perceptual analysis relied on only 10. The repeated W3 evaluations enabled within-text analyses, but the Authorial-experiential presence versus schematic generality dimension now requires replication using a larger, more varied set of stimuli.

The corpus also varied in terms of topic and communicative context. The Experience, Opinion, and Emotion variables offered only broad control over content differences, so the semantic and stylometric origin signatures should be viewed as characteristics of this particular corpus, not as universal indicators of AI-generated language.

In addition, the participants were professional linguists. Their focus on textual form and their expectations regarding AI-generated writing may differ from those of other reader groups, so the perceptual patterns reported here should not be assumed to generalize unchanged to broader populations.

Lastly, the three regimes differed at the same time in participant cohort, procedure, informational context, and judgment sequence. They were not randomized levels of a single manipulation. As a result, the cross-regime findings document cue reweighting

linked to the different regimes, but they do not show which specific element caused it. A more rigorous causal test would keep the stimuli and participant population constant while manipulating the evaluative frame or source information.

CONCLUSION

Perceived AI-likeness couldn't be traced back in a simple way to the actual source of the text, or to any fixed set of linguistic markers. Human-written and AI-generated texts showed strong differences in the way meaning was distributed and in their stylometric profiles, which means that real authorship did leave a measurable linguistic imprint in this corpus. However, the key dimensions that most clearly distinguished the two source categories didn't, by themselves, fully account for AI attribution.

Individual cues also didn't carry a stable attributional meaning across different evaluation settings. Several semantic and stylometric indicators even flipped their direction: the same measurable feature could be linked to AI attribution in one regime and to Human attribution in another. So, an objective signal distinguishing Human from AI doesn't automatically translate into a fixed perceptual "code" that people use when judging the source.

Cohort 3 suggests a different organizing principle. Here, AI attribution seemed to follow a latent contrast between an author's lived, experiential presence and more schematic, generalized writing. Readers were less likely to label a text as AI-generated when they could picture an author who believes in what they're saying – someone emotionally engaged and grounded in real experience. They were more likely to label it as AI-generated when the writing felt generalized, obvious, and template-like. Importantly, this sense of authorial presence is a perception, not a guarantee about the text's true origin.

As a result, perceived AI-likeness is best understood in this study as a property of the text as it is perceived, not as a

straightforward proxy for machine authorship. The study separates that perceptual property empirically from actual origin and identifies one way in which it was organized in readers' experience. Future work should therefore move beyond linguistic markers of actual AI origin and examine the linguistic signature of perceived AI-likeness itself: which textual features make writing seem AI-generated, how these cues vary across genres, communicative settings, and characteristics of authors and readers, and how they relate to trust, credibility, and later evaluation. An especially important question is what happens when a text is believed to be AI-generated, whether that belief is correct or not.

Statement on the Use of Generative AI:

During the preparation of this manuscript, the authors used Wordvice AI (Wordvice, model/version not specified by the provider) for English-language proofreading, including spelling, grammar, and stylistic editing. The tool was used in relation to the English-language text of the manuscript and did not perform data analysis or make methodological or interpretive decisions. All AI-assisted edits were critically reviewed and, where necessary, revised by the authors. The authors verified the accuracy of factual information, references, and other relevant content and take full responsibility for the final version of the manuscript.

Supplementary materials: Supplementary Tables S1–S12, together with the original Russian W3 text-perception questionnaire and its English translation, are publicly available at: <https://github.com/Litvinova1984/when-human-writing-looks-artificial-supplement/releases/tag/v1.0>

Заявление об использовании генеративного ИИ: При подготовке настоящей рукописи авторы использовали Wordvice AI (Wordvice; модель/версия разработчиком не указана) для языковой вычитки англоязычного текста, включая исправление орфографии, грамматики и стилистики. Инструмент применялся только при редактировании англоязычного

текста рукописи и не использовался для анализа данных, принятия методологических решений или интерпретации результатов. Все изменения, внесённые с использованием системы, были критически проверены и при необходимости отредактированы авторами. Авторы проверили достоверность фактической информации, ссылок и иных элементов содержания и несут полную ответственность за окончательную версию рукописи.

Дополнительные материалы:

Дополнительные таблицы S1–S12, а также оригинальный русский опросник W3 по восприятию текста и его английский перевод доступны для публичного просмотра по адресу: <https://github.com/Litvinova1984/when-human-writing-looks-artificial-supplement/releases/tag/v1.0>

References

- Altay, S. and Gilardi, F. (2024). People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation, *PNAS Nexus*, 3 (10), pgae403. <https://doi.org/10.1093/pnasnexus/pgae403> (In English)
- Chen, R., Xiong, S., He, J. and Ross, G. J. (2026). Stylometric detection of AI-generated texts: evidence from human and machine-written essays, *Digital Scholarship in the Humanities*, fqag064. <https://doi.org/10.1093/llc/fqag064> (In English)
- Demir, N. Y. (2026). Human perceptions of register alignment and functional appropriateness: ChatGPT-generated vs human-authored texts, *Digital Scholarship in the Humanities*, fqag084. <https://doi.org/10.1093/llc/fqag084> (In English)
- Fiedler, A. and Döpke, J. (2025). Do humans identify AI-generated text better than machines? Evidence based on excerpts from German theses. *International Review of Economics Education*, 49, 100321. <https://doi.org/10.1016/j.iree.2025.100321> (In English)
- Hau, M. F., Vulpius, J. and Tønnesen, S. (2026). AI literacy as metacommunicative agility: exploring users' interpretive strategies in detecting and evaluating AI and human

authorship, AI & Society.
<https://doi.org/10.1007/s00146-026-03125-3> (In English)

Jakesch, M., Hancock, J. T. and Naaman, M. (2023). Human heuristics for AI-generated language are flawed, in *Proceedings of the National Academy of Sciences*, 120 (11), e2208839120.

<https://doi.org/10.1073/pnas.2208839120> (In English)

Jia, H., Appelman, A., Wu, M. and Bien-Aimé, S. (2024). News bylines and perceived AI authorship: Effects on source and message credibility. *Computers in Human Behavior: Artificial Humans*, 2 (2), 100093. (In English)

<https://doi.org/10.1016/j.chbah.2024.100093>

Latini, N., Bråten, I. and Strømsø, H. I. (2026). Mind Over Model? Students' Evaluation and Use of ChatGPT-Generated Versus Human-Generated Texts. *Reading Research Quarterly*, 61 (1), e70087. <https://doi.org/10.1002/rq.70087> (In English)

Lermann Henestrosa, A. and Kimmerle, J. (2024). The Effects of Assumed AI vs. Human Authorship on the Perception of a GPT-Generated Text, *Journalism and Media*, 5 (3), 1085–1097. <https://doi.org/10.3390/journalmedia5030069> (In English)

Litvinova, T., Molchanova, V., Dekhnich, O., Seredin, P. (2025). Describing Images: Linguistic Signatures of AI and Humans, in *Proceedings of the 38th Conference of Open Innovations Association FRUCT*, Helsinki, Finland. (In English)

Litvinova, T., Zavarzina, V., Panicheva, P. (2024). Individual Differences in the Most Frequent Content Word Usage as a New Type of Features in the Authorship Profiling Task, in *Proceedings of The International Society for Technology Education and Science*, ICRES 2024, 1901–1924. (In English)

Liu, S., Chu, H., Yuan, S., Sun, Y. and Men, L. R. (2026). The more you think you know: AI experience, literacy, heuristics, and perceived AI authorship. *AI & Society*, 41 (3), 2199–2211. <https://doi.org/10.1007/s00146-025-02643-w> (In English)

Marklová, A., Vinš, O., Vokáčová, M. and Milička, J. (2026). The author is dead, but what if they never lived? A reception experiment on Czech AI- and human-authored poetry. *Digital Scholarship in the Humanities*, fqag067. <https://doi.org/10.1093/llc/fqag067> (In English)

Markowitz, D. M., Hancock, J. T. and Bailenson, J. N. (2024). Linguistic Markers of Inherently False AI Communication and Intentionally False Human Communication: Evidence From Hotel Reviews. *Journal of Language and Social Psychology*, 43 (1), 63–82. <https://doi.org/10.1177/0261927X231200201> (In English)

Mizumoto, A., Yasuda, S. and Tamura, Y. (2024). Identifying ChatGPT-generated texts in EFL students' writing: Through comparative analysis of linguistic fingerprints, *Applied Corpus Linguistics*, 4 (3), 100106. <https://doi.org/10.1016/j.acorp.2024.100106> (In English)

O'Sullivan, J. (2025). Stylometric comparisons of human versus AI-generated creative writing, *Humanities and Social Sciences Communications*, 12, 1708. <https://doi.org/10.1057/s41599-025-05986-3> (In English)

Proksch, S., Schühle, J., Streeb, E., Weymann, F., Luther, T. and Kimmerle, J. (2024). The impact of text topic and assumed human vs. AI authorship on competence and quality assessment, *Frontiers in Artificial Intelligence*, 7, 1412710.

<https://doi.org/10.3389/frai.2024.1412710> (In English)

Szabó, G., Krizsai, F., and Deme, A. (2026). The invisible author: Citizen sociolinguistic perspectives on identifying human and AI-generated narrative texts, *Social Sciences & Humanities Open*, 13, 102646. <https://doi.org/10.1016/j.ssaho.2026.102646> (In English)

Stoltz, D. S., Taylor, M. A. (2022). text2map: R Tools for Text Matrices, *Journal of Open Source Software*, 7 (72), 3741. (In English)

Taylor, M. A., Stoltz, D. S. (2021). Integrating Semantic Directions with Concept Mover's Distance to Measure Binary Concept Engagement, *Journal of Computational Social Science*, 4, 231–242. (In English)

Tudino, G. and Qin, Y. (2024). A corpus-driven comparative analysis of AI in academic discourse: Investigating ChatGPT-generated academic texts in social sciences, *Lingua*, 312, 103838.

<https://doi.org/10.1016/j.lingua.2024.103838> (In English)

Wang, Y., Xing, R., Mansurov, J. et al. (2026). Is Human-Like Text Liked by Humans?

Multilingual Human Detection and Preference Against AI, *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), 14043–14076. <https://doi.org/10.18653/v1/2026.acl-long.639> (In English)

Список литературы

Altay S. People are skeptical of headlines labeled as AI-generated, even if true or human-made, because they assume full AI automation / S. Altay, F. Gilardi // *PNAS Nexus*. 2024. T. 3. № 10. Art. pgae403. DOI: 10.1093/pnasnexus/pgae403

Chen R. Stylometric detection of AI-generated texts: evidence from human and machine-written essays / R. Chen, S. Xiong, J. He, G. J. Ross // *Digital Scholarship in the Humanities*. 2026. Art. fqag064. DOI: 10.1093/llc/fqag064

Demir N. Y. Human perceptions of register alignment and functional appropriateness: ChatGPT-generated vs human-authored texts / N. Y. Demir // *Digital Scholarship in the Humanities*. 2026. Art. fqag084. DOI: 10.1093/llc/fqag084

Fiedler A. Do humans identify AI-generated text better than machines? Evidence based on excerpts from German theses / A. Fiedler, J. Döpke // *International Review of Economics Education*. 2025. T. 49. Art. 100321. DOI: 10.1016/j.iree.2025.100321

Hau M. F. AI literacy as metacommunicative agility: exploring users' interpretive strategies in detecting and evaluating AI and human authorship / M. F. Hau, J. Vulpius, S. Tønnesen // *AI & Society*. 2026. DOI: 10.1007/s00146-026-03125-3

Jakesch M. Human heuristics for AI-generated language are flawed / M. Jakesch, J. T. Hancock, M. Naaman // *Proceedings of the National Academy of Sciences*. 2023. T. 120. № 11. Art. e2208839120. DOI: 10.1073/pnas.2208839120

Jia H. News bylines and perceived AI authorship: Effects on source and message credibility / H. Jia, A. Appelman, M. Wu, S. Bien-Aimé // *Computers in Human Behavior: Artificial Humans*. 2024. T. 2. № 2. Art. 100093. DOI: 10.1016/j.chbah.2024.100093

Latini N. Mind Over Model? Students' Evaluation and Use of ChatGPT-Generated Versus Human-Generated Texts / N. Latini,

I. Bråten, H. I. Strømsø // *Reading Research Quarterly*. 2026. T. 61. № 1. Art. e70087. DOI: 10.1002/rrq.70087

Lermann Henestrosa A. The Effects of Assumed AI vs. Human Authorship on the Perception of a GPT-Generated Text / A. Lermann Henestrosa, J. Kimmerle // *Journalism and Media*. 2024. T. 5. № 3. C. 1085–1097. DOI: 10.3390/journalmedia5030069

Litvinova T. Describing Images: Linguistic Signatures of AI and Humans / T. Litvinova, V. Molchanova, O. Dekhnich, P. Seredin // *Proceedings of the 38th Conference of Open Innovations Association FRUCT*. Helsinki, Finland, 2025.

Litvinova T. Individual Differences in the Most Frequent Content Word Usage as a New Type of Features in the Authorship Profiling Task / T. Litvinova, V. Zavarzina, P. Panicheva // *Proceedings of The International Society for Technology Education and Science, ICRES 2024*. 2024. Pp. 1901–1924.

Liu S. The more you think you know: AI experience, literacy, heuristics, and perceived AI authorship / S. Liu, H. Chu, S. Yuan, Y. Sun, L. R. Men // *AI & Society*. 2026. T. 41. № 3. Pp. 2199–2211. DOI: 10.1007/s00146-025-02643-w

Marklová A. The author is dead, but what if they never lived? A reception experiment on Czech AI- and human-authored poetry / A. Marklová, O. Vinš, M. Vokáčová, J. Milička // *Digital Scholarship in the Humanities*. 2026. Art. fqag067. DOI: 10.1093/llc/fqag067

Markowitz D. M. Linguistic Markers of Inherently False AI Communication and Intentionally False Human Communication: Evidence From Hotel Reviews / D. M. Markowitz, J. T. Hancock, J. N. Bailenson // *Journal of Language and Social Psychology*. 2024. T. 43. № 1. Pp. 63–82. DOI: 10.1177/0261927X231200201

Mizumoto A. Identifying ChatGPT-generated texts in EFL students' writing: Through comparative analysis of linguistic fingerprints / A. Mizumoto, S. Yasuda, Y. Tamura // *Applied Corpus Linguistics*. 2024. T. 4. № 3. Art. 100106. DOI: 10.1016/j.acorp.2024.100106

O'Sullivan J. Stylometric comparisons of human versus AI-generated creative writing / J. O'Sullivan // *Humanities and Social Sciences Communications*. 2025. T. 12. Art. 1708. DOI: 10.1057/s41599-025-05986-3

Proksch S. The impact of text topic and assumed human vs. AI authorship on competence and quality assessment / S. Proksch, J. Schühle, E. Streeb, F. Weymann, T. Luther, J. Kimmerle // *Frontiers in Artificial Intelligence*. 2024. T. 7. Art. 1412710. DOI: 10.3389/frai.2024.1412710

Szabó G. The invisible author: Citizen sociolinguistic perspectives on identifying human and AI-generated narrative texts / G. Szabó, F. Krizsai, A. Deme // *Social Sciences & Humanities Open*. 2026. T. 13. Art. 102646. DOI: 10.1016/j.ssaho.2026.102646

Stoltz D. S. text2map: R Tools for Text Matrices / D. S. Stoltz, M. A. Taylor // *Journal of Open Source Software*. 2022. T. 7. № 72. Art. 3741

Taylor M. A. Integrating Semantic Directions with Concept Mover's Distance to Measure Binary Concept Engagement / M. A. Taylor, D. S. Stoltz // *Journal of Computational Social Science*. 2021. T. 4. Pp. 231–242.

Tudino G. A corpus-driven comparative analysis of AI in academic discourse: Investigating ChatGPT-generated academic texts in social sciences / G. Tudino, Y. Qin // *Lingua*. 2024. T. 312. Art. 103838. DOI: 10.1016/j.lingua.2024.103838

Wang Y. Is Human-Like Text Liked by Humans? Multilingual Human Detection and Preference Against AI / Y. Wang, R. Xing, J. Mansurov [et al.] // *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics*. Vol. 1, Long Papers. 2026. Pp. 14043–14076. DOI: 10.18653/v1/2026.acl-long.639

All authors have read and approved the final manuscript.

Все 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.

Татьяна Александровна Литвинова, доктор филологических наук, профессор кафедры возрастной и социальной психологии, ведущий научный сотрудник, Белгородский государственный университет, Белгород, Россия.

Tatiana A. Litvinova, Doctor of Philological Sciences (D.Sc. in Philology), Professor, Department of Developmental and Social Psychology; Leading Research Fellow, Belgorod State National Research University, Belgorod, Russia.

Ольга Витальевна Дехнич, кандидат филологических наук, доцент, доцент кафедры английской филологии и межкультурной коммуникации Белгородского государственного национального исследовательского университета, Россия.

Olga V. Dekhnich, Candidate of Philological Sciences (Cand. in Philology), Associate Professor of the Department of English Philology and Cross-Cultural Communication, Belgorod State National Research University, Belgorod, Russia.

РАЗДЕЛ I. ТЕОРИЯ ЯЗЫКА SECTION I. THEORY OF LANGUAGE

УДК 81'42

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

Раренко М. Б.¹

Языковые индикаторы персуазивности
в социальной рекламе

¹ Институт научной информации по общественным наукам РАН (ИНИОН РАН)
Нахимовский проспект, д. 51/21, 117418, г. Москва, Россия
E-mail: rarenko@rambler.ru
ORCID: 0000-0003-3601-6832

*Статья поступила 27 июля 2026 г.; принята 15 сентября 2026 г.;
опубликована 30 сентября 2026 г.*

Аннотация: Актуальность. Статья посвящена исследованию персуазивных языковых индикаторов социальной рекламы, одной из речевых практик (жанров) дискурса социальной гармонии, рассматриваемого в противовес агрессивному дискурсу как совокупность речевых практик, способствующих стабилизации общественных отношений и снижению социальной напряженности. Актуальность работы обусловлена возрастающей ролью дискурсивных средств, в частности, в социальной рекламе, направленных на убеждение целевой аудитории в необходимости изменить ее установки, мнения, поведение, в условиях цифровизации и информационной избыточности.

Проблематика. В статье представлены языковые персуазивные индикаторы, используемые в современной социальной рекламе, а также показан механизм их воздействия на целевую аудиторию. Материалом исследования выступает социальная реклама, трактуемая как особая разновидность персуазивного метадискурса, интегрирующего убеждающую интенцию с социально значимым содержанием.

Цель исследования представляется в выявлении, систематизации и характеристике языковых индикаторов персуазивности в социальной рекламе как речевой практике дискурса социальной гармонии, а также определении их роли в воздействии на целевую аудиторию и культурную обусловленность их выбора.

Методы. Для решения поставленных задач в работе использован комплекс методов, обусловленный многоуровневым характером объекта исследования. Выбор методологического инструментария определялся необходимостью сочетания макроанализа дискурсивных практик с микроанализом конкретных языковых средств. Были использованы: метод лингвистического описания, который был базовым для настоящего исследования и применялся на всех этапах работы, метод концептуального моделирования, метод классификации и типологизации, сопоставительный (контрастивный) метод, метод лингвостилистического анализа для выявления и описания конкретных языковых средств реализации персуазивности, а также метод интерпретативного анализа. На заключительном этапе работы применялся метод теоретического синтеза, позволивший объединить результаты, полученные на разных уровнях анализа, в

целостную картину функционирования персуазивных языковых индикаторов в дискурсе социальной гармонии. Синтез осуществляется как в горизонтальном измерении (обобщение данных, полученных разными методами), так и в вертикальном (интеграция теоретических положений и эмпирических наблюдений в общие выводы).

Результаты. Разграничиваются две персуазивные стратегии – убеждение, апеллирующее к рациональной сфере адресата, и уговаривание (суггестивная стратегия), ориентированное на эмоциональное восприятие. В рамках каждой стратегии выделены и охарактеризованы конкретные речевые тактики. Особое внимание уделяется языковым средствам реализации персуазивности – персуазивным индикаторам: инклюзивным конструкциям, обращениям к адресату, метадискурсивным маркерам, индикаторам социальной модальности, средствам позитивной вежливости, а также эмотивной лексике и имплицатурам.

Выводы. В выводах подчёркивается, что система персуазивных индикаторов носит многоуровневый характер. Представлена типология персуазивных языковых индикаторов социальной рекламы. На основе сопоставительного анализа русскоязычной, немецкоязычной и итальянской социальной рекламы делается вывод о культурной обусловленности выбора персуазивных средств: русскоязычная реклама тяготеет к психолингвистическим приёмам, немецкоязычная – к рациональной аргументации, итальянская – к игровым стратегиям. Эффективность социальной рекламы в дискурсе социальной гармонии напрямую зависит от учёта культурных особенностей целевой аудитории.

Ключевые слова: Дискурс социальной гармонии; Персуазивность; Социальная реклама; Стратегии убеждения и уговаривания; Языковые индикаторы; Культурная специфика; Эмотивная лексика

Информация для цитирования: Раренко М. Б. Языковые индикаторы персуазивности в социальной рекламе // Научный результат. Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 3. С. 32–51. DOI: 10.18413/2313-8912-2026-12-3-0-2

UDC 81'42

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

Maria B. Rarenko¹ 

| Linguistic indicators of invasiveness in social advertising

¹ Institute of Scientific Information for Social Sciences of the Russian Academy of Sciences, Moscow, Russian Federation,
E-mail: rarenco@rambler.ru
ORCID: 0000-0003-3601-6832

Received 28 July 2026; accepted 15 September 2026; published 30 September 2026

Abstract: Relevance. The article is devoted to the study of persuasive linguistic indicators in social advertising, one of the speech practices (genres) of the discourse of social harmony, which is considered to be in contrast to aggressive discourse as a set of speech practices aimed at stabilizing social relations and reducing social tension. **The relevance** of the work is explained by the growing role of discursive means, in particular in social advertising, which are aimed at persuading the target audience of the

need to change their attitudes, opinions, and behavior in the context of digitalization and information overload.

Problems. The article presents linguistic persuasive indicators used in modern social advertising and also shows the mechanism of their impact on the target audience. **The research material** is social advertising, which is interpreted as a special type of persuasive metadiscourse integrating a persuasive intention with socially significant content.

The aim of the study is to identify, systematize, and characterize linguistic indicators of persuasion in social advertising as a speech practice within the discourse of social harmony, as well as to determine their role in influencing the target audience and the cultural factors that determine their choice.

Methods. Several methods were used to address the tasks, which were determined by the multi-level nature of the research object. The choice of methodology tools was determined by the need to combine macroanalysis of discursive practices with microanalysis of specific language means. The following methods were used: the method of language description, which served as the foundation for this study and was applied at all stages of the work; the method of conceptual modeling; the method of classification and typologization; the comparative (contrastive) method; the method of language-stylistic analysis, aimed at identifying and describing specific language means of implementing persuasiveness; and the method of interpretive analysis. At the final stage of the work, the method of theoretical synthesis was applied, which made it possible to combine the results obtained at different levels of analysis into a comprehensive description of the functioning of persuasive linguistic indicators in the discourse of social harmony. Synthesis is carried out both in the horizontal dimension (generalization of data obtained using different methods) and in the vertical dimension (integration of theoretical concepts and empirical observations into general conclusions).

Results. Two types of persuasive strategies are distinguished: persuasion, which appeals to the rational sphere of the addressee, and coaxing (a suggestive strategy), which is focused on emotional perception. Within each strategy, speech tactics are identified and characterized. Special attention is paid to the language means of realizing persuasiveness, that is persuasive indicators: inclusive constructions, appealing to the addressee, metadiscursive markers, indicators of social modality, means of positive politeness, as well as emotive vocabulary and implicatures.

Conclusions. The conclusions emphasize that the system of persuasive indicators is multi-level in nature. A typology of persuasive language indicators of social advertising is presented. A comparative analysis of social advertising in Russian, German and Italian reveals that the choice of persuasive means is culturally conditioned: Russian social advertising tends to use psycholinguistic techniques, German social advertising is characterised by rational argumentation, and Italian social advertising uses more playful strategies. In general, the effectiveness of social advertising in the discourse of social harmony directly depends on taking into account the cultural characteristics of the target audience.

Keywords: Discourse of social harmony; Persuasion; Social advertising; Strategies of persuasion and coaxing; Linguistic indicators; cultural specificity; emotive vocabulary

How to cite: Rarenko, M. B. (2026). Linguistic indicators of invasiveness in social advertising, *Research Result. Theoretical and Applied Linguistics*, 12 (3), 32–51. DOI: 10.18413/2313-8912-2026-12-3-0-2

Введение

Данная статья продолжает исследования в области дискурса социальной гармонии, под которым мы понимаем дискурс в противовес агрессивному, в рамках которого появляются новые речевые практики, направленные на стабилизацию отношений в обществе, и посвящена изучению языковых индикаторов персуативности в современной социальной рекламе. Исследование представляется актуальным, поскольку в условиях информационного перенасыщения, цифровизации и социальной нестабильности возрастает значимость дискурсивных практик, направленных на снижение социальной напряженности и гуманизацию общества (Раренко, 2020; Раренко, 2026). Социальная реклама является одной из таких практик, а персуазивность – **неотъемлемой** ее характеристикой.

Научная новизна исследования видится нами в том, что персуазивные индикаторы социальной рекламы впервые анализируются не как набор разрозненных языковых средств, а как целостная многоуровневая система. Такой подход позволяет выявить иерархические связи между отдельными маркерами убеждения и рассматривать их совокупное воздействие на адресата. Как справедливо отмечают исследователи, язык не только служит средством коммуникации, но и является инструментом передачи информации, убеждения и оказания влияния на аудиторию (Annisa Zahrani, Muhammad Hamzah, Marpaung, Simarmata, 2025; Ertimur, Chen, 2026; Han, Ni, Amon, Necaie, Han, Lee, 2025; Lathiva Istavida, Joko Wibowo, 2026; Wasif Hussain, Panggabean, 2024).

Среди языковых индикаторов дискурса социальной гармонии нам хотелось бы особое внимание обратить на персуазивные ввиду их высокой

значимости.

Персуазивность (от лат. *persuasio* – убеждение) мы рассматриваем как системное свойство высказывания, направленное на изменение ментальных установок реципиента с целью побуждения его к совершению посткоммуникативного действия в интересах адресанта, а персуазивную коммуникацию, вслед за А.В. Голодновым, определяем как «особый тип ментально-речевого взаимодействия индивидов, реализующий попытку преимущественно вербального воздействия адресанта на установку реципиента» (Голоднов, 2010: 2).

В качестве материала исследования мы предлагаем обратиться к одной из практик дискурса социальной гармонии – социальной рекламе на том основании, что рекламный дискурс в целом рассматривается как одна из ключевых разновидностей персуазивной коммуникации, а социальная реклама в этом контексте представляет собой особый тип риторического метадискурса, интегрирующего убеждающую интенцию с социально значимым содержанием (более подробно см.: (Isakova, 2007; Герасименко, Николаева, Темирева, 2016; Игошина, 2008; Малюга, Попова, 2018; Тянь, 2021 и др.)).

Представим принципы работы социальной рекламы в виде наглядной таблицы, в основании которой лежит классический риторический треугольник Аристотеля, но с перераспределением акцентов.

Если же представить социальную рекламу как особый вид дискурса, используя формулу, то она будет выглядеть так:

SR = (Ethos × Pathos) + (Media Code) – (Profit),

где «минус Profit» – это главный дискурсивный маркер, превращающий рекламу в социальную, а не коммерческую.

Таблица 1. Принципы работы социальной рекламы
Table 1. Principles of social advertising

СОЦИАЛЬНЫЙ АДРЕСАТ (Общество/Гражданин)		
ЭТОС	ПАФОС	ЛОГОС
Доверие	Эмпатия	Статистика
СОЦИАЛЬНЫЙ ПРОБЛЕМАТОР (Заказчик: НКО/Государство)		
МЕДИА-КАНАЛ		

Методы и материал

В работе использован комплекс методов, обусловленный многоуровневым характером объекта исследования; выбор методологического инструментария определялся необходимостью сочетания макроанализа дискурсивных практик с микроанализом конкретных языковых средств. Дискурсивный анализ (метод лингвистического описания дискурса) стал базовым для настоящего исследования и применялся на всех этапах работы. Дискурсивный анализ позволил идентифицировать дискурс социальной гармонии как особый тип дискурса, противопоставленный агрессивному дискурсу; описать системные свойства данного дискурса (целеполагание, субъектно-объектные отношения, хронотоп, модальность); выявить функциональную специфику социальной рекламы как одной из ключевых речевых практик дискурса социальной гармонии. Метод структурно-семантического описания с элементами коммуникативно-прагматического был задействован при последовательном описании структурных и содержательных характеристик исследуемого материала на коммуникативно-прагматическом уровне.

Для наглядной репрезентации теоретических положений применялся метод концептуального моделирования, позволивший представить сложные дискурсивные феномены в операционализированном,

схематизированном виде. Систематизация эмпирического материала осуществлялась посредством метода классификации и типологизации.

Для выявления культурно-специфических особенностей персуазивных языковых индикаторов применяется сопоставительный метод. Материалом для сопоставления послужили 37 текстов социальной рекламы за период с 2012 по 2024 гг. из открытых источников и личных наблюдений автора на русском (19 текстов), немецком (17 текстов) и итальянском языках (11 текстов) (преимущественно антитабачной тематики, что обеспечивает сопоставимость по предметно-референтной области). Сравнение проводилось по следующим параметрам: 1) доминантный тип персуазивного воздействия (психолингвистический / когнитивно-логический / социально-игровой); 2) целевая аудитория и характер адресации; 3) эмоциональная тональность и модус; 4) приоритетные языковые индикаторы и способы их реализации; 5) риторическая функция и формула персуазивности; 6) тип конструируемой социальной гармонии. Контрастивный анализ позволил сделать вывод о культурной обусловленности выбора персуазивных средств и о необходимости учета лингвокультурного фактора при создании эффективной социальной рекламы.

На микроуровне текста применялся метод лингвостилистического анализа,

направленный на выявление и описание конкретных языковых средств реализации персуазивности. Данный метод позволил нам зафиксировать конкретные языковые маркеры персуазивности и описать их функциональную нагрузку в дискурсе социальной гармонии на примере социальной рекламы.

Для раскрытия глубинных смыслов, не эксплицированных в поверхностной структуре текста, нами использовался метод интерпретативного анализа, который позволил перейти от формального описания языковых средств к пониманию их стратегической функции в дискурсе социальной гармонии.

На заключительном этапе работы мы применили метод теоретического синтеза, позволивший объединить результаты, полученные нами на разных уровнях анализа, в целостную картину функционирования персуазивных языковых индикаторов в дискурсе социальной гармонии. Синтез осуществлялся как в горизонтальном измерении (при обобщении данных, полученных разными методами), так и в вертикальном (для интеграции теоретических положений и перевода эмпирических наблюдений в общие выводы).

Результаты

1. Структура персуазивного воздействия

В рамках рекламной коммуникации, на наш взгляд, заслуживают внимания две стратегии, различающиеся по способу достижения персуазивной цели и по характеру восприятия информации адресатом: стратегия убеждения и стратегия уговаривания. Стратегия убеждения представляет собой совокупность речевых действий, воздействующих на рациональную сферу адресата – его разум и логику. Целью данной стратегии становится побуждение

к действию в открытой форме, когда сам факт воздействия осознается реципиентом, однако содержание сообщения подвергается критическому осмыслению. Языковая реализация стратегии убеждения предполагает использование аргументов, обращенных к интеллектуальной сфере: логических доводов, фактов, статистических данных, ссылок на авторитетные источники. В рекламном дискурсе факт воздействия представляется «открытым» (более подробно см.: (Ху, 2023)), однако адресант часто стремится скрыть прямой императив «Купи!» или, в случае социальной рекламы, «Сделай!», «Поступи так!». Стратегия уговаривания (или суггестивная стратегия) ориентирована на эмоциональную и чувственную сферу адресата. В отличие от убеждения, уговаривание реализуется в скрытой форме, когда реципиенту трудно критически оценить поступающую информацию, поскольку она воспринимается на эмоциональном уровне. Суггестивная коммуникация, как подчеркивает Г. А. Мирошниченко, «реализуется включением в содержание рекламного сообщения эмотивного нарратива, возбуждающего желание реципиента», и директивных речевых актов в форме установки, «направляющей действия потребителя в интересах бренда» (Мирошниченко, 2019: 266). Важно подчеркнуть, что в социальной рекламе, в отличие от коммерческой, стратегия убеждения и стратегия уговаривания не противопоставляются как взаимоисключающие, а скорее взаимодополняют друг друга. Социальная реклама, реализуемая в рамках дискурса социальной гармонии, требует как рационального осознания социальной проблемы (стратегия убеждения), так и эмоционального сопереживания (стратегия уговаривания), что создает комплексное персуазивное воздействие.

Таблица 2. Критериальное противопоставление стратегий
Table 2. Critical juxtaposition of strategies

Критерий	Стратегия УБЕЖДЕНИЯ	Стратегия УГОВАРИВАНИЯ
Объект воздействия	Рациональная сфера (разум, сознание)	Эмоциональная сфера (чувства, подсознание)
Форма предъявления	Открытая (адресат осознает воздействие)	Скрытая (адресат не осознает механизм воздействия)
Способ обработки информации	Логический анализ, критическая оценка	Суггестия, внушение, эмпатия
Тип аргументов	Факты, статистика, ссылки, логические цепочки	Метафоры, нарративы, образы, эмотивная лексика
Иллокутивная цель	Доказать → Убедить → Побудить	Вовлечь → Сопережить → Побудить
Императивность	Часто завуалирована, но логически выводима	Скрыта за эмоциональным нарративом
Пример языковых маркеров	«120 000 человек умирают ежегодно» (цифровое представление)	«Твоя мечта может рассыпаться в прах» (образное представление)

2. Речевые тактики в рамках персуазивных стратегий

В рамках стратегии убеждения в рекламном дискурсе выделим следующие тактики: 1) тактику апелляции к удовлетворению базовых потребностей реципиента; 2) тактику апелляции к опыту; 3) тактику апелляции к авторитету. В самом общем виде, тактика апелляции к удовлетворению базовых потребностей реципиента представляет собой обращение к фундаментальным потребностям человека (безопасность, здоровье, комфорт) и реализуется через рациональные аргументы о полезности и практической значимости предлагаемого действия. В социальной рекламе эта тактика проявляется в призывах к

здоровому образу жизни, безопасности на дорогах, экологической и социальной ответственности. Тактика апелляции к опыту в основе своей имеет ссылку на личный или коллективный опыт, который подтверждает эффективность предлагаемого поведенческого паттерна. В дискурсе социальной гармонии, в частности, социальной рекламе, эта тактика может реализовываться через обращение к исторической памяти, культурным традициям и общепринятым нормам. Тактика апелляции к авторитету состоит в использовании ссылок на мнение экспертов, общественных деятелей или признанных организаций. В социальной рекламе данной тактикой часто сопровождаются

кампании, проводимые при поддержке государственных институтов или общественных организаций.

В рамках стратегии уговаривания выделим тактики, апеллирующие преимущественно к эмоциональной сфере: 1) тактика апелляции к социальным и духовным потребностям; 2) тактика апелляции к личности человека; 3) тактика «надевания маски». Так, тактика апелляции к социальным и духовным потребностям реализуется посредством обращения к потребности в признании, самореализации, принадлежности к группе, что особенно значимо для дискурса социальной гармонии. В социальной рекламе эта тактика реализуется через инклюзивные конструкции и позитивную вежливость. Тактика апелляции к личности человека представляет собой персонифицированное обращение, акцентирующее уникальность и ценность каждого адресата. Данная тактика способствует формированию личной ответственности и эмпатического

восприятия социальных проблем. Тактика «надевания маски» создает у адресата ощущения, что он добровольно и самостоятельно принимает решение, тогда как на самом деле его поведение направляется скрытыми коммуникативными установками. Эта тактика тесно связана с механизмами манипуляции, когда коммуникатор не отдавал реципиенту никаких команд, ни о чем его не просил, но тот почему-то сам захотел что-то сделать для него. Тактика обращения к эмоциям реципиента ориентирована на актуализацию чувственного восприятия социальной проблемы. В зависимости от коммуникативной задачи используются различные эмоциональные модусы – от страха (как предупреждения о негативных последствиях) до надежды и доверия (как основы для позитивных изменений).

Выделим риторические топосы (общие места) социальной рекламы. Ниже в Таблице 3 представлена схема риторической аргументации:

Таблица 3. Риторические топосы социальной рекламы

Table 3. Rhetorical toposes of social advertising

1. Топос угрозы (Страх)	<i>«Курение убивает»</i>	Апелляция к инстинкту самосохранения
2. Топос стыда (Социальное давление)	<i>«Ты не такой, как все?»</i>	Апелляция к конформизму
3. Топос жертвы (Эмпатия):	<i>«Помоги больному ребенку»</i>	Апелляция к состраданию
4. Топос героя (Идентификация)	<i>«Стань донором – спаси жизнь»</i>	Апелляция к высшему "Я"
5. Топос общности судьбы (Экзистенция)	<i>«Мы – одна страна/планета»</i>	Апелляция к коллективному бессознательному

3. Языковые средства реализации персуазивных стратегий в социальной рекламе

По аналогии с анализом рекламных текстов, проведенным А.В. Голодновым, который позволил выявить типичные инвариантные имиджи коммерческой рекламы, в основе которых лежат различные типы оценки, можно выделить имиджи, регулярно встречающиеся в социальной рекламе: 1) имидж "хорошее

дело"; 2) имидж "полезное действие"; 3) имидж "благородное дело". Так, имидж "хорошее дело" основан на общей оценке, выражаемой общеоценочными предикатами (*хорошо, замечательно, правильно, верно*). Персуазивный потенциал общей оценки определяется ее «семантической опустошенностью», позволяющей реципиенту самостоятельно заполнить смысловой вакуум различными признаками, что делает сообщение

релевантным для широкой аудитории. Имидж «полезное действие» акцентирует утилитарную и телеологическую оценку и оценивается как полезное в определенной проблемной ситуации реципиента, а результат описывается как наиболее эффективный; проявляется в демонстрации социальной и личной выгоды от просоциального поведения. Имидж «благородное дело» связывает действие реципиента с удовлетворением его эмоциональных потребностей реципиента и находит выражение в позитивных эмоциональных модусах, демонстрирующих радость, удовлетворение и благополучие как результат социально ответственного и социально одобряемого поведения. Выделение "имиджей" при анализе персуазивных стратегий в социальной рекламе способствует систематизации разрозненных креативных решений и выявлению глубинных механизмов воздействия, выходящих за пределы ситуативного контекста конкретной кампании. Так, для всех текстов социальной рекламы, направленных на поддержание любого из имиджей, характерна абсолютизация положительной оценки, т.е. расположение определенных характеристик в области максимума оценочной шкалы. К языковым средствам выражения абсолютно положительной оценки относятся: 1) лексические единицы, обозначающие высшую степень оценки (*абсолютно, безусловно, совершенно, максимально*); 2) интенсификаторы (*исключительно, особенно, крайне*); композиты, экспрессивно выражающие гиперболизированную оценку; 3) лексические единицы с семантикой

генерализации и универсальности (*всегда, везде, каждый*).

4. Гибридизация жанров как персуазивная стратегия

Для речевых практик дискурса социальной гармонии, к которым относится социальная реклама, особый интерес представляет использование гибридных жанров (более подробно см.: (Чернявская, Горшкова, 2025)) как инструмента персуазивной коммуникации. Гибридизация жанра становится коммуникативной стратегией, позволяющей осуществить воздействие на адресата через совмещение признаков двух и более текстовых жанров. Жанровая полифония в социальной рекламе создает эффект «свой-чужой» трансформации: форма привычного, повседневного общения служит для передачи социально значимого содержания, что снижает психологическую защиту адресата и повышает восприимчивость к убеждающему воздействию. Текст социальной рекламы может быть оформлен как личный разговор, инструкция, рецепт, энциклопедическая статья или интернет-чат и пр., в зависимости от целевой аудитории, что способствует снижению коммуникативных барьеров и повышению доверия к сообщению. Данная стратегия особенно эффективна сегодня, когда традиционные жанры официальной коммуникации часто ассоциируются с манипуляцией и не вызывают доверия.

Социальная реклама располагается на пересечении трех осей, где она занимает срединное (гибридное) положение, что и нашло отражение в Таблице 4.

Таблица 4. Жанровая полифония социальной рекламы
Table 4. Genre polyphony of social advertising

Ось	Коммерческий полюс	Социальная реклама	Пропагандистский полюс
Цель	Покупка товара	Изменение поведения/установки	Формирование лояльности к власти
Адресат	Целевая аудитория (сегмент)	Человек как нравственный субъект	Электорат/Масса
Хронотоп	Здесь и сейчас (сиюминутная выгода)	Будущее (отсроченный результат)	Вечное настоящее (миф)
Модальность	Императив + соблазн	Деонтическая (долженствование)	Абсолютная истина

5. Типология персуазивных языковых индикаторов дискурса социальной гармонии (на примере социальной рекламы)

Ключевую роль в достижении социальной гармонии играют языковые средства, направленные на установление диалогических отношений¹ между адресантом и адресатом, и, как следствие, выступающие в качестве персуазивных языковых индикаторов. К числу таких индикаторов относятся: 1) инклюзивные конструкции, 2) разные типы обращений к адресату, 3) метадискурсивные маркеры, 4) индикаторы социальной модальности, 4) индикаторы позитивной вежливости, 5) эмотивная лексика и имплицатуры².

¹ Под "диалогическими отношениями" в социальной рекламе мы понимаем способ построения коммуникации, при котором адресант (создатель рекламы) и адресат (зритель/слушатель) взаимодействуют не как «говорящий» и «пассивный слушатель», а как равноправные участники смыслового обмена. Вместо того чтобы "приказывать", информировать или манипулировать (монолог), социальная реклама выстраивается так, чтобы у реципиента возникла внутренняя рефлексия, ответная мысль или эмоциональный отклик, то есть чтобы он стал соавтором процесса осмысления проблемы.

² В рамках данного исследования статистических замеров не проводилось, поскольку на данном этапе задача исследования состояла в выявлении и типологизации персуазивных языковых индикаторов в текстах социальной рекламы.

Использование местоимений первого лица множественного числа (мы, наш, с нами) создает эффект общности, солидарности и совместной ответственности за решение социальной проблемы. *Инклюзивный язык*, как показывают исследования, укрепляет социальные нормы и формирует чувство принадлежности к группе. В социальной рекламе такой подход находит широкое применение: «Мы – за безопасные дороги»³, «Наш выбор – здоровое будущее», «Вместе мы сможем больше». Использование личных местоимений второго лица (ты, вы, твой), форм повелительного наклонения и риторических вопросов создает эффект прямого диалога, сокращает дистанцию между коммуникантами и повышает вовлеченность реципиента. Исследования показывают, что социальная реклама, построенная как прямое обращение к каждому члену общества, обладает более высоким персуазивным потенциалом. К метадискурсивным маркерам, т.е. языковым средствам, организующим взаимодействие в процессе коммуникации и способствующим установлению гармоничных отношений, мы, в первую очередь, относим вводные конструкции ("как известно", "должно быть",

³ Здесь и далее примеры приводятся из открытых источников и личных наблюдений автора статьи.

"очевидно", "общепризнано", "не подлежит сомнению"), уточняющие и пояснительные обороты, а также средства, выражающие отношение говорящего к содержанию высказывания. В социальной рекламе метадискурсивные маркеры выполняют функцию «мостика» между адресантом и адресатом, создавая атмосферу доверия и взаимоуважения. Социальная модальность в рекламном социальном дискурсе реализуется через систему языковых средств, с помощью которых адресант выражает свое отношение к социальной норме и формирует у адресата соответствующее восприятие действительности.

В рамках дискурса социальной гармонии выделяются следующие средства выражения социальной модальности⁴:

- 1) модальные глаголы и конструкции,
- 2) обобщающие высказывания,
- 3) этические обращения.

Употребление модальных глаголов долженствования ("должен", "обязан", "необходимо", "следует") позволяет адресанту транслировать социальные нормы и ожидания, формировать у адресата представление о должном, социально одобряемом поведении. Одновременно использование модальных глаголов возможности ("можешь", "способен", "в твоих силах") акцентирует личную ответственность и субъектность адресата, что способствует принятию социально значимых решений. Отмечается, что модальные глаголы в социальной рекламе выступают ключевым средством убеждения, так как они эксплицируют отношение говорящего к описываемой ситуации и формируют соответствующее отношение у слушающего. Использование обобщающих высказываний ("каждый из нас", "все мы знаем", "как принято в

обществе") актуализирует социальный консенсус и нормативность. Социальное доказательство, как лингвистический механизм, опирается на принцип «так поступают все», что побуждает адресата следовать общепринятым моделям поведения. Этические обращения, т.е. языковые средства, апеллирующие к моральным ценностям и нравственным принципам, в социальной рекламе реализуются через лексику с семантикой справедливости, добра, ответственности, что способствует формированию у адресата внутренней мотивации для изменения поведения.

В основе конструирования социальной гармонии лежат стратегии вежливости, интерпретируемые как надежные лингвистические индикаторы социальных отношений. Позитивная вежливость направлена на подчеркивание солидарности и общности с адресатом, демонстрацию уважения и признания его ценности. К числу таких индикаторов относятся: 1) комплиментарные конструкции, 2) разные типы выражения заботы, 3) хеджи (средства смягчения).

Комплиментарные конструкции, т.е. высказывания, выражающие одобрение и положительную оценку адресата, его действий или выбора, в социальной рекламе представляют собой фразы подобные следующим: «Ты делаешь правильный выбор», «Каждый из нас способен на добрый поступок», «Здоровый образ жизни – это твой выбор». Выражение заботы осуществляется посредством разных языковых средств, демонстрирующих внимание адресанта к потребностям и интересам адресата. Фразы «Позаботьтесь о своих близких», «Ваше здоровье – в ваших руках», «Не пренебрегайте своим здоровьем и здоровьем своих близких» не только информируют, но и создают эмоциональную связь между адресантом и адресатом. Языковые средства, снижающие категоричность высказывания и минимизирующие коммуникативные риски, известные как хеджи, в социальной рекламе

⁴ Под "социальной модальностью" мы понимаем разновидность прагматической модальности, которая фиксирует отношение между участниками коммуникации с точки зрения их социального статуса, ролей, дистанции и нормативных ожиданий общества.

реализуются через вводные конструкции, наречия и глаголы со значением предположения. Они позволяют смягчить иллокутивную силу высказывания и сохранить «лицо» собеседника, что особенно важно при обсуждении социально чувствительных тем.

Особую группу персуазивных языковых индикаторов представляют эмотивная лексика и импликатуры. Эмотивная лексика является мощным

персуазивным инструментом в социальной рекламе. Исследования показывают, что использование слов с ярко выраженной эмоциональной окраской способствует более глубокому воздействию на ментальную сферу реципиента (см. например: (Голоднов, 2001; Голоднов, 2002)). Ниже в Таблице 5 представлена детальная схема соответствия индикаторов риторическим интенциям.

Таблица 5. Индикаторы и соответствующие им риторические интенции в социальной рекламе

Table 5. Indicators and their corresponding rhetorical intentions in social advertising

Языковой индикатор	Риторическая функция	Схематическая формула действия	Пример (Русскоязычный контекст)
1. Инклюзивные конструкции	Тактика солидаризации (Этос)	Слияние субъекта и объекта действия. (Я + Ты) = Мы	«Нам не все равно», «Мы за чистый город». (Исключается "я начальник, ты дурак").
2. Обращения к адресату	Акт признания (Фатика)	Персонализация безличной проблемы. Абстракция переходит в Личность	«Друг!», «Водитель!», «Мама, позвони мне». Устанавливает контакт до аргументации.
3. Метадискурсивные маркеры	Управление вниманием (Логос + Этикет)	Вербализация процесса мышления. Смотри, следовательно, Подумай, следовательно, Сделай вывод	«Согласитесь», «Очевидно, что...», «Представьте себе». Это приглашение к со-размышлению.
4. Социальная модальность	Деонтическая трансформация (Логос)	Снижение категоричности императива. Должен, следовательно, Принято / Важно / Полезно	Вместо «Не пей» используется «Здоровым быть выгодно». Давление нормы заменяется ценностным ориентиром.
5. Позитивная вежливость	Стратегия "сохранения лица" (Пафос + Этос)	Одобрение адресата до того, как ему указали на недостатки. Ты ценен, следовательно, Сделай лучше	«Ты – наша надежда», «Умные люди сортируют мусор». (Лесть как презумпция положительности).
6. Эмотивная лексика + Импликатуры	Эвокация и Догадка (Пафос + Глубинный смысл)	Прямое чувство в совокупности со скрытым выводом. Текст (Shock) приводит к Подтексту (Выводу)	Эмотив: «Больно». Импликатура: (если не пристегнешься, будет так же больно твоим детям). Вывод адресат делает самостоятельно.

6. Эмоциональные модусы дискурса социальной гармонии

В дискурсе социальной гармонии выделяются два основных (и противоположных друг другу) эмоциональных модуса: позитивная эмоциональная тональность и негативная эмоциональная тональность. Позитивная эмоциональная тональность создается посредством употребления лексики, передающей оптимизм, надежду, доверие, радость. так, исследователи отмечают, что в социальной рекламе итальянского медиадискурса, например, преобладает позитивный эмоциональный фон, что социальная реклама ориентирована преимущественно на молодежную аудиторию. Позитивная тональность социальной гармонии способствует формированию у адресата ощущения безопасности и стабильности, что является важным условием социальной гармонии. Негативная эмоциональная тональность реализуется посредством привлечения лексики, актуализирующей

чувство страха, тревогу, озабоченность. Этот модус используется как предупреждение реципиентов о негативных последствиях социально опасного поведения. Однако в дискурсе социальной гармонии негативная эмоциональность, как правило, компенсируется позитивными конструкциями, предлагающими пути решения проблемы.

Эмотивные импликатуры – скрытые эмоциональные смыслы, передаваемые через контекст и культурные конвенции. В юридическом и социально-ориентированном дискурсе, как отмечают исследователи, эмотивные импликатуры транслируют чувство разделяемых ценностей, справедливости и солидарности. Социальная гармония в дискурсе удерживается не статикой, а динамическим равновесием двух модусов. Схематически это выглядит как колебание маятника, где негатив – это сигнал тревоги, а позитив – это ресурс для восстановления.

Схема 1. Базовая бинарная структура (Маятник гармонии)

Scheme 1. Basic binary structure (Harmony pendulum)

ДИСКУРС СОЦИАЛЬНОЙ ГАРМОНИИ	
НЕГАТИВНАЯ ТОНАЛЬНОСТЬ (Модус Проблемы)	ПОЗИТИВНАЯ ТОНАЛЬНОСТЬ (Модус Решения)
Страх / Тревога / Озабоченность	Оптимизм / Надежда / Доверие
"Больно" / "Опасно" / "Красиво"	"Выживем" / "Стыдно" / "Спасибо"
КОМПЕНСАТОРНЫЙ БАЛАНС (Негатив привлекает внимание -> Позитив дает энергию)	
ЭМОТИВНЫЕ ИМПЛИКАТУРЫ (Скрытый смысл: Солидарность, Справедливость)	

7. Культурно-специфические особенности языковых индикаторов дискурса социальной гармонии (на примере социальной рекламы)

Сопоставительные исследования социальной рекламы в различных

лингвокультурах выявляют как универсальные, так и культурно-специфические черты персуазивных языковых индикаторов⁵.

⁵ В этой статье мы хотели бы отметить эту закономерность и обратить внимание

Русскоязычная социальная реклама тяготеет к психолингвистическим приемам воздействия на ментальную сферу целевой аудитории. Это проявляется в акцентировании эмоциональной составляющей сообщения, использовании ярких метафор и экспрессивных средств, обращении к культурным архетипам и коллективным представлениям. Так, в социальной рекламе «Родите ли? Курение вызывает бесплодие», построенной на **каламбуре** ("родители" / "родите ли?") и предостерегающей будущих родителей от последствий курения, в основе воздействия на целевую аудиторию лежит игра слов, которая привлекает внимание к проблеме, воздействуя на эмоциональную сферу реципиентов.

Немецкоязычная социальная реклама, напротив, ориентирована преимущественно на языковые приемы воздействия, что выражается в большей логической структурированности текста, использовании рациональных аргументов и строгой лексической организации. Такое различие обусловлено культурной средой и спецификой восприятия убеждающих сообщений в различных лингвосообществах. В рекламе, направленной против курения "Rauchen verursacht Krebs" (Курение вызывает рак), превентивной по своей характеру, нейтральной по тону, содержится фактическая информация. Отмечается, что "для немецкоязычной социальной рекламы больше характерно рациональное убеждение, включающее опору на факты, статистику или логические аргументы. "Jedes Jahr sterben 120.000 Menschen an den Folgen des Rauchens" (Каждый год 120 000 человек умирает от последствий курения), передаваемая в нейтральном стиле" (Зубкова, Рябова, Шатохина, 2025: 230). "Можно сделать вывод, что русскоязычная социальная реклама формально более нацелена на индивидуального потребителя,

но в массовом масштабе, что приближает её к немецкоязычной. Однако активизация поведенческих реакций в русскоязычной и немецкоязычной аудитории осуществляется по-разному. Русскоязычная социальная реклама апеллирует к эмоциональной составляющей воздействия, а немецкоязычная социальная реклама опосредована больше когнитивной, в соответствии с чем и предопределяется выбор соответствующих языковых средств выражения персуазивности" (Зубкова, Рябова, Шатохина, 2025: 231). "Русскоязычная реклама тяготеет к психолингвистическим, а немецкоязычная – к лингвистическим приёмам воздействия на ментальную сферу целевой аудитории" (Зубкова, Рябова, Шатохина, 2025: 233).

Итальянская социальная реклама, адресованная, в первую очередь, молодежи, характеризуется использованием позитивного эмоционального фона, грубоватого юмора, игровых стратегий и акцентированием национально-культурных особенностей (более подробно см.: (Бибикова, Курилов, 2025)). Такой подход позволяет преодолеть психологическую защиту аудитории и создать более доверительную атмосферу общения. Направленная на борьбу с курением итальянская реклама для повышения эмоциональности высказывания использует форму прямого обращения к реципиентам, разговорную лексику, восклицательные конструкции, отрицательный императив: "E voi che siete così giovani, non fatevi fregare. Non mandate in fumo i vostri sogni! Il fumo uccide! Difendetevi!" (А вы, что так молоды, не дайте вас одурачить! Не дайте рассыпаться в прах вашим мечтам! Курение убивает! Защищайтесь!)⁶.

В приложении в Таблице 6 представлены особенности языковых индикаторов дискурса социальной гармонии на примере антитабачной

исследователей на перспективность дальнейших изысканий в данном направлении.

⁶ Пример заимствован из работы: (Бибикова, Курилов, 2025: 12).

русскоязычной, немецкоязычной и италийскоязычной социальной рекламы.

Выводы

В заключение отметим, что персуазивные языковые индикаторы дискурса социальной гармонии в социальной рекламе представляют собой многоуровневую систему, включающую: 1) индикаторы межличностного взаимодействия (инклюзивные конструкции, обращения к адресату, метадискурсивные маркеры), которые организуют общение, создавая атмосферу диалога и взаимопонимания, сокращая социальную дистанцию между коммуникантами; 2) индикаторы социальной модальности (модальные глаголы, обобщения и социальное доказательство, этические обращения), которые транслируют социальные нормы и ожидания, формируя у адресата представление о должном поведении и социально приемлемых практиках; 3) индикаторы позитивной вежливости (комплиментарные конструкции, выражения заботы, хеджи), которые позволяют смягчить категоричность сообщения и сохранить «лицо» адресата, демонстрируя уважение и признание его ценности); 4) эмотивную лексику и имплицатуры, которые, воздействуя на чувственную сферу реципиента, способствуют формированию глубинных установок и ценностных ориентаций.

Существенной характеристикой социальной рекламы как речевой практики дискурса социальной гармонии является ее культурная обусловленность. Сопоставительные исследования показывают, что в различных лингвокультурах акцент делается на разных типах персуазивных индикаторов. Так, русскоязычная реклама тяготеет к психолингвистическим приемам, немецкоязычная – к рационально-логическим аргументам, итальянская – к игровым стратегиям и юмору. Это свидетельствует о том, что эффективность

социальной рекламы напрямую зависит от учета культурных особенностей целевой аудитории.

Заявление об использовании генеративного ИИ:

При подготовке настоящей рукописи автор использовал сервис <https://chat.deepseek.com/>. Инструмент применялся для построения схем и таблиц на материале проведенных нами исследований. Результаты, полученные с использованием системы, были критически проверены и при необходимости отредактированы автором. Автор проверил достоверность фактических данных, ссылок и иных элементов содержания и несет полную ответственность за окончательную версию рукописи.

Statement on the Use of Generative AI: During the preparation of this manuscript, the author used the service <https://chat.deepseek.com/>. The tool was used to build tables and schemes based on the research we conducted. All AI-assisted outputs were critically reviewed and, where necessary, revised by the author. The author verified the accuracy of factual information, references, and other relevant content and take full responsibility for the final version of the manuscript.

Список литературы

- Бибикина А. М., Курилов А. М. Лингвопрагматические аспекты медицинской социальной рекламы в итальянском медиадискурсе // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2025. Вып. 11 (905). С. 9–15.
- Герасименко В.Н., Николаева Я.Д., Темирева К.А. Социальная реклама в интернете как способ конструирования социальной реальности // Социально-гуманитарные проблемы образования и профессиональной самореализации (Социальный инженер-2016). Сборник материалов Всероссийского научного форума молодых исследователей. 2016. С. 33–35.
- Голоднов А. В. Персуазивная коммуникация: стратегии и тактики воздействия (на материале современной

немецкоязычной рекламы. Санкт-Петербург: Астерион, 2010. 244 с.

Голоднов А.В. Лексические маркеры персуазивности в рекламных слоганах // Разноуровневые характеристики лексических единиц. Ч. 1.: Лексика и фразеология. Терминология. Смоленск: Изд-во СГПУ, 2001. С. 80–86.

Голоднов А.В. Лингвостилистическая интерпретация коммуникативных стратегий персуазивности в текстах коммерческой рекламы // Studia Linguistica - П. Проблемы когнитивной семантики. СПб.: Изд-во РШУ, 2002. С. 114–119.

Зубкова Я. В., Рябова М. Э., Шатохина О. А. Языковые средства выражения персуазивности в русскоязычной и немецкоязычной социальной рекламе // Российский социально-гуманитарный журнал. 2025. №2. URL: www.evestnik-mgou.ru

Игошина Т.С. Социальная реклама как форма социальной коммуникации и объект проектирования // Альманах современной науки и образования. 2008. № 4–2. С. 94–97.

Малюга Е. Н., Попова К. В. Лингвопрагматика речевых стратегий в социальной рекламе // Вестник Московского государственного областного университета (электронный журнал). 2018. № 4. URL: www.evestnik-mgou.ru (дата обращения: 15.09.2026).

Мирошниченко Г.А. Внушение в рекламе: особенности коммуникативного механизма // Вестник Российского университета дружбы народов. Серия: Литературоведение. Журналистика. 2019. Т. 24. № 2. С. 264–272.

Раренко М. Б. Концептуальное поле дискурса социальной гармонии // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2026. Вып. 5 (911). С. 76–83.

Раренко М.Б. Лингвокультурные параметры вербализации социальной гармонии // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2020. Вып. 12 (841). С. 147–157.

Тянь К. Социальная реклама как способ социальной коммуникации в новых медиа // Успехи гуманитарных наук. 2021. № 7. С. 227–230.

Ху Ц. Реализация стратегии убеждения и стратегии уговаривания в российской телевизионной рекламе (на материале видеороликов о продуктах питания) // Мир науки. Социология, филология, культурология. 2023. Т. 14. № 4. URL: <https://sfk-mn.ru/PDF/28FLSK423.pdf> (дата обращения: 15.09.2026).

Чернявская В.Е., Горшкова Н.Э. Гибридизация жанра в персуазивной коммуникации // Слово.ру: балтийский акцент. 2025. Т. 16, №1. С. 74–89.

Annisa Zahrani, Muhammad Hamzah, Marpaung F. S. and Simarmata A. A. A Speech Act Analysis of Persuasion and Evaluation in Marques Brownlee's "M4 MacBook Air Review: Too Easy!" // Journal of Applied Linguistics (JoAL). 2026. Vol. 5 No. 2, January. Pp. 48–59. <https://doi.org/10.52622/joal.v5i2.571>

Ertimur B., Chen S. Toward Destigmatization of Mental Health: The Institutional Role of Advertising // Journal of Advertising. 2026. Vol. 55 (1). Pp. 81–97. <https://doi.org/10.1080/00913367.2025.2593665>

Han J., Ni Y., Amon M. J., Necaie A., Han M., Lee M. Multimodal Physiological Responses Predict Advertisement Effectiveness // Journal of Advertising, 2025. Vol. 54 (5). Pp. 693–712. <https://doi.org/10.1080/00913367.2025.2524185>

Isakova I.A. Social Advertising As Cultural Phenomenon // ЧЕЛОВЕК. ОБЩЕСТВО. КУЛЬТУРА. СОЦИАЛИЗАЦИЯ. Материалы всероссийской студенческой конференции, посвященной 450-летию добровольного вхождения Башкирии в состав России. 2007. С. 48–50.

Lathiva Istavida, Joko Wibowo. Persuasive Illocutionary Speech Acts in Canva Advertisements on the Ptalform: A Linguistic Approach // Journal of Applied Linguistics. 2026. Vol. 5 No. 2. Pp. 106–114. <https://doi.org/10.52622/joal.v5i2.607>

Wasif Hussain, Panggabean H. Language and Power Dynamics in Automobile Advertising: A Multimodal Discourse Analysis of Word Choice and Semiotics // Journal of Applied Linguistics. 2024. Vol. 4 No. 1. Pp. 71–81. <https://doi.org/10.52622/joal.v4i1.242>

References

Bibikova, A. M., Kurilov, A. M. (2025). Lingvopragmaticheskie aspekty meditsinskoy sotsialnoy reklamy v italyanskom mediadiskurse

[Linguopragmatic aspects of medical social advertising in the Italian media discourse], *Bulletin of the Moscow State Linguistic University. Humanities*, Issue 11 (905), 9–15. (In Russian)

Gerasimenko, V. N., Nikolaeva, Ya. D. and Temireva, K. A. (2016). Sotsialnaya reklama v internete kak sposob konstruirovaniya sotsialnoy realnosti [Social advertising on the Internet as a way of constructing social reality], *Socio-humanitarian problems of education and professional self-realization (Social Engineer-2016)*, Collection of materials of the All-Russian Scientific Forum of Young Researchers, 33–35. (In Russian)

Golodnov, A. V. (2010). Persuazivnaya kommunikatsiya: strategii i taktiki vozdeystviya (na materiale sovremennoy nemetskoyazychnoy reklamy [Persuasive communication: strategies and tactics of influence (based on the material of modern German-language advertising], Asterion, St. Petersburg, 244 p. (In Russian)

Golodnov, A. V. (2001). Leksicheskie markery persuazivnosti v reklamnykh sloganakh [Lexical markers of persistence in advertising slogans], *Multilevel characteristics of lexical units. Part 1.: Vocabulary and phraseology. Terminology*, Smolensk: SSPU Publishing House, 80–86. (In Russian)

Golodnov, A. V. (2002). Lingvostilisticheskaya interpretatsiya kommunikativnykh strategiy persuazivnosti v tekstakh kommercheskoy reklamy [Linguostylistic interpretation of communicative strategies of persistence in commercial advertising texts], *Studia Linguistica - P. Problems of cognitive semantics*, Publishing House of the Russian Academy of Medical Sciences, St. Petersburg, 114–119. (In Russian)

Zubkova, Ya. V., Ryabova, M. E. and Shatokhina, O. A. (2025). Yazykovye sredstva vyrazheniya persuazivnosti v russkoyazychnoy i nemeckoyazychnoy sotsialnoy reklame [Linguistic means of expressing persistence in Russian and German-language social advertising], *Russian Social and Humanitarian Journal*, No. 2 [Online], available at: www.evestnik-mgou.ru (Accessed 15.09.2026) (In Russian)

Igoshina, T. S. (2008). Sotsialnaya reklama kak forma sotsialnoy kommunikatsii i ob'ekt proektirovaniya [Social advertising as a form of social communication and an object of design], *Almanac of Modern Science and Education*, No. 4–2, 94–97. (In Russian)

Malyuga, E. N., Popova, K. V. (2018). Lingvopragmatika rechevykh strategiy v sotsialnoy reklame [Linguopragmatics of speech strategies in social advertising], *Bulletin of Moscow State Regional University (electronic journal)*. № 4 [Online], available at: www.evestnik-mgou.ru (Accessed 15.09.2026) (In Russian)

Miroshnichenko, G. A. (2019). Vnushenie v reklame: osobennosti kommunikativnogo mekhanizma [Suggestion in advertising: features of the communicative mechanism], *Bulletin of the Peoples Friendship University of Russia. Series: Literary Studies. Journalism*, Vol. 24. No. 2, 264–272. (In Russian)

Rarenko, M. B. (2026). Концептуальное поле дискурса социальной гармонии [The conceptual field of the discourse of social harmony], *Bulletin of Moscow State Linguistic University. The Humanities*, Issue 5 (911), 76–83. (In Russian)

Rarenko, M. B. (2020). Lingvokulturnye parametry verbalizatsii sotsialnoy garmonii [Linguocultural parameters of verbalization of social harmony], *Bulletin of Moscow State Linguistic University. Humanities*, Issue 12 (841), 147–157. (In Russian)

Tyan, K. (2021). Sotsialnaya reklama kak sposob sotsialnoy kommunikatsii v novykh media [Social advertising as a way of social communication in new media], *Successes of the Humanities*, No. 7, 227–230. (In Russian)

Hu, C. (2023). Realizatsiya strategii ubezhdeniya i strategii ugovarivaniya v rossiyskoy televizionnoy reklame (na materiale videorolikov o produktakh pitaniya) [Implementation of urging and persuasion strategies in Russian television advertising (based on videos about food)], *Mir Nauki. Sociology, philology, cultural studies*, Vol. 14. No. 4 [Online], available at: <https://sflk-mn.ru/PDF/28FLSK423.pdf> (Accessed 15.09.2026) (In Russian)

Chernyavskaya, V. E., Gorshkova, N. E. (2025). Gibridizatsiya zhanra v persuazivnoy kommunikatsii [Hybridization of genre in the aggressive communication], *Word.ru: Baltic Accent*, Vol. 16, No. 1, 74–89. (In Russian)

Annisa Zahrani, Muhammad Hamzah, Marpaung, F. S. and Simarmata, A. A. (2025). A Speech Act Analysis of Persuasion and Evaluation in Marques Brownlee's "M4 MacBook Air Review: Too Easy!", *Journal of Applied Linguistics*, 5 (2), 48–59. <https://doi.org/10.52622/joal.v5i2.571> (In English)

Ertimur, B. and Chen, S. (2026). Toward Destigmatization of Mental Health: The Institutional Role of Advertising, *Journal of Advertising*, 55 (1), 81–97. <https://doi.org/10.1080/00913367.2025.2593665> (In English)

Han, J., Ni, Y., Amon, M. J., Necaise, A., Han, M. and Lee, M. (2025). Multimodal Physiological Responses Predict Advertisement Effectiveness, *Journal of Advertising*, 54 (5), 693–712. <https://doi.org/10.1080/00913367.2025.2524185> (In English)

Isakova, I. A. (2007). Social Advertising As Cultural Phenomenon, *MAN. SOCIETY. CULTURE. SOCIALIZATION*, Materials of the All-Russian student conference dedicated to the

450th anniversary of Bashkirias voluntary incorporation into Russia, 2007, 48–50. (In Russian)

Lathiva Istavida and Joko Wibowo. (2026). Persuasive Illocutionary Speech Acts in Canva Advertisements on the Ptalform: A Linguistic Approach, *Journal of Applied Linguistics*, 5 (2), 106–114. <https://doi.org/10.52622/joal.v5i2.607> (In English)

Wasif Hussain and Panggabean, H. (2024). Language and Power Dynamics in Automobile Advertising: A Multimodal Discourse Analysis of Word Choice and Semiotics, *Journal of Applied Linguistics*, 4 (1), 71–81. <https://doi.org/10.52622/joal.v4i1.242> (In English)

ПРИЛОЖЕНИЕ

Таблица 6. Культурно-специфические особенности языковых индикаторов дискурса социальной гармонии (на примере антитабачной социальной рекламы)

Table 6. Culturally specific features of language indicators of the discourse of social harmony (using anti-tobacco social advertising as an example)

Параметр сравнения	Русскоязычная социальная реклама	Немецкоязычная социальная реклама	Итальянская социальная реклама
Доминантный тип воздействия	Психолингвистический (воздействие на эмоционально-образную сферу)	Лингвистический / Когнитивный (воздействие на рационально-логическую сферу)	Социально-игровой / Эмотивный (воздействие через доверие и идентификацию с молодежной средой)
Целевая аудитория (вектор)	Массовый индивидуальный адресат («каждый» через коллективные архетипы)	Абстрактный рациональный субъект («гражданин», статистическая единица)	Молодежная аудитория («свой», ровесник)
Модус тональности	Гибридный (эмоциональный шок + скрытая надежда)	Нейтрально-информационный (фактологическая строгость, отсутствие пафоса)	Ярко-позитивный / Фамильярный (юмор, ирония, энергия)
Приоритетные персуазивные индикаторы	Яркие метафоры Экспрессивные средства Обращение к архетипам Каламбур (языковая игра)	Логическая структурированность Факты и статистика Строгая лексическая организация Нейтральная модальность	Прямые обращения (ты-адресация) Разговорная лексика Восклицательные конструкции Отрицательный императив + инклюзивы
Способ подачи аргумента	Ассоциативно-образный Смысл выводится через игру слов и эмоциональную реакцию	Прямой / Эксплицитный Смысл выводится через цифры и причинно-следственные связи	Диалогово-игровой Смысл выводится через доверие и предупреждение «свой своим»

Отношение к адресату	Апелляция к ментальной сфере (ценностям, страхам, архетипам)	Апелляция к когнитивной сфере (разуму, осознанию рисков)	Апелляция к эмоционально-волевой сфере (желанию жить ярко, не быть обманутым)
Языковая реализация (примеры)	1. «Родите ли? Курение вызывает бесплодие» Прием: Каламбур (омонимия), провокационный вопрос, удар по репродуктивному инстинкту. 2. «Мамочка, не переживай, я не узнаю, что моя астма из-за тебя» Прием: Обращение к самому дорогому для ребенка человеку («мамочка»), эмоциональный шантаж, ирония, эксплуатация чувства вины и родительского инстинкта. Ребенок берет на себя роль «прощающего», что усиливает удар по сознанию курящей матери. 3. «СОРВИСЬ С КРЮЧКА! Курильщик зависит от 6570 сигарет в год» Прием: Восклицание, императив, метафора наркотической зависимости («крючок»), пугающая статистика как приговор. Сочетание прямого призыва к действию и неопровержимого факта. 4. «Купи себе рак легких» Прием: Метафора, оксюморон (покупка болезни как товара), черный юмор. Разрушение привычных моделей восприятия.	1. «Jedes Jahr sterben 120.000 Menschen an den Folgen des Rauchens» Прием: Статистика, нейтральная лексика, безличная конструкция, факт как приговор. 2. «Mighty Tasty» (Camel) Прием: Английский слоган в немецком контексте, дерзкий вызов критикам. Кампания сопровождалась сценами, нарочито нарушающими табу (пепел в еде), с предупреждением «Viewer discretion advised». Контр-позиционирование: отказ от защитной риторики в пользу агрессивной уверенности. 3. «Viewer discretion advised» (в контексте антитабачных и табачных кампаний) Прием: Заимствованный из кинематографической культуры дисклеймер, создающий интригу и провокацию. В контексте рекламы сигарет Camel использовался для создания ореола «запретного» контента. 4. «Rauchen lässt Ihre Haut altern» Прием: Персонализированное обращение («Ihre»), конкретизация абстрактного вреда через видимый и лично значимый эффект (старение кожи), научная авторитетность без запугивания. Удар по эстетическому самолюбию вместо страха смерти. 5. «Rauchen kann zur Erblindung führen» Прием: Статистика, нейтральная лексика. Расширение привычного спектра угроз (лёгкие, сердце) на неожиданную сферу - зрение.	1. «E voi che siete così giovani, non fatevi fregare... Il fumo uccide! Difendeteви!» Прием: Ты-обращение, инклюзивный императив, просторечие, восклицание, апелляция к мечте («sogni»). 2. «Invece di accendere la sigaretta accendi il cervello» Прием: Антитеза, игра слов, разговорный императив. Противопоставление автоматического действия («зажечь сигарету») и сознательного усилия («включить мозг») создает каламбур с оттенком насмешки. 3. «Chi non fuma sta una favola!» Прием: Каламбур, разговорная игра с фразеологизмом. Буквально «тот, кто не курит, – это сказка», с отсылкой к итальянскому выражению «стать как в сказке» (т.е. чувствовать себя прекрасно). 4. «A Natale fatti un bel regalo. Smetti di fumare» Прием: Апелляция к празднику и культуре подарков, инклюзивный

		Формулировка «kann... führen» создаёт дистанцию и одновременно неотвратимость. Апелляция к одному из базовых страхов - потерять зрение.	императив, позитивная мотивация вместо запугивания. Отказ от курения подается как акт заботы о себе. 5. «Il fumo uccide: difenditi!» Прием: Двухчастная конструкция: констатация факта («курение убивает») плюс императив-призыв («защити себя»). Военная метафора («difenditi») превращает курильщика из виновника в жертву, нуждающуюся в защите.
Риторическая функция	Фатическая + Эмотивная (Удержать внимание через удивление / шок)	Информативная + Логическая (Убедить через неопровержимость данных)	Конативная + Поэтическая (Побудить к действию через экспрессию и близость)
Формула персуазивности (схематично)	<i>Шок (Игра слов) → Осознание угрозы → Эмоциональный отклик</i>	<i>Цифра → Логическое умозаключение → Рациональный выбор</i>	<i>Комплимент (Ты молод/крут) → Запрет (Не дай себя обмануть) → Энергия протеста против смерти</i>
Тип социальной гармонии (итоговый)	Гармония через общий страх и преодоление его вместе.	Гармония через общественное согласие с фактами (порядок и безопасность).	Гармония через общий позитивный драйв («мы не хотим терять мечты»).

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.

Мария Борисовна Раренко, кандидат филологических наук, ведущий научный

сотрудник отдела языкознания, заведующий отделом языкознания, Институт научной информации по общественным наукам РАН (ИНИОН РАН), Москва, Российская Федерация.

Maria B. Rarenko, Candidate of Philology, Leading Researcher, Head of the Department of Linguistics, Institute of Scientific Information for Social Sciences, Russian Academy of Sciences, Moscow, Russian Federation.

РАЗДЕЛ II. ПРИКЛАДНАЯ ЛИНГВИСТИКА SECTION II. APPLIED LINGUISTICS

UDC 81'42

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

Vladimir V. Bochkarev¹ 
Anton S. Afanasev² 
Rinat A. Bakirov³ 

Spearman's rank distance as a tool for stylometric attribution: a comparative analysis on heterogeneous corpora

¹ Kazan (Volga Region) Federal University,
18 Kremlyovskaya St., Kazan, 420008, Russia
E-mail: vbochkarev@mail.ru
ORCID: 0000-0001-8792-1491

² Kazan (Volga Region) Federal University,
18 Kremlyovskaya St., Kazan, 420008, Russia
E-mail: a.s.afanasyev@mail.ru
ORCID: 0000-0003-0571-833X

³ Kazan (Volga Region) Federal University,
18 Kremlyovskaya St., Kazan, 420008, Russia
Institute of Russian Literature (Pushkin House)
4 Makarov Emb., Saint Petersburg, 199034, Russia
E-mail: r1nt@ya.ru
ORCID: 0000-0001-6693-1076

*Статья поступила 17 июня 2026 г.; принята 15 сентября 2026 г.;
опубликована 30 сентября 2026 г.*

Abstract: This paper presents an empirical comparison of a metric based on Spearman's rank correlation coefficient with traditional and relatively new distance measures (Burrows's Delta, Cosine Delta, Labbé distance, Soergel distance, and Jensen-Shannon divergence, the latter receiving particular attention) used in stylometric text attribution. The relevance of the study lies in the need to search for and identify a stylometric tool (a distance measure and its "settings") that is robust to within-author variability and feature space sparsity. The research problem stems from methodological uncertainty in choosing an optimal distance measure for stylometric attribution. Despite the development of distance-based approaches, the effectiveness of metrics critically depends on corpus structure, the length of the analyzed fragments, and feature space parameters. In the absence of a single reliable metric for attributing complex text corpora, this study for the first time substantiates the use of Spearman's rank distance as a valid tool for stylometric analysis. The empirical basis of the study comprises four text corpora heterogeneous in terms of time, genre, and language (49 authors in total, 899 literary texts). The study included unified text preprocessing (tokenization, lowercasing, wordform-level analysis) and systematic parameter tuning for the six metrics. For each combination (author, metric, fragment length), the optimal configuration of parameters (number of frequent words; absence/presence of standardization) was determined,

providing the maximum true positive rate in attribution using the k-NN method, $k=1$. The results show that Spearman's rank distance demonstrates the greatest robustness and universality, maintaining high accuracy under genre heterogeneity, variability in text length, and the presence of stylistic anomalies. It is established that Spearman's rank distance systematically outperforms frequency-based and entropy-based metrics in robustness under conditions of genre heterogeneity and can serve as a reliable tool for stylometric attribution.

Keywords: Quantitative literary studies; Corpus linguistics; Stylometry; Distance measures; Spearman's rank distance; Jensen-Shannon divergence

Acknowledgments: The study was supported by the Russian Science Foundation grant No. 24-18-00570, <https://rscf.ru/project/24-18-00570/>.

How to cite: Bochkarev, V. V., Afanasev, A. S., Bakirov, R. A. (2026). Spearman's rank distance as a tool for stylometric attribution: a comparative analysis on heterogeneous corpora, *Research Result. Theoretical and Applied Linguistics*, 12 (3), 52–79. DOI: 10.18413/2313-8912-2026-12-3-0-3

UDC 81'42

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

Бочкарев В. В.¹
Афанасьев А. С.²
Бакиров Р. А.³

Ранговое расстояние Спирмена как инструмент
стилометрической атрибуции: сравнительный анализ
на разнородных корпусах

¹ Казанский (Приволжский) федеральный университет
ул. Кремлевская, 18, Казань, 420008, Россия
E-mail: vbochkarev@mail.ru
ORCID: 0000-0001-8792-1491

² Казанский (Приволжский) федеральный университет
ул. Кремлевская, 18, Казань, 420008, Россия
E-mail: a.s.afanasyev@mail.ru
ORCID: 0000-0003-0571-833X

³ Казанский (Приволжский) федеральный университет
ул. Кремлевская, 18, Казань, 420008, Россия
Институт русской литературы (Пушкинский Дом)
наб. Макарова, 4, Санкт-Петербург, 199034, Россия
E-mail: r1nt@ya.ru
ORCID: 0000-0001-6693-1076

Received 17 June 2026; accepted 15 September 2026; published 30 September 2026

Аннотация: В статье представлено эмпирическое сравнение метрики, основанной на расчете коэффициента ранговой корреляции Спирмена, с традиционными и относительно новыми мерами расстояния (Дельта Берроуза, косинусное расстояние, расстояние Лаббе, расстояние Зергеля, дивергенция Йенсена-Шеннона, которой уделено особое внимание), используемых при стилометрической атрибуции текста. Актуальность исследования заключается в необходимости поиска и обнаружения стилометрического инструмента (меры

расстояния и ее «настроек»), устойчивого к внутриавторской вариативности и разреженности признакового пространства. Проблема исследования обусловлена методологической неопределенностью в выборе оптимальной меры расстояния для стилометрической атрибуции. Несмотря на развитие distance-based подходов эффективность метрик критически зависит от структуры корпуса, длины анализируемых фрагментов и параметров признакового пространства. В условиях отсутствия единой надежной метрики для атрибуции сложных текстовых массивов данное исследование впервые обосновывает использование расстояния Спирмена как валидного инструмента стилометрического анализа. Эмпирической базой исследования стали четыре разнородных по времени, жанрам и языку написания корпуса текстов (всего 49 авторов, 899 художественных текстов). Исследование включало унифицированную предобработку текстов (токенизация, приведение к нижнему регистру, анализ на уровне словоформ) и систематический перебор параметров для шести метрик. Для каждой комбинации (автор, метрика, длина фрагмента) определялась оптимальная конфигурация параметров (количество частотных слов; отсутствие/наличие стандартизации), обеспечивающая максимальную долю истинно положительных результатов при атрибуции методом k-NN, k=1. Результаты показали, что расстояние Спирмена демонстрирует наибольшую робастность и универсальность, сохраняя высокую точность при жанровой неоднородности, вариативности объема текстов и наличии стилистических аномалий. Установлено, что ранговое расстояние Спирмена систематически превосходит частотные и энтропийные метрики по робастности в условиях жанровой гетерогенности и может служить надежным инструментом стилометрической атрибуции.

Ключевые слова: Квантитативное литературоведение; Корпусная лингвистика; Стилметрия; Меры расстояния; Ранговое расстояние Спирмена; Дивергенция Йенсена-Шеннона

Благодарности: Исследование выполнено при поддержке гранта Российского научного фонда № 24-18-00570, <https://rscf.ru/project/24-18-00570/>.

Информация для цитирования: Бочкарев В. В., Афанасьев А. С., Бакиров Р. А. Ранговое расстояние Спирмена как инструмент стилометрической атрибуции: сравнительный анализ на разнородных корпусах // Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 3. С. 52–79. DOI: 10.18413/2313-8912-2026-12-3-0-3

Introduction. The systematic use of stylometric text analysis dates back to the second half of the 19th century. Without delving into a detailed historical excursus (referring the reader on this matter to works by Grzybek (Grzybek, 2014), Martynenko (Martynenko, 2014; Martynenko, 2015), Kovalev (Kovalev, 2024)), we move directly to the beginning of the 21st century and recall that a turning point in the development of stylometric analysis was the publication of J. Burrows' now-classic work (Burrows, 2002), in which he proposed the Delta method for determining stylistic proximity between texts

and their attribution. Burrows' Delta (otherwise known as standardized Manhattan distance) is still one of the most popular distance measures in stylometric analyses (Skorinkin and Orekhov, 2023; Dviniatin and Kovalev, 2024; O'Sullivan, 2025; Mikros et al 2026), but is periodically (and, in our view, justifiably) subjected to constructive criticism and attempts at improvement (see, e.g., Hoover, 2004; Argamon, 2008; Smith and Aldridge, 2011; Rybicki and Eder, 2011), which stimulates the search for and development of alternative distance measures

that could overcome certain limitations of the classical Delta.

Currently, alongside standardized Manhattan distance, various metrics from the group of distance-based approaches are widely used in stylometry: unstandardized Manhattan distance, Euclidean distance, Cosine Delta, Canberra distance, Eder's Delta, Min-Max, and some others. Given this diversity of metrics, a natural question arises: which (or perhaps which one) of them demonstrates the greatest reliability and stability when solving stylometric problems?

Several research works compare distance measures used in stylometric studies.

A research team led by F. Jannidis compared 14 different distance measures (Jannidis et al., 2015) and concluded that Cosine Delta significantly surpasses the others in terms of accuracy and stability. A later article explains the advantage of cosine similarity over other distance measures: the process of normalising vector length mitigates the impact of deviation from the norm (Evert et al., 2017).

M. Kocher and J. Savoy (Kocher and Savoy, 2019), using French and English literature, compared Tanimoto, Matusita, Clark, and Cosine Delta for the authorship attribution task. It was found that the best results were achieved using 4-6-character letter n-grams as features and the Tanimoto, Matusita, and Clark distance metrics, which systematically outperformed Cosine Delta in accuracy.

D. Stanikunas, J. Mandravickaite, and T. Krilavicius (Stanikunas et al., 2017), analyzing Lithuanian texts (transcripts of parliamentary debates), compared 15 distance and similarity metrics, including Burrows's Delta, Eder's Delta, Eder's Simple Delta, Binomial Index, and Cosine Delta. The researchers concluded that Eder's Simple Delta demonstrates consistently high performance and can be used for general stylometric analysis.

O. Alieva (Alieva, 2024) conducts a comparative analysis of some of the metrics listed above and other distance measures that

have also been sporadically used for stylometric investigations and have proven their worth: Labbé distance (Manhattan distance between full vectors of relative frequencies), Tanimoto distance (which, in our interpretation, O. Alieva uses to describe Soergel distance), and Jeffrey's divergence (symmetrized Kullback-Leibler divergence). According to Alieva's experiment, "when classifying ancient Greek texts, the best results are achieved using cosine similarity on standardized frequency values, Tanimoto distance on relative frequency values, and Labbé distance on absolute frequency values" (Alieva, 2024: 23).

As we can see, a single universally successful metric does not exist: all studies confirm that the effectiveness of metrics strongly depends on language, genre, and frequency normalization method, and no universal solution has been proposed.

Conceptually significant for us is the study by D. Pronin and E. Kazartsev (Pronin and Kazartsev, 2026), who, using literary corpora in English, German, French, and Russian, compared classical metrics of the Delta family (Burrows's Delta, Cosine Delta, Eder's Delta) with two new measures – Rank-Turbulence Delta and Jensen-Shannon Delta. These metrics are based on the interpretation of standardized frequency vectors as probabilistic or rank distributions, making it possible to apply divergence measures from information theory and complex systems analysis. The article proves that Rank-Turbulence Delta demonstrates accuracy comparable to Cosine Delta, and Jensen-Shannon Delta systematically matches or surpasses Burrows's Delta in accuracy. One also cannot ignore an earlier work (Grieve, 2007), where the author considers rank methods as a robust alternative to Euclidean and Manhattan distances when working with sparse data.

Although the potential of Jensen-Shannon Delta (in our study, we will call the Jensen-Shannon Delta the Jensen-Shannon Divergence – JSD) has begun to be studied in stylometry (Pronin and Kazartsev, 2026), the

distance based on Spearman's rank correlation coefficient has not yet appeared in comparative experiments. Therefore, the aim of this study is to conduct an empirical comparison of distance metrics based on Spearman's rank correlation and JSD with metrics traditional for stylometric research (Manhattan distance, Cosine Delta, Labbé distance, and Soergel distance) under conditions of genre heterogeneity, variability in text size, and changes in feature space size.

Materials and Methods.

Corpus Matritals.

For our experiment, we used four text corpora.

The "Greek" corpus includes 57 texts by 17¹ ancient Greek authors and was borrowed from the study by O. Alieva (Alieva, 2024). The composition of the corpus and source data are available in an open repository:

<https://github.com/locusclassicus/compareDis t>.

The "Sholokhov" and "Rowling" corpora were formed to reproduce two classical stylometric cases: confirming M. Sholokhov as the author of the epic novel "And Quiet Flows the Don" and identifying J. Rowling under the pseudonym R. Galbraith.

The "Sholokhov" corpus is based on the corpus prepared by N. Velikanova and B. Orekhov (Velikanova and Orekhov, 2019: 79). We retained 11 authors but increased the number of works to 33 (M. Bulgakov: "The White Guard", "The Master and Margarita", "Heart of a Dog"; Vs. Ivanov: "Armoured Train 14-69", "Blue Sands", stories; F. Kryukov: "Group B", "Swell", "To the Source of Healings", "Mother", "Squall"; L. Leonov: "Badgers", "The Thief", "The End of a Small Man"; N. Ostrovsky: "How the Steel Was Tempered", "Born of the Storm"; A. Platonov: "The Foundation Pit",

"Chevengur", "The Juvenile Sea"; V. Seversky: "Don on Crutches"; A. Serafimovich: "The Iron Flood", stories; A. Fadeev: "The Rout", "The Spill", "The Birth of the Amgun Regiment"; D. Furmanov: "Red Landing", "Mutiny", "Chapaev"; M. Sholokhov: "Don Stories", "They Fought for Their Country", "Virgin Soil Upturned", "The Fate of a Man", "And Quiet Flows the Don")².

The 11 authors included in the "Rowling" corpus were taken by us from the article by B. Orekhov (Orekhov, 2024: 2-4), but, as with "Sholokhov", the number of works was increased to enhance sample representativeness. Our corpus contains 54 literary texts: G. M. Burnet: "The Disappearance of Adèle Bedeau", "His Bloody Project", "Case Study", "Benbecula"; H. Coben: "Deal Breaker", "Drop Shot", "Fade Away", "One False Move", "Gone for Good", "No Second Chance", "Tell No One"; T. French: "In the Woods", "The Likeness", "Faithful Place", "Broken Harbour", "The Secret Place", "The Trespasser", "The Searcher", "The Hunter", "The Witch Elm"; R. Galbraith: "The Cuckoo's Calling", "The Silkworm", "Career of Evil", "Lethal White", "Troubled Blood", "The Ink Black Heart", "The Running Grave", "The Hallmarked Man"; C. Lewis: "The Lion, the Witch and the Wardrobe", "Prince Caspian: The Return to Narnia", "The Voyage of the Dawn Treader", "The Silver Chair", "The Horse and His Boy", "The Magician's Nephew", "The Last Battle"; J. Rowling: "Harry Potter and the Philosopher's Stone", "Harry Potter and the Chamber of Secrets", "Harry Potter and the Prisoner of Azkaban", "Harry Potter and the Goblet of Fire", "Harry Potter and the Order of the Phoenix", "Harry Potter and the Half-Blood Prince", "Harry Potter and the Deathly Hallows", "The Casual Vacancy"; B. Sanderson: "Mistborn: The Final Empire", "The Well of Ascension", "The Hero of

¹ The graphs represent 16 authors of the corpus: Antisthenes consistently showed NaN in our calculations.

² The curve for V. Seversky is not presented in the graphs because his oeuvre is represented by only one novel, "Don on Crutches".

Agas”; J. Stroud: “The Amulet of Samarkand”; “The Golem’s Eye”; S. Turton: “The Seven Deaths of Evelyn Hardcastle”, “The Devil and the Dark Water”, “The Last Murder at the End of the World”; J. Tolkien: “The Lord of the Rings”; F. Warrington: “A Blackbird in Silver”, “The Dark Arts of Blood”).

Regarding these three corpora, it was important for us to see how Spearman’s rank distance (hereinafter referred to as Spearman distance) and JSD perform on existing data.

The “Science fiction and rustic prose” (“SF&RP”) corpus is a hyper-corpus combining 755 works by 10 authors of Soviet science fiction and rustic prose: A. Belyaev – 46, I. Efremov – 27, A. and B. Strugatsky – 67, K. Bulychev – 210, A. Kazantsev – 90; V. Astafyev – 97, V. Belov – 17, B. Mozhaev – 38, V. Rasputin – 25, V. Shukshin – 138. This corpus was used to assess the robustness of metrics to very large, cross-genre collections. There is an imbalance in the representation of texts by different authors (from 17 to 210 texts). This factor was considered when interpreting the results and is discussed in the limitations section.

The total vocabulary of the “Greek” corpus was 82955 words, “Rowling” – 60294, “Sholokhov” – 184686, and “SF&RP” – 454049.

Methodology.

Having analyzed previous work, we selected metrics that showed the best results in accuracy and stability. These are Labbé and Soergel distances, Cosine Delta, Burrows’ Delta (Manhattan distance with preliminary standardization of frequencies), and Jensen-Shannon divergence (hereinafter referred to as JSD), to which we will pay special attention in our study. We will compare these with Spearman distance.

These 6 metrics can be divided into three fundamentally different groups based on the type of information used. The first and most representative group consists of traditionally used metrics – Manhattan distance, Cosine Delta and Labbé distance. These approaches work directly with relative

frequencies of wordforms, measuring linear or angular discrepancies between frequency vectors. The second group comprises the probabilistic and information-theoretic measures of the Soergel distance and the JSD. They interpret frequency vectors as probability distributions and assess their divergence through entropy characteristics, making them sensitive to the shape of the distribution, not just the magnitude of frequencies. Finally, the third, separate group is represented by Spearman’s rank metric. Its fundamental difference lies in operating not with the frequencies themselves, but with word ranks – ordinal numbers in a sorted list. This fundamentally changes the stability of the metric: the rank approach is insensitive to extreme values and outliers and does not assume normality of distribution.

We will not discuss the traditional metrics that have been repeatedly used for stylometric analysis in detail, but will refer to existing research. The application of standardized Manhattan distance is described in detail by J. Burrows (Burrows, 2002) and critically evaluated in (Argamon, 2008). Cosine Delta (among other metrics) was the subject of analytical review in (Evert et al., 2017), and regarding Labbé distance, we refer to the work of the author himself (Labbé and Labbé, 2006) (the results of its application can be found in (Bortolato, 2016)). The Labbé distance is the Manhattan distance between the full vectors of relative frequencies (without employing standardization) of all words occurring in the compared texts.

We will discuss Spearman distance, JSD, and Soergel distance in more detail.

In statistics, Spearman’s rank correlation coefficient (ρ) is used to assess the strength of association between two variables (Spearman, 1904). Unlike the Pearson correlation coefficient, it is not the values of two quantities x_i and y_i that are compared directly, but their ranks ri and ri , that is, the numbers in a list sorted in ascending (or descending) order. The coefficient is then calculated by the formula:

$$\rho = 1 - \frac{6}{n(n-1)^2} \sum_{i=1}^n (r_i^x - r_i^y)^2$$

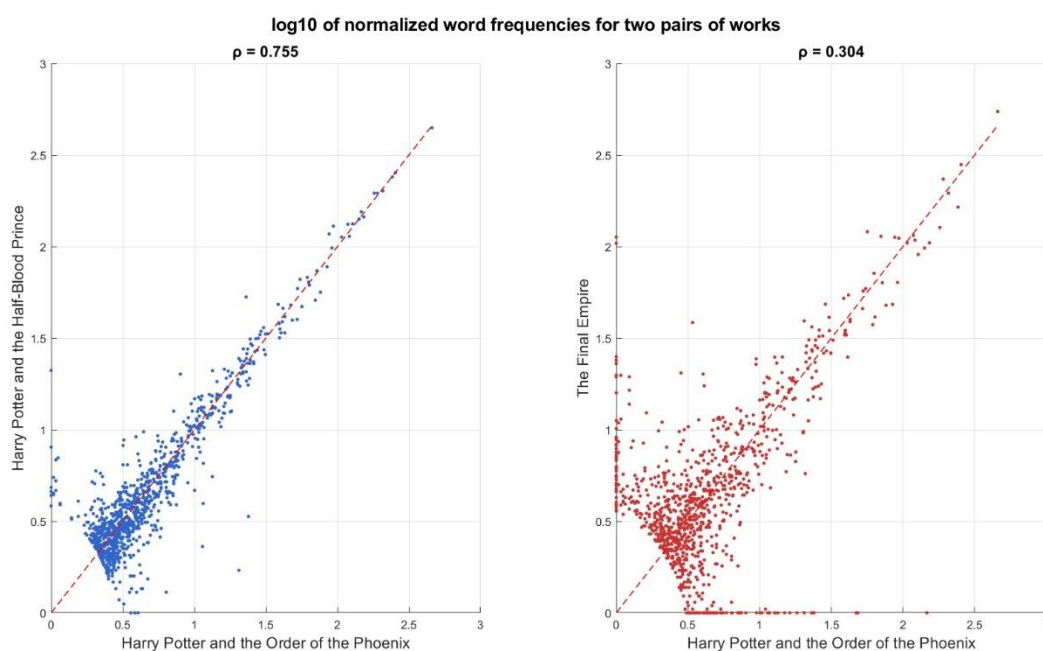
While the Pearson correlation coefficient and Cosine Delta provide an adequate estimate of the degree of similarity/difference only when the distribution of fluctuations in the data is close to normal, Spearman's rank correlation coefficient can be effectively applied to cases with a wide variety of probability distributions. This is extremely important for stylometric tasks, as the specifics of an author's word usage in their works is a complex and currently unexplained phenomenon arising from the intersection of the author's unique psychology, cognitive abilities, social experience, and conscious aesthetic intentions.

To illustrate the operation of Spearman's rank correlation, let us take two

pairs of texts. The first pair – J. Rowling's novels "Harry Potter and the Order of the Phoenix" and "Harry Potter and the Half-Blood Prince"; the second pair – J. Rowling's "Harry Potter and the Order of the Phoenix" and B. Sanderson's "The Final Empire". With 1000 features, Spearman's coefficient between word frequencies in the two novels by J. Rowling was $\rho = 0.755$. In the left graph (Figure 1), a clear linear trend is observed: the points are located near and along the line $y = x$. In the right graph, comparing texts by different authors, the coefficient drops to $\rho = 0.304$. The points form a sparse cloud, significantly deviating from the bisector, indicating weak agreement of lexical profiles. Such a contrasting visual divergence clearly demonstrates the discriminatory ability of the metric: it clearly records stylistic closeness within an author's corpus and just as confidently marks inter-author differences.

Figure 1. Spearman's rank correlation for Rowling/Rowling and Rowling / Sanderson novels

Рисунок 1. Ранговая корреляция Спирмена романов Дж. Роулинг / Дж. Роулинг и Дж. Роулинг / Б. Сандерсон



For stylometric analysis, it is more convenient to use not the Spearman coefficient ρ itself, but the value $1 - \rho$ which we will call Spearman distance. This

transformation is driven by the desire to uniformly describe all metrics used in stylometric analysis.

Jensen-Shannon divergence (JSD) (Lin, 1991; Endres and Schindelin, 2003) is used as a measure of dissimilarity between two probability distributions. Let two probability vectors $P = (p_1, p_2 \dots p_n)$ and

$$JSD(P, Q) = H\left(\frac{P + Q}{2}\right) - \frac{1}{2}\{H(P) + H(Q)\}$$

Here $H(P)$ is the entropy of distribution P , defined as

$$H(P) = - \sum_i p_i \log_2 p_i$$

JSD values range from 0 to 1. Values close to 0 indicate a high degree of similarity between the compared distributions, while values close to 1 indicate large differences. A successful example of using JSD within stylometric analysis can be found in (Arefin et al., 2014). Jensen-Shannon divergence has properties similar to symmetrized Kullback-Leibler divergence, also used in stylometric works (Savoy, 2020: 39-42). However, in the practical application of Kullback-Leibler divergence, one might encounter a problem if one of the frequencies in the compared distributions is 0 – in this case, the divergence value becomes infinite. Jensen-Shannon divergence is free from this drawback.

The Soergel distance is a generalization of the Jaccard distance to the case of comparing two probability distributions (Levy et al., 2025; Willett et al., 1998). The Soergel distance is calculated using the following formula:

$$1 - \frac{\sum_i \min(p_i, q_i)}{\sum_i \max(p_i, q_i)}$$

In the literature, this quantity can also be found under the names Jaccard distance, Ruzicka distance, and Tanimoto distance. When the two compared distributions completely coincide, the sums in the numerator and denominator are equal, and the Soergel distance becomes 0. The largest distance value will be observed when, in each pair p_i, q_i , one of the quantities is 0 (this situation, in relation to stylometric tasks, can

$Q = (q_1, q_2 \dots q_n)$ be given. The values of Jensen-Shannon divergence are calculated using the following formula:

occur if the lexicons of the compared texts do not intersect). In this case, the sum in the numerator is 0, and the distance value is 1.

For Manhattan, Cosine Delta, and Spearman distance, we tested both modes – with standardization of word frequencies (since it helps mitigate the impact of outliers and differences in frequency scale) and without it. For JSD, Soergel distance, and Labbé distance, standardization is not applied: after standardization, numbers can be both positive and negative, which precludes the calculation/application of these metrics; Labbé distance by definition operates on full vectors of relative frequencies without transformations.

The JSD and Soergel distance calculations were carried out using user scripts in the MATLAB language. The Manhattan distance, Cosine Delta, Labbé distance and Spearman distance were calculated using the MATLAB system's built-in functions.

Text Preprocessing and Experimental Protocol.

All texts underwent a unified preprocessing procedure implemented in the MATLAB environment using functions from the Text Analytics Toolbox.

At the first stage, tokenization was performed, preserving punctuation markers and removing numeric sequences. Next, case folding was applied. Stop-word filtering was not used to preserve the full stylistic signal. The analysis was conducted at the wordform level. For the ancient Greek corpus, it was critically important to retain polytonic diacritics. Since the built-in linguistic models of the Text Analytics Toolbox do not support Ancient Greek, lemmatization and

morphological tagging, as in O. Alieva's work, were not performed.

For each studied corpus, its lexicon was compiled, and the total occurrence of each wordform in the corpus was counted. The analysis was based on the most frequent words (MFW), so the list of wordforms was then ranked in descending order of word frequency across the entire corpus. Regarding the choice of the optimal number of features, we repeated all calculations with the number of features varying in the range of 100-3200 in steps of 50.

The texts presented in the corpora have different lengths. To enable comparisons for all authors and works under the same conditions, all texts were divided into fragments of fixed length (with 50% overlap). The range of fragment lengths varied depending on the characteristics of the corpus: for "Greek" – from 625 to 7071 words, for "Sholokhov" – from 1250 to 40000 words, for "Rowling" – from 2500 to 20000 words, for "SF&RP" – from 2500 to 40000 words. Fragment extraction was performed on a logarithmic grid with a multiplier of 2 (i.e., each subsequent length is obtained by multiplying the previous one by ≈ 1.414), which ensures uniform coverage of the length scale on a logarithmic scale and increased sensitivity to changes in accuracy on short segments. After that, for each fragment, the relative frequencies of each wordform present in the corpus were calculated, and then, for each compared metric and number of features, the distances from each fragment to all other fragments of the same length were computed.

Authorship attribution was carried out using the k-nearest neighbors method (k-NN) with $k=1$.

The choice of $k=1$ is standard for stylometric studies using the nearest neighbor method (see, for example, Evert et al., 2017; Jannidis et al., 2015). At $k=1$, the attribution result is maximally sensitive to the choice of metric. Increasing k "blurs" the contribution of a specific metric, averaging the decision across several neighbors. For each fragment of a given work, the most similar fragment

(i.e., the one with the minimum distance to it) belonging to other works in the corpus was found. Attribution was considered correct if the most similar fragment belonged to another work by the same author.

The primary measure of attribution accuracy in this study was the true positive rate (TPR). It reflects the absolute effectiveness of the metric – what proportion of fragments of a given author, on average, the metric attributes correctly. TPR was calculated using the formula:

$$TPR = \frac{N_{true\ positive}}{N_{total}}$$

where $N_{true\ positive}$ is the number of fragments correctly attributed by the algorithm as belonging to the given author, and N_{total} is the total number of fragments of this author subjected to the attribution procedure.

The choice of TPR as the primary metric is driven by the objective of stylometric text attribution: determining what proportion of fragments from a given author are correctly attributed to them. TPR directly measures this value. Accuracy and F1-score are more appropriate for binary/multiclass classification problems with a fixed set of classes, while in stylometry, the number of classes (in this case, authors) varies from study to study.

Results and Discussion

Dependence of Attribution Accuracy on MFW and Text Length.

Before performing evaluation of the metrics across the entire data, it is advisable to establish how each metric responds to changes in feature space size and the length of the analyzed segments. Such analysis allows us to identify parameter ranges where metrics yield the highest accuracy.

For this experiment, we selected three multilingual corpora: "Greek", "Rowling", and "SF&RP", and within each corpus – one author: Plato, J. Rowling, and the Strugatsky brothers. This choice is motivated by the following.

Plato can be considered a benchmark of stylistic uniqueness, since the corpus of his works is homogeneous (all written within a narrow chronological period and belonging to the same genre – philosophical dialogue), and the works themselves possess highly unique stylistic and lexical markers (dialogic form, philosophical terminology, specific particles).

J. Rowling was chosen due to the genre binary nature of her corpus (fantasy, realistic novel) and the associated classical stylometric case (see more on this in section 4.4).

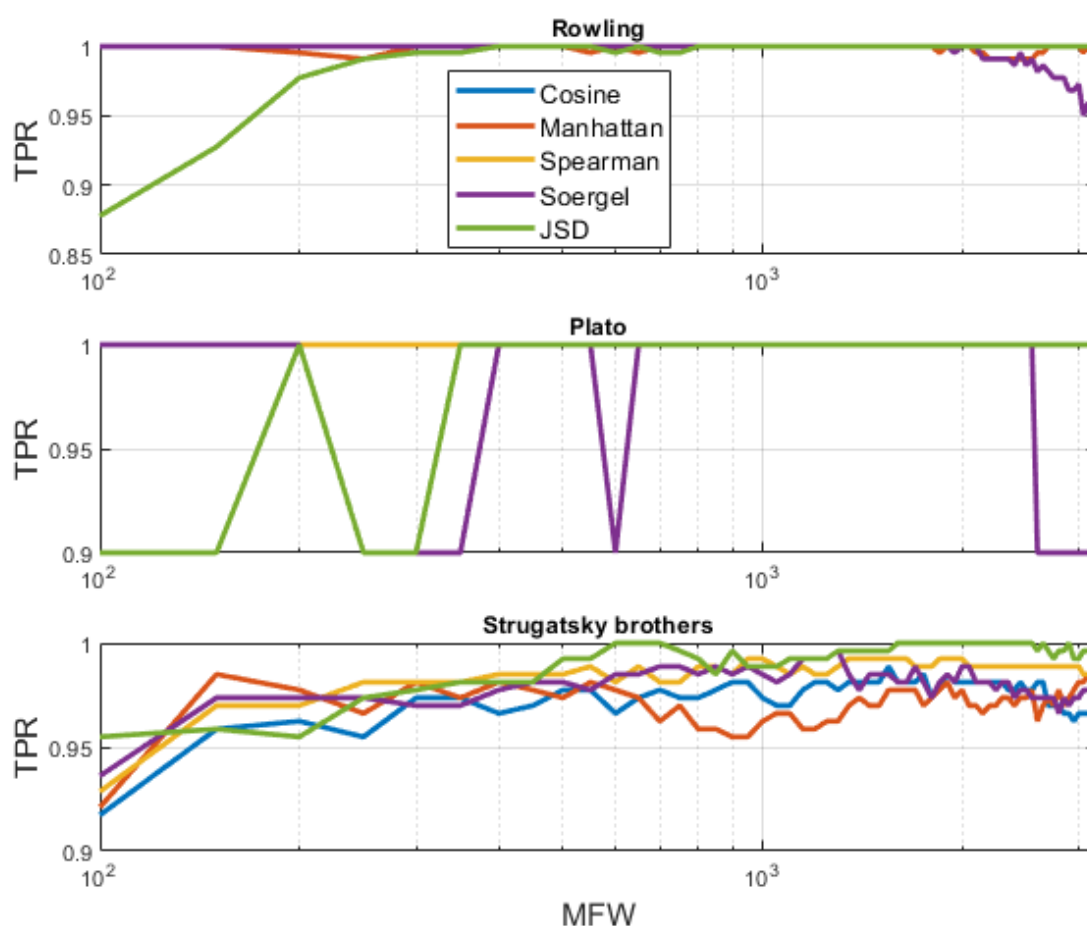
A. and B. Strugatsky represent great research interest, as they create certain difficulties for attribution. The complexity is explained by co-authorship (see stylometric investigations by V. Staroverova (Staroverova, 2025)) and chronological span,

broad genre diversity (novel, novella, short story, and their thematic modifications), as well as the inclusion of the Strugatsky brothers' works within the Russian-language hyper-corpus.

The experiment was conducted at a fixed fragment length of 10000 words, except for Plato, for whom the fragment length was set to 7071 words. The number of MFW features varied from 100 to 3200 in steps of 50. Three metrics – Spearman distance, Cosine Delta, and Manhattan distance – were used with preliminary frequency standardization, the others without standardization. The obtained TPR values for the 5 metrics (excluding Labbé distance) are presented in Figure 2.

Figure 2. Dependence of text attribution accuracy on MFW at a fixed text fragment length.

Рисунок 2. Зависимость точности атрибуции текста от MFW при фиксированной длине текстового фрагмента



First, we note that in many cases, TPR values reach 1.000. Any decision-making algorithm ensures decisions with some non-zero error probability (although under favorable conditions this probability can be extremely small). However, in a real experiment, we estimate TPR from finite sample sizes. The number of text fragments to be attributed may be insufficient for problematic cases to appear in the experiment (this especially applies to the region of long lengths, where the number of fragments becomes relatively small). Therefore, empirical TPR values in many cases are equal to 1.

For Rowling, all metrics achieve an accuracy of 1.000 already at MFW=400. Spearman distance, Cosine Delta, and JSD maintain this level across the entire MFW range. The Soergel distance shows a slight decrease in accuracy at MFW > 2500, likely due to noise introduced by rare words.

For Plato, three metrics – Spearman distance, Cosine Delta, and Manhattan distance – maintain an accuracy of 1.000 regardless of the size of the MFW list. A

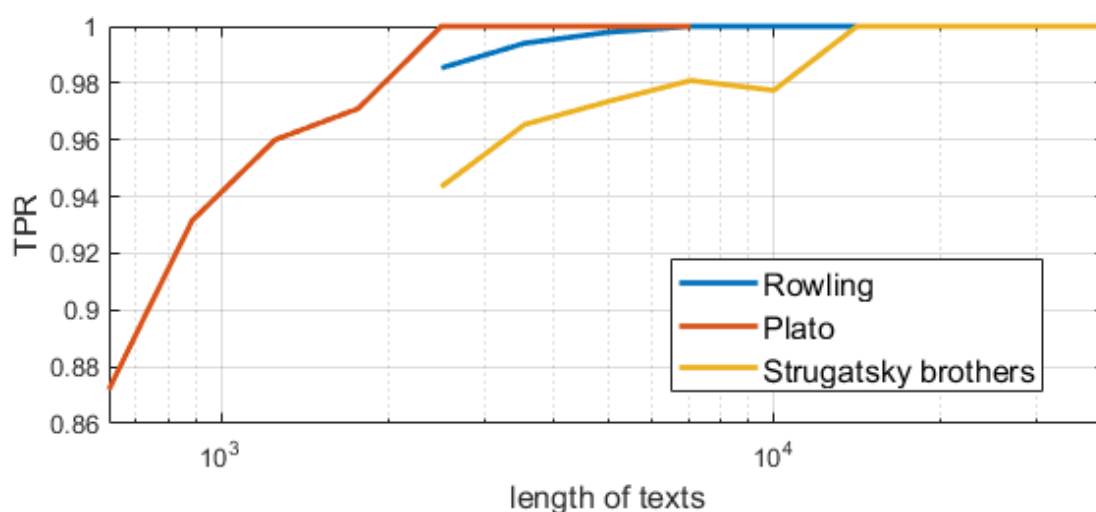
different picture emerges for the Soergel distance and the JSD: under certain feature configurations, their indicators drop to 0.900, which, in our opinion, reflects the increased sensitivity of these measures to the lexical composition of the sample.

The case of the Strugatsky brothers proved to be the most intriguing: JSD confidently reaches 1.000 at 600 and 1700-2200 features, surpassing its competitors. None of the other four metrics reached the maximum value on this case at the given fragment length (10000 words). This observation allows us to consider the entropy approach as a promising tool for texts with complex authorial structure and genre heterogeneity.

The Labbé distance works with the full vocabulary, and the MFW parameter is not set for it. However, its accuracy is sensitive to fragment length (see Figure 3). For the homogeneous Plato, 1.000 is recorded already at 2500 words, for the bi-genre Rowling – after 7000 words, and for the heterogeneous Strugatsky brothers, stabilization occurs only on segments exceeding 12000 words.

Figure 3. Dependence of text attribution accuracy on text fragment length for the Labbé distance

Рисунок 3. Зависимость точности атрибуции текста от длины текстового фрагмента для расстояния Лаббе



Comparison of Attribution Accuracy for Individual Authors.

The MFW and the application/non-application of frequency standardization are

among the tunable parameters, the choice of which is fixed by the researcher at the

experimental design stage. When comparing the effectiveness of different metrics, it seems natural that the optimal values of these two parameters will be selected for each metric independently. Within this study, we adhere to an approach where, for each metric, author, and fragment length, the configuration of parameters that maximizes attribution accuracy is determined. This strategy allows us to evaluate the potential of distance measures under optimal conditions for each of them.

Thus, for each corpus, author, metric, and fragment length, we choose whether to perform standardization or not, as well as the optimal MFW value, and then perform attribution of all fragments. Such calculations were performed for all authors of each of the four corpora with all 6 metrics. The attribution accuracies obtained during the experiment for a total of 46 authors (the figures represent the results for the authors with at least two works in the corpora) are shown in Figures 1-4 of the Appendix.

Spearman distance demonstrates the least dependence on authorial, corpus, and linguistic specifics. The key result of the experiment: when selecting the optimal value of MFW with a text length increase, the use of the given metric allows us obtaining a value of 1.000 for all authors in all four corpora. No exceptions in terms of failure to reach the accuracy ceiling were recorded. The difference between “simple” and “complicated” authors lies not in the fundamental possibility of reliable attribution but in the length of the text that allows one reaching it.

For stylistically homogeneous authors (Rasputin, Plotinus, French, Ivanov), the maximum value is recorded at small fragment lengths and maintained over a wide range of MFW. For authors with pronounced intra-author variability (Belov, Demosthenes, Kryukov, Warrington), the metric also reaches 1.000 but requires longer segments and more precise selection of MFW (most often 600-800). On short fragments, these authors show increased dispersion of indicators, but with

sufficient fragment size, the sample word frequencies stabilize, and the TPR reaches 1.000.

Manhattan distance and Cosine Delta behave almost identically. Manhattan distance reached 1.000 for all authors; Cosine Delta did not cope with Warrington texts (the highest TPR for this author was 0.913). For authors with a homogeneous set of texts in the corpus, the curve stabilizes already with a short length of a text fragment and maintains the maximum value across the entire subsequent range. For authors with pronounced intra-corpus variability, a TPR value of 1 is achieved in the region of longer segments. In case of Astafiev, Serafimovich, and Efremov, Manhattan distance started better than Spearman distance, but in most cases it acted as the one catching up (and always did catch up). The use of standardization noticeably accelerates the achievement of maximum accuracy on short fragments, while on ultra-long segments the difference between the standardized and non-standardized variants disappears.

The Labbé distance shows the sharpest polarisation. For stylistically balanced authors (Rasputin, Galen, Lewis, Fadeev), the curve rises and reaches 1.000 sufficiently quickly, making the metric highly effective on short, homogeneous texts. For the “complex” authors Plutarch, Platonov, Kryukov, the curve reaches a local maximum and then gradually decreases. For Turton and Belov, a monotonic increase is observed with maximum values of 0.586 and 0.833, respectively, even at the maximum fragment length. In several cases (4 observations in the “Sholokhov” corpus, 3 in “SF&RP”, 6 in “Greek”), we see a “rollback” effect: the curve first increases, reaches a local maximum at medium lengths, and then sharply drops. This effect is explained by the metric working with the full corpus vocabulary: on long fragments, rare wordforms and genre shifts accumulate, which disrupt the internal homogeneity of the segments and distort the stylistic signal,

leading to a sharp decrease in accuracy after reaching a local peak.

The Soergel distance curve on homogeneous authors (Mozhaev, Plotinus, Stroud) goes up and quickly reaches 1.000 at minimal lengths. Regarding "complex" authors (Belov, Turton, Platonov), accuracy gets stuck in the range of 0.550-0.850, with the curve not leveling off even when fragment length increases to the upper bounds of the scale.

JSD behaves differently from other metrics: either it quickly reaches 1.000, or it gets stuck at low values; there are almost no intermediate options.

On homogeneous corpora, JSD works almost instantly. For Lucian, Plutarch, Rasputin, and 5 authors of the "Rowling" corpus, the curve jumps to 1.000 already at the first points. Moreover, it often overtakes even Spearman distance and Manhattan. Apparently, for stable texts, the entropy signal is so clear that the metric does not require a long text fragment.

The situation with heterogeneous authors is different. For the same Belov and

Turton, JSD does not rise above 0.833 and 0.793, respectively: genre differences within the work of a single author create an entropy noise that the metric cannot filter out.

When compared with other metrics, JSD works more rigidly: Spearman distance and Manhattan distance tend to 1 smoothly, albeit at different speeds, while JSD either shoots immediately or does not shoot at all. For short, genre-pure texts, this is an excellent tool, but in heterogeneous corpora with internal style dynamics, it often fails.

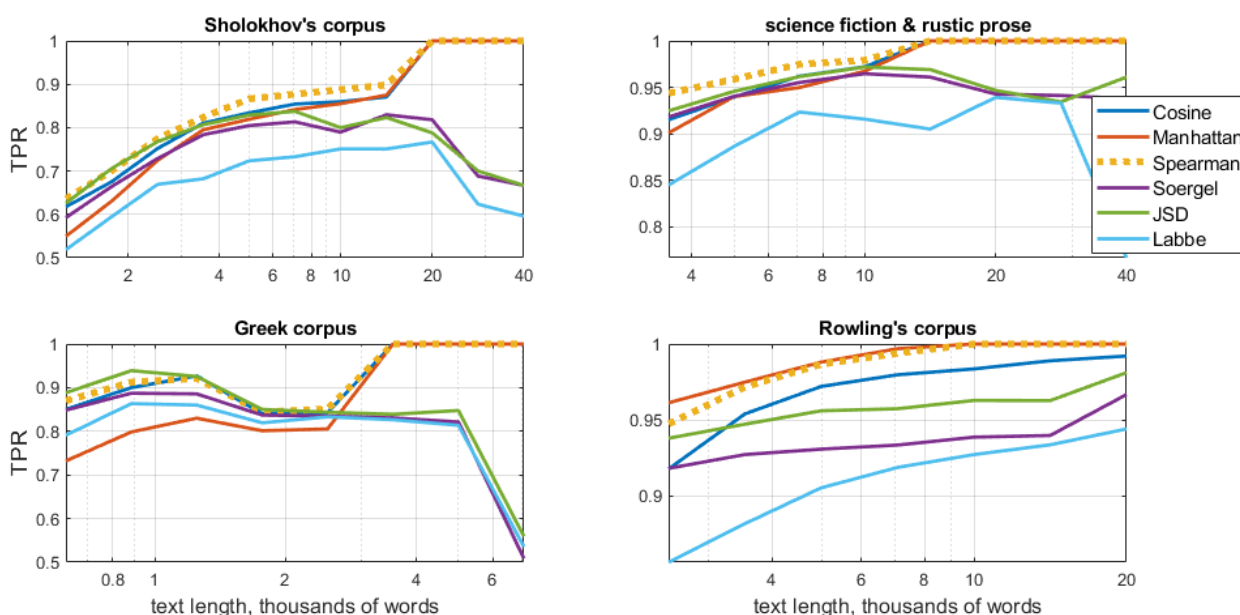
Comparison of Attribution Accuracy for Corpora as a Whole.

In order to present the results in a more convenient and generalized form, for each of the four corpora we averaged the obtained TPR values across all authors included in that corpus. When calculating the average, all authors are weighted equally, regardless of how many of their works are represented in the corpus.

Analysis of the average attribution accuracy reveals a clear hierarchy of metrics (see Figure 4).

Figure 4. Average TPR for 4 corpora

Рисунок 4. Среднее TPR для 4 корпусов



Spearman distance demonstrates the greatest universality, systematically reaching an average accuracy value of 1.000 in all four corpora. Manhattan distance and Cosine Delta show similar dynamics: on homogeneous data, their curves almost merge with Spearman distance, while on heterogeneous materials, they lag only slightly on short segments, also reaching maximum accuracy with increasing text fragment length.

The Soergel distance and the JSD show ambiguous results. Both metrics on the “Rowling” and “SF&RP” corpora grow monotonically, confidently staying in the corridor > 0.900 , while on “Sholokhov” and “Greek”, after rising to 0.850-0.920 on small and medium fragments, they exhibit a noticeable decline to 0.550-0.700.

The Labbé distance shows the least stability growing smoothly only on the “Rowling” corpus and reaching the value of 0.944, while on the other three corpora, its average accuracy curve drops sharply.

For a summary assessment of metrics

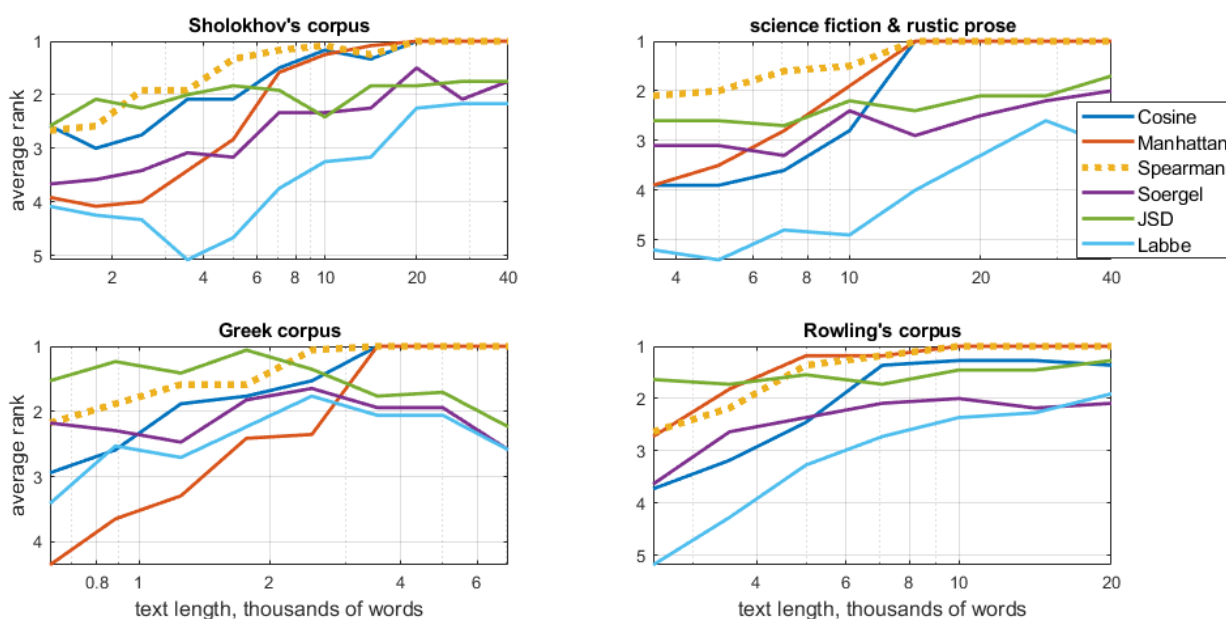
under inter-author variability, average ranks of the metrics were also calculated. They show the place the selected metric occupies among all six metrics in the list ordered by decreasing accuracy.

Ranking was carried out as follows: for each author represented in the corpus at a given text fragment length, the TPR values were ordered in descending order; the metric with the highest accuracy was assigned rank 1, the next – 2, and so on. If several metrics showed the same TPR, each was assigned the best of the numbers they occupy in the ordered list. For example, if three metrics showed the same TPR and occupied positions 2-4 in the ordered list, each was assigned rank 2. This rule was adopted to achieve a more transparent interpretation. In this case, for example, an average rank equal to 1 means that the metric for all authors either surpassed the other metrics or showed a result no worse than any other.

The average rank graphs (Figure 5) make the hierarchy of metrics even clearer.

Figure 5. Average rank for 6 metrics in different corpora

Рисунок 5. Средний ранг для 6 метрик в разных корпусах



As can be seen, Spearman distance is among the leaders on all corpora, regardless of text length; in most cases, it occupies the

first position. Manhattan distance in three out of four corpora noticeably lags behind Spearman distance on short text fragments but

catches up with the leader on all four corpora when the context increases. Cosine Delta, due to a drop on Warrington, does not share the top positions on the “Rowling” corpus. The Labbé distance systematically closes the ranking (ranks 4-5), especially in the corpus of Soviet prose, where the gap from the top metrics reaches 3-4 units.

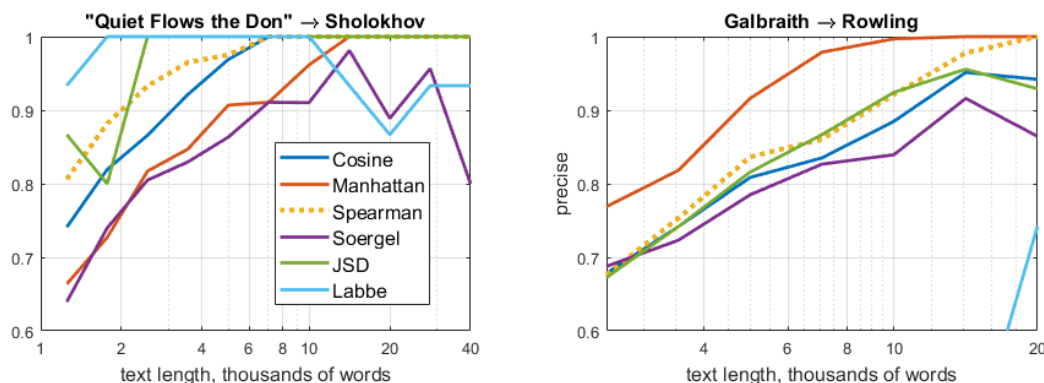
JSD exhibits consistently strong performance on short texts. It outperforms all competing metrics on the “Greek”, “Sholokhov”, and “Rowling” corpora at small sample sizes, and is surpassed only by Spearman distance on the “SF&RP” corpora. However, with increasing fragment length, it drops to the fourth rank. The Soergel distance behaves more reservedly: it rarely rises above 3rd or 4th position, demonstrating smooth but slow growth (the exception is “Greek”, where its curve falls).

The final part of our study is devoted to assessing the *accuracy of metrics in solving two well-known stylometric problems*: establishing the authorship of the novel “And Quiet Flows the Don” (see about this: Velikanova and Orekhov, 2019; Maslinsky, 2022) and identifying the real writer hiding behind the pseudonym R. Galbraith. For each case, we compared the effectiveness of 5 metrics on fragments of different lengths with the optimal MFW values selected for each case, and the Labbé distance on the full vocabulary.

The experiment on attributing disputed works in two heterogeneous cases revealed fundamental differences in the behavior of stylometric metrics depending on the quantitative and qualitative characteristics of the corpus (Figure 6).

Figure 6. Dependence of authorship attribution accuracy on text length for six stylometric metrics

Рисунок 6. Зависимость точности атрибуции авторства от длины текста для шести стилометрических метрик



Analysis of the graph for the “Sholokhov” corpus demonstrates unexpected behavior of JSD. Contrary to generalized expectations, JSD reaches the threshold value of 1.000 in the range of 2300 words, and, unlike the Labbé distance, which reaches 1.000 earlier than others, it does not show a subsequent decrease in accuracy on very long fragments. This behavior is explained by the specifics of the corpus material: with relative genre and size balance of the texts (in the corpus, the only exceptions in size and genre

were the story “The Fate of a Man”; “Don Stories” were not divided but went as a single text array), the entropy signal proves to be sufficiently stable, allowing the information-theoretic measure to instantly capture the authorial invariant. At the same time, rank and frequency metrics increase in a more gradual manner, only reaching their maximum by 7,000–12,000 words. In contrast, despite initially reaching 1.000, Labbé distance loses accuracy after 14,000 words due to the accumulation of stylistic noise and disruption

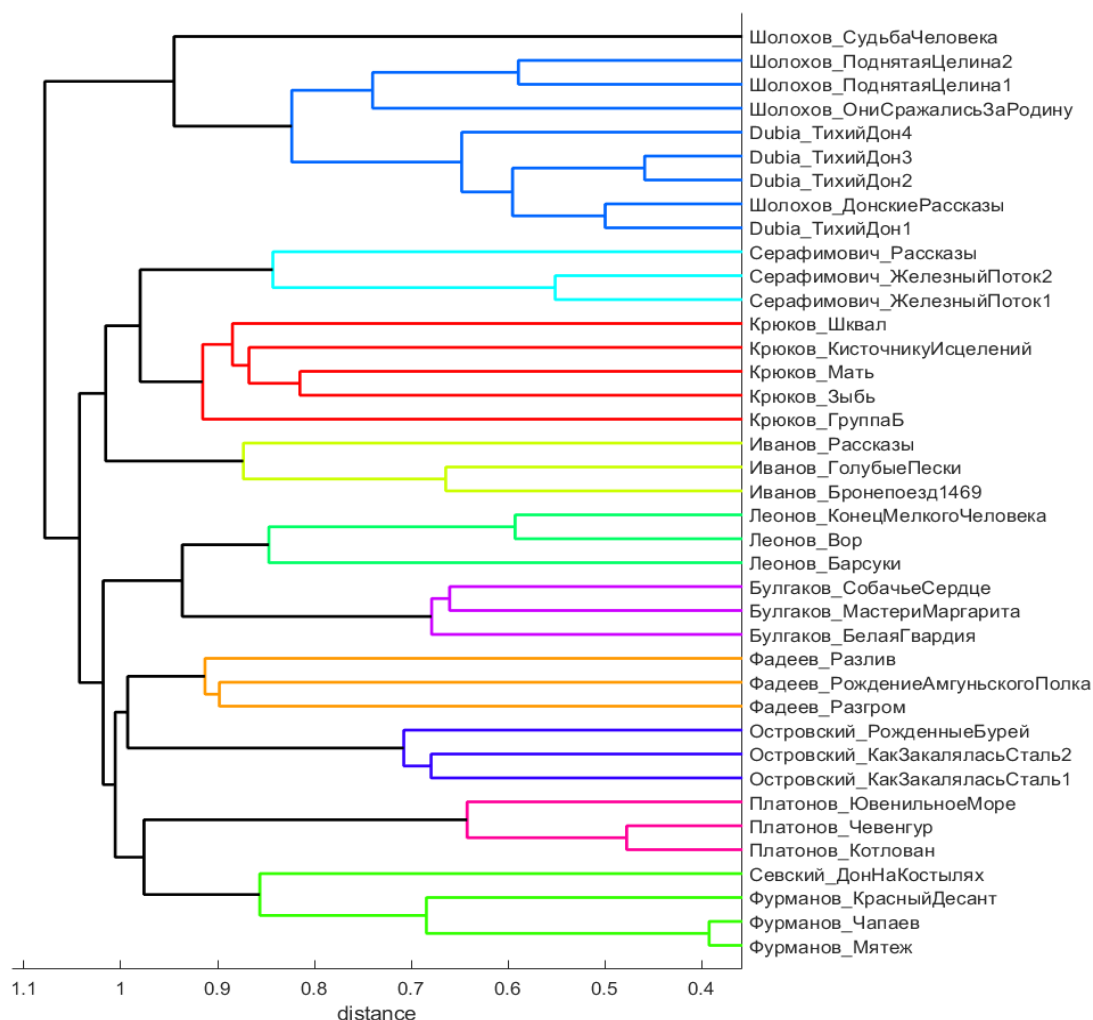
to the internal homogeneity of long segments. This pattern confirms that, in a balanced corpus, JSD can serve not only as an alternative, but also as an advanced attribution tool that combines high convergence speed with robustness to context expansion.

A different picture emerges in the “Rowling” case, where the pronounced asymmetry in text sizes (Rowling’s large-scale novels versus Galbraith’s more compact detective stories) creates specific conditions for attribution. The Labbé distance in these conditions is practically inapplicable: on fragments up to 10000 words, its accuracy approaches zero, and only at larger values (14142-20000 words) is moderate growth up

to 0.750 recorded. This empirical result confirms the theoretical proposition that comparable text lengths are necessary for the correct application of the Labbé distance. Absolute leadership in this corpus is demonstrated by Manhattan distance, reaching a value of 1.000 already at the 10000-word mark and maintaining it with further length increase. Spearman distance also shows high effectiveness, although it reaches the limit value only by 20000 words. The other metrics – Cosine Delta, Soergel distance, and Jensen-Shannon divergence – demonstrate significantly lower results under size imbalance conditions.

Figure 7. Dendrogram of the “Sholokhov” corpus, Spearman distance (with frequency standardization), MFW=2700

Рисунок 7. Дендрограмма корпуса “Шолохов”, расстояние Спирмена (со стандартизацией частот), MFW=2700



Based on the results obtained using Spearman distance, we decided to replicate the dendrogram for the “Sholokhov” corpus presented in the article “Digital Textology: Text Attribution on the Example of M. A. Sholokhov’s Novel “And Quiet Flows the Don”” by N. Velikanova and B. Orekhov. The researchers note: “A certain strangeness is observed only in the fact that “Don Stories” and Sholokhov’s later texts ended up in different parts of the tree. At the same time, this is quite natural: the novel was written over 16 years, the author’s style should have changed during this time” (Velikanova and Orekhov, 2019: 78). Spearman groups the block of “Don Stories”, “And Quiet Flows the Don”, and later works by M. Sholokhov (“Virgin Soil Uplifted”, “They Fought for Their Country”) into one cluster (see Figure 7). With increased distance, the only story in this corpus, “The Fate of a Man”, joins this cluster.

Discussion. The comparison of 6 metrics on four heterogeneous corpora reveals a stable pattern: Spearman distance ensures the highest accuracy of authorship attribution regardless of the genre structure of the sample, fragment length, and MFW. Unlike frequency and entropy approaches, the metric does not show critical failures when moving from stylistically aligned authorial corpora to hyper-corpora with pronounced intra-author variability. JSD behaves in two ways: on homogeneous texts, it quickly captures the authorial invariant; however, under conditions of genre or chronological heterogeneity, its discriminatory ability sharply decreases. Traditional measures (Manhattan distance, Cosine Delta) maintain high effectiveness, but their stability depends on preliminary standardization and precise selection of MFW, whereas Labbé and Soergel metrics prove sensitive to the accumulation of rare wordforms and imbalance in text size.

Such robustness of Spearman distance is associated with the fact that the metric operates not with relative or standardized frequencies of lexical units, but with their ranks, i.e., positions in a list ordered by

descending or ascending frequency. The transition to a rank scale mitigates the impact of extreme outliers and allows obtaining stable results for cases with different probability distributions of empirical word frequencies. Thanks to this, the stylistic signal is preserved even under significant intra-corpus shifts. Jensen-Shannon divergence, measuring the divergence of probability masses, interprets genre and thematic variations as statistical noise under the same conditions, which explains its instability on heterogeneous arrays.

Our results refine the findings of F. Jannidis and colleagues (Jannidis et al., 2015), who recorded the advantage of Cosine Delta and its robustness to the expansion of the MFW list due to vector length normalization. We have shown that the rank approach maintains comparable accuracy but demonstrates a broader area of parametric stability in heterogeneous corpora, where Cosine Delta requires more rigid “tuning”.

The conclusions obtained confirm the results of the study by D. Pronin and E. Kazartsev (Pronin and Kazartsev, 2026) on the effectiveness of entropy measures on controlled samples but point to their limitations when moving to heterogeneous arrays, where Spearman’s rank metric proves more robust.

The results of O. Alieva (Alieva, 2024), obtained on the homogeneous corpus of ancient Greek prose, demonstrate the high effectiveness of Cosine Delta, Tanimoto distance, and Labbé distance. The first metric in our studies also showed high results, but Tanimoto/Soergel and Labbé distance noticeably lost to all others: Tanimoto/Soergel loses discriminatory ability under imbalance in text size, and Labbé distance exhibits a “rollback” effect on long fragments. For “Sholokhov”, “Rowling”, and “SF&RP”, this decline can be explained by genre heterogeneity, accumulation of rare wordforms, and different text size, to which these measures are mathematically sensitive. The discrepancy between the results for the

ancient Greek corpus can be explained by differences in the experimental protocol.

Reviews of quantitative attribution methods (Koppel et al., 2009) and studies of the role of frequency features (Hoover, 2002) show that the effectiveness of a metric is determined by the structure of the corpus. Our results clarify these conclusions: under conditions of genre heterogeneity and volume imbalance, the rank approach proves more robust than frequency and entropy measures.

The totality of results allows us to formulate recommendations oriented towards the corpus structure and research task:

1. For heterogeneous genre collections with size imbalance, Spearman distance is preferred. Shift to the rank scale mitigates the influence of outliers in empirical frequency values and maintains informativeness even with a sparse feature space;

2. For homogeneous corpora, stylistically uniform texts, and short texts, it is advisable to use JSD. With optimal selection of MFW (600-2000 features), the metric provides fast convergence, outpacing frequency approaches;

3. The Labbé and Soergel distances require preliminary size-balance verification. Their application to heterogeneous samples without filtering is risky: the metrics are sensitive to zero frequencies and asymmetry of text lengths.

As we have seen, there is no single metric that yields the best results under all conditions. The question arises about an objective procedure, free from the researcher's personal preferences, for choosing a metric to solve specific tasks. In practice, a situation often occurs where the set of possible candidates for authorship of the text being attributed is known. Under such conditions, it is natural to start work by assembling a corpus of texts by the candidate authors, conducting a simulation similar to that described above, and evaluating the accuracy of different metrics in attributing texts specifically for this set of alternatives. Such a preliminary procedure will allow an

informed choice of the most suitable tool for solving a specific authorship attribution task.

Despite the results obtained, the study has a number of limitations:

The imbalance in author representation within the corpora (especially in “SF&RP”) could have led to a bias in accuracy estimates in favor of authors with larger sample sizes (this concerns primarily Belov with his 17 texts in our hyper-corpus);

The focus on literary texts is due to the fact that stylometric attribution has historically developed precisely in this area and is the most well-developed methodologically. The applicability of the compared distance measures to non-literary genres – such as academic writing, journalistic texts, or social media posts – remains an open question and requires separate investigation.

The corpora we studied demonstrate similar patterns. Further expansion of validation across genres, languages, and historical periods will allow us to clarify the boundaries of applicability of rank and information-theoretic metrics;

The fixed set of MFW (100-3200) does not cover the entire spectrum of possible feature space configurations.

Conclusion. Our study has once again demonstrated that a universal metric does not exist in stylometry; under different conditions, different metrics may prove most useful. However, the distance based on Spearman's rank correlation coefficient more often than other metrics shows the best results, and thus represents the most robust tool for stylometric attribution of heterogeneous text collections. The transition to a rank scale ensures robustness to strong outliers in empirical frequencies and maintains informativeness even under conditions of genre heterogeneity, chronological dispersion, and size imbalance. Unlike frequency and probabilistic metrics, Spearman distance achieves maximum accuracy for all authors with adequate parameter tuning. Jensen-Shannon divergence

demonstrates the conditional nature of its effectiveness: on stylistically homogeneous texts and short fragments, it can surpass basic approaches; however, under conditions of pronounced heterogeneity, the metric's accuracy decreases. Traditionally employed measures remain reliable with sufficient text size whereas Labbé and Soergel metrics exhibit systematic sensitivity to the structural features of the corpus.

Statement on the Use of Generative

AI: The authors declare that no generative artificial intelligence or AI-assisted technologies requiring disclosure under the Journal's policy were used in conducting the research or preparing this manuscript.

Заявление об использовании генеративного ИИ: Авторы заявляют, что при проведении исследования и подготовке настоящей рукописи генеративный искусственный интеллект и ИИ-ассистированные технологии, подлежащие раскрытию в соответствии с политикой журнала, не использовались.

References

- Alieva, O. V. (2024). Distance measures for determining the authorship of ancient Greek texts, *Tsifrovye gumanitarnye issledovaniya* [Digital Humanities Research], 1 (1), 8–33. DOI: 10.31860/cgi-2024-1-8-33 (In Russian)
- Dvinyatin, F. N., Kovalev, B. V. (2024). “Sirin is not inferior to Leonov”: the Delta method for stylometric analysis of Russian novels of the interwar period, *Шаги / Steps* [Shagi/Steps], 10 (3), 207–229. DOI: 10.22394/2412-9410-2024-10-2-207-229. (In Russian)
- Martynenko, G. Ya. (2014). Stylometry: emergence and development in the context of interdisciplinary interaction. Part 1. First steps: 19th century, *Strukturnaya i prikladnaya lingvistika* [Structural and Applied Linguistics], 10, 3–23. (In Russian)
- Martynenko, G. Ya. (2015). Stylometry: emergence and development in the context of interdisciplinary interaction. Part 2. First half of the 20th century: expansion of interdisciplinary contacts of stylometry, *Strukturnaya i prikladnaya lingvistika* [Structural and Applied Linguistics], 11, 9–28. (In Russian)
- Maslinskiy, K. A. (2022). Refined digital textology: once again on the question of the authorship of the novel “And Quiet Flows the Don”, *Russkaya literatura* [Russian Literature], 1, 247–254. DOI: 10.31860/0131-6095-2022-1-247-254. (In Russian)
- Staroverova, P. V. (2025). The Strugatskys separated: Stylometric identification of individual authorial fragments in a collective text, *Tsifrovye gumanitarnye issledovaniya* [Digital Humanities Research], 1, 24–39. DOI: 10.31860/cgi-2025-1-24-39. (In Russian)
- Velikanova, N. P. and Orekhov, B. V. (2019). Digital textology: text attribution using the example of M. A. Sholokhov's novel “And Quiet Flows the Don”, *Mir Sholokhova* [Sholokhov World], 1 (11), 70–82. (In Russian)
- Arefin, A. S., Vimieiro, R., Riveros, C., Craig, H. and Moscato, P. (2014). An information theoretic clustering approach for unveiling authorship affinities in Shakespearean era plays and poems, *PLoS One*, 9 (10), e111445. DOI: 10.1371/journal.pone.0111445. (In English)
- Argamon, S. (2008). Interpreting Burrows' delta: Geometric and probabilistic foundations, *Literary and Linguistic Computing*, 23 (2), 131–147. DOI: 10.1093/lc/fqn003. (In English)
- Bortolato, C. (2016). Intertextual Distance of Function Words as a Tool to Detect an Author's Gender: A Corpus-Based Study on Contemporary Italian Literature, *Glottometrics*, 34, 28–43. (In English)
- Burrows, J. (2002). ‘Delta’: a Measure of Stylistic Difference and a Guide to Likely Authorship, *Literary and Linguistic Computing*, 17 (3), 267–287. DOI: 10.1093/lc/17.3.267. (In English)
- Endres, D. M. and Schindelin, J. E. (2003). A new metric for probability distributions, *IEEE Transactions on Information Theory*, 49 (7), 1858–1860. DOI: 10.1109/TIT.2003.813506. (In English)
- Evert, S., Proisl, T., Jannidis, F., Reger, I., Pielström, S., Schöch, C. and Vitt, T. (2017). Understanding and explaining Delta measures for authorship attribution, *Digital Scholarship in the Humanities*, 32, 4–16. DOI: 10.1093/lc/fqx023. (In English)
- Grieve, J. (2007). Quantitative authorship attribution: An evaluation of techniques, *Literary*

and *Linguistic Computing*, 22 (3), 251–270. DOI: 10.1093/lc/fqm020. (In English)

Grzybek, P. (2014). The Emergence of Stylometry: Prolegomena to the History of Term and Concept, available at: http://www.peter-grzybek.eu/science/publications/2014/grzybek_2014_stylometry.pdf (Accessed 05.04.2026). (In English)

Hoover, D. (2002). Frequent word sequences and statistical stylistics, *Literary and Linguistic Computing*, 17 (2), 157–180. DOI: 10.1093/lc/17.2.157. (In English)

Hoover, D. (2004). Testing Burrows's Delta, *Literary and Linguistic Computing*, 19 (4), 453–475. DOI: 10.1093/lc/19.4.453. (In English).

Jannidis, F., Pielström, S., Schöch, C. and Vitt, T. (2015). Improving Burrows' Delta – An empirical evaluation of text distance measures, available at: https://www.researchgate.net/publication/280086768_Improving_Burrows'_Delta-An_empirical_evaluation_of_text_distance_measures (Accessed 5 April 2026). (In English)

Koppel, M., Schler, J. and Argamon, S. (2009). Computational methods in authorship attribution, *Journal of the American Society for Information Science and Technology*, 60 (1), 9–26. DOI: 10.1002/asi.20961. (In English)

Kocher, M. and Savoy, J. (2019). Evaluation of text representation schemes and distance measures for authorship linking, *Digital Scholarship in the Humanities*, 34 (1), 189–207. DOI: 10.1093/lc/fqy013. (In English)

Kovalev, B. V. (2024). From Classics to Digital Philology: On the Origin and Growth of Stylometry, *Philologia Classica*, 19 (2), 347–360. DOI: 10.21638/spbu20.2024.211. (In English)

Labbé, C. and Labbé, D. (2006). A Tool for Literary Studies: Intertextual Distance and Tree Classification, *Literary and Linguistic Computing*, 21 (3), 311–326. DOI: 10.1093/lc/fqi063. (In English)

Levy, A., Shalom, B. R. and Chalamish, M. (2025). A guide to similarity measures and their data science applications, *J Big Data*, 12, Art. 188. DOI: 10.1186/s40537-025-01227-1. (In English)

Lin, J. (1991). Divergence measures based on the Shannon entropy, *IEEE Transactions on Information Theory*, 37 (1), 145–151. DOI: 10.1109/18.61115. (In English)

Mikros, G., Sosoni, V., Manousakis, V., and Polychroniou, K. (2026). Burrows' Delta as a

Convergent Validator: Stylometric Analysis for Complementary Machine Translation Evaluation. *Journal of Quantitative Linguistics*, 1–29. DOI: 10.1080/09296174.2026.2612931. (In English)

Orekhov, B. V. (2024). Does Burrows' Delta really confirm that Rowling and Galbraith are the same author?, available at: <https://publications.hse.ru/pubs/share/direct/950068199.pdf> (Accessed 5 April 2026). DOI: 10.48550/arXiv.2407.10301. (In English)

O'Sullivan, J. (2025). Stylometric comparisons of human versus AI-generated creative writing. *Humanities and Social Sciences Communications*, 12, 1708. DOI: 10.1057/s41599-025-05986-3. (In English)

Pronin, D. and Kazartsev, E. (2026). Rank-Turbulence Delta and Interpretable Approaches to Stylometric Delta Metrics, available at: <https://arxiv.org/pdf/2604.19499> (Accessed 5 April 2026). (In English)

Rybacki, J. and Eder, M. (2011). Deeper Delta across genres and languages: do we really need the most frequent words?, *Literary and Linguistic Computing*, 26 (3), 315–321. DOI: 10.1093/lc/fqr031. (In English)

Savoy, J. (2020). Machine Learning Methods for Stylometry: Authorship Attribution and Author Profiling, Springer International Publishing, Cham, Switzerland. DOI: 10.1007/978-3-030-53360-1. (In English)

Skorinkin, D. and Orekhov, B. (2023). Hacking stylometry with multiple voices: Imaginary writers can override authorial signal in Delta. *Digital Scholarship in the Humanities*, 38 (3), 1247–1264. DOI: 10.1093/lc/fqad012. (In English)

Smith, P. and Aldridge, W. (2011). Improving Authorship Attribution. Optimizing Burrows' Delta Method, *Journal of Quantitative Linguistics*, 18 (1), 63–88. DOI: 10.1080/09296174.2011.533591. (In English)

Spearman, C. (1904). The proof and measurement of association between two things, *American Journal of Psychology*, 15 (1), 72–101. DOI: 10.2307/1412159. (In English)

Stanikunas, D., Mandravickaite, J. and Krilavicius, T. (2017). Comparison of distance and similarity measures for stylometric analysis of Lithuanian texts, available at: <https://portalcris.vdu.lt/server/api/core/bitstreams/423f881d-b231-4a02-b509-80fd53205028/content> (Accessed 28 April 2026). (In English)

Willett, P., Barnard, J. and Downs, G. (1998). Chemical similarity searching, *J Chem Inform Comput Sci*, 38, 983–996. DOI: 10.1021/CI9800211. (In English)

Литература

Алиева О. В. Меры расстояния для определения авторства древнегреческих текстов // Цифровые гуманитарные исследования. 2024. № 1 (1). С. 8–33. DOI: 10.31860/cgi-2024-1-8-33

Великанова Н. П., Орехов Б. В. Цифровая текстология: атрибуция текста на примере романа М. А. Шолохова «Тихий Дон» // Мир Шолохова. 2019. № 1 (11). С. 70–82.

Двнятин Ф. Н., Ковалев Б. В. «Сирин не уступает Леонову»: метод Delta для стилиметрического анализа русских романов межвоенного периода // Шаги/Steps. 2024. 10 (3). С. 207–229. DOI: 10.22394/2412-9410-2024-10-2-207-229

Мартыненко Г. Я. Стилиметрия: возникновение и становление в контексте междисциплинарного взаимодействия. Часть 1. Первые шаги: XIX век // Структурная и прикладная лингвистика. 2014. № 10. С. 3–23.

Мартыненко Г. Я. Стилиметрия: возникновение и становление в контексте междисциплинарного взаимодействия. Часть 2. Первая половина XX в.: расширение междисциплинарных контактов стилиметрии // Структурная и прикладная лингвистика. 2015. № 11. С. 9–28.

Маслинский К. А. Уточненная цифровая текстология: еще раз к вопросу об авторстве романа «Тихий Дон» // Русская литература. 2022. № 1. С. 247–254. DOI: 10.31860/0131-6095-2022-1-247-254

Староверова П. В. Стругацкие разделенные. Стилиметрическое выявление индивидуальных авторских фрагментов в коллективном тексте // Цифровые гуманитарные исследования. 2025. № 1. С. 24–39. DOI: 10.31860/cgi-2025-1-24-39

Arefin A. S. An information theoretic clustering approach for unveiling authorship affinities in Shakespearean era plays and poems / Arefin A. S., Vimieiro R., Riveros C., Craig H., Moscato P. // PLoS One. 2014. Vol. 9, № 10. e111445. <https://doi.org/10.1371/journal.pone.0111445>

Argamon S. Interpreting Burrows' delta: Geometric and probabilistic foundations //

Literary and Linguistic Computing. 2008. Vol. 23, № 2. Pp. 131–147. DOI:10.1093/lc/fqn003

Bortolato C. Intertextual Distance of Function Words as a Tool to Detect an Author's Gender: A Corpus-Based Study on Contemporary Italian Literature // Glottometrics. 2016. № 34. Pp. 28–43.

Burrows J. 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship // Literary and Linguistic Computing. 2002. Vol. 17, № 3. Pp. 267–287. DOI:10.1093/lc/17.3.267

Endres D. M., Schindelin J. E. A new metric for probability distributions // IEEE Transactions on Information Theory. 2003. 49 (7). Pp. 1858–1860. DOI:10.1109/TIT.2003.813506

Evert S. Understanding and explaining Delta measures for authorship attribution / Evert S., Proisl T., Jannidis F., Reger I., Pielström S., Schöch C., Vitt T. // Digital Scholarship in the Humanities. 2017. Vol. 32. Pp. 4–16. DOI:10.1093/lc/fqx023

Grieve J. Quantitative authorship attribution: An evaluation of techniques // Literary and Linguistic Computing. 2007. 22 (3). P. 251–270. DOI: 10.1093/lc/fqm020

Grzybek P. The Emergence of Stylometry: Prolegomena to the History of Term and Concept. URL: http://www.peter-grzybek.eu/science/publications/2014/grzybek_2014_stylometry.pdf (дата обращения: 05.04.2026).

Hoover D. Frequent word sequences and statistical stylistics // Literary and Linguistic Computing. 2002. 17 (2). Pp. 157–180. DOI: 10.1093/lc/17.2.157

Hoover D. Testing Burrows's Delta // Literary and Linguistic Computing. 2004. Vol. 19, № 4. Pp. 453–475. DOI:10.1093/lc/19.4.453

Jannidis F. Improving Burrows' Delta – An empirical evaluation of text distance measures / Jannidis F., Pielström S., Schöch C., Vitt T. // Book of Abstracts of the Digital Humanities Conference 2015. URL: https://www.researchgate.net/publication/280086768_Improving_Burrows'_Delta-An_empirical_evaluation_of_text_distance_measures (дата обращения: 05.04.2026).

Koppel M., Schler J., Argamon S. Computational methods in authorship attribution // Journal of the American Society for Information Science and Technology. 2009. 60 (1). Pp. 9–26. DOI: DOI:10.1002/asi.20961

Kocher M. Evaluation of text representation schemes and distance measures for authorship

linking / Kocher M., Savoy J. // Digital Scholarship in the Humanities. 2019. Vol. 34. № 1. Pp. 189–207. DOI:10.1093/lc/fqy013

Kovalev B. V. From Classics to Digital Philology: On the Origin and Growth of Stylometry // Philologia Classica. 2024. Vol. 19. № 2. Pp. 347–360. DOI: 10.21638/spbu20.2024.211

Labbé C. A Tool for Literary Studies: Intertextual Distance and Tree Classification / Labbé C., Labbé D. // Literary and Linguistic Computing. 2006. Vol. 21. № 3. Pp. 311–326. DOI:10.1093/lc/fqi063

Levy A. A guide to similarity measures and their data science applications / Levy A., Shalom B. R., Chalamish M. // J Big Data. 2025. Vol. 12. Art. 188. DOI: 10.1186/s40537-025-01227-1

Lin J. Divergence measures based on the Shannon entropy // IEEE Transactions on Information Theory. 1991. 37 (1). Pp. 145–151. DOI: 10.1109/18.61115

Mikros G. Burrows' Delta as a Convergent Validator: Stylometric Analysis for Complementary Machine Translation Evaluation / Mikros G., Sosoni V., Manousakis V., Polychroniou K. // Journal of Quantitative Linguistics. 2026. Pp. 1–29. DOI: 10.1080/09296174.2026.2612931

Orekhov B. V. Does Burrows' Delta really confirm that Rowling and Galbraith are the same author? URL: <https://publications.hse.ru/pubs/share/direct/950068199.pdf> (дата обращения: 05.04.2026). DOI: 10.48550/arXiv.2407.10301

O'Sullivan J. Stylometric comparisons of human versus AI-generated creative writing // Humanities and Social Sciences Communications. 2025. Vol. 12. Art. 1708. DOI: 10.1057/s41599-025-05986-3

Pronin D. Rank-Turbulence Delta and Interpretable Approaches to Stylometric Delta Metrics / Pronin D., Kazartsev E. URL: <https://arxiv.org/pdf/2604.19499> (дата обращения: 24.04.2026).

Rybicki J. Deeper Delta across genres and languages: do we really need the most frequent words? / Rybicki J., Eder M. // Literary and Linguistic Computing. 2011. Vol. 26. № 3. Pp. 315–321. DOI: 10.1093/lc/fqr031

Savoy J. Machine Learning Methods for Stylometry: Authorship Attribution and Author Profiling. Cham: Springer International

Publishing, 2020. DOI: 10.1007/978-3-030-53360-1

Skorinkin D. Hacking stylometry with multiple voices: Imaginary writers can override authorial signal in Delta / Skorinkin D., Orekhov B. // Digital Scholarship in the Humanities. 2023. Vol. 38. № 3. Pp. 1247–1264. DOI: 10.1093/lc/fqad012

Smith P. Improving Authorship Attribution. Optimizing Burrows' Delta Method / Smith P., Aldridge W. // Journal of Quantitative Linguistics. 2011. Vol. 18. № 1. Pp. 63–88. DOI: 10.1080/09296174.2011.533591

Spearman C. The proof and measurement of association between two things // American Journal of Psychology. 1904. 15 (1). Pp. 72–101. DOI: 10.2307/1412159

Stanikunas D. Comparison of distance and similarity measures for stylometric analysis of Lithuanian texts / Stanikunas D., Mandravickaite J., Krilavicius T. 2017. URL: <https://portalcris.vdu.lt/server/api/core/bitstreams/423f881d-b231-4a02-b509-80fd53205028/content> (дата обращения: 28.04.2026).

Willett P. Chemical similarity searching / Willett P., Barnard J., Downs G. // J Chem Inform Comput Sci. 1998. Vol. 38. Pp. 983–996. DOI: 10.1021/C19800211

All authors have read and approved the final manuscript.

Все 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.

Vladimir V. Bochkarev, engineer, Research Laboratory “Multidisciplinary Textual Studies”, Kazan (Volga Region) Federal University, Kazan, Russia.

Владимир Владимирович Бочкарев, инженер, научно-исследовательская лаборатория «Мультидисциплинарные исследования текста», Казанский (Приволжский) федеральный университет, Казань, Россия.

Anton S. Afanasev, Doctor of Philology, Associate Professor, Dean of the Leo Tolstoy Higher School of Russian Philology and Culture,

Professor at the Department of Russian Literature and Methods of Teaching It, Kazan (Volga Region) Federal University, Kazan, Russia.

Антон Сергеевич Афанасьев, доктор филологических наук, доцент, декан Высшей школы русской филологии и культуры им. Льва Толстого, профессор кафедры русской литературы и методики ее преподавания, Казанский (Приволжский) федеральный университет, Казань, Россия.

Rinat A. Bakirov, Candidate of Philology, Associate Professor, Department of Russian Literature and Methods of Its Teaching, Kazan (Volga Region) Federal University, Kazan,

Russia; Research Fellow, Laboratory of Digital Literature and Folklore Studies, Institute of Russian Literature (Pushkin House), St. Petersburg, Russia.

Бакиров Ринат Альбертович, кандидат филологических наук, доцент кафедры русской литературы и методики ее преподавания, Казанский (Приволжский) федеральный университет, Казань, Россия; научный сотрудник лаборатории цифровых исследований литературы и фольклора Института русской литературы (Пушкинский Дом), Санкт-Петербург, Россия.

Appendix

Figure 8. Best probability of correct attribution (TPR) for the “SF&RP” corpus

Рисунок 8. Наилучшая вероятность правильной атрибуции (TPR) по корпусу «SF&PR»

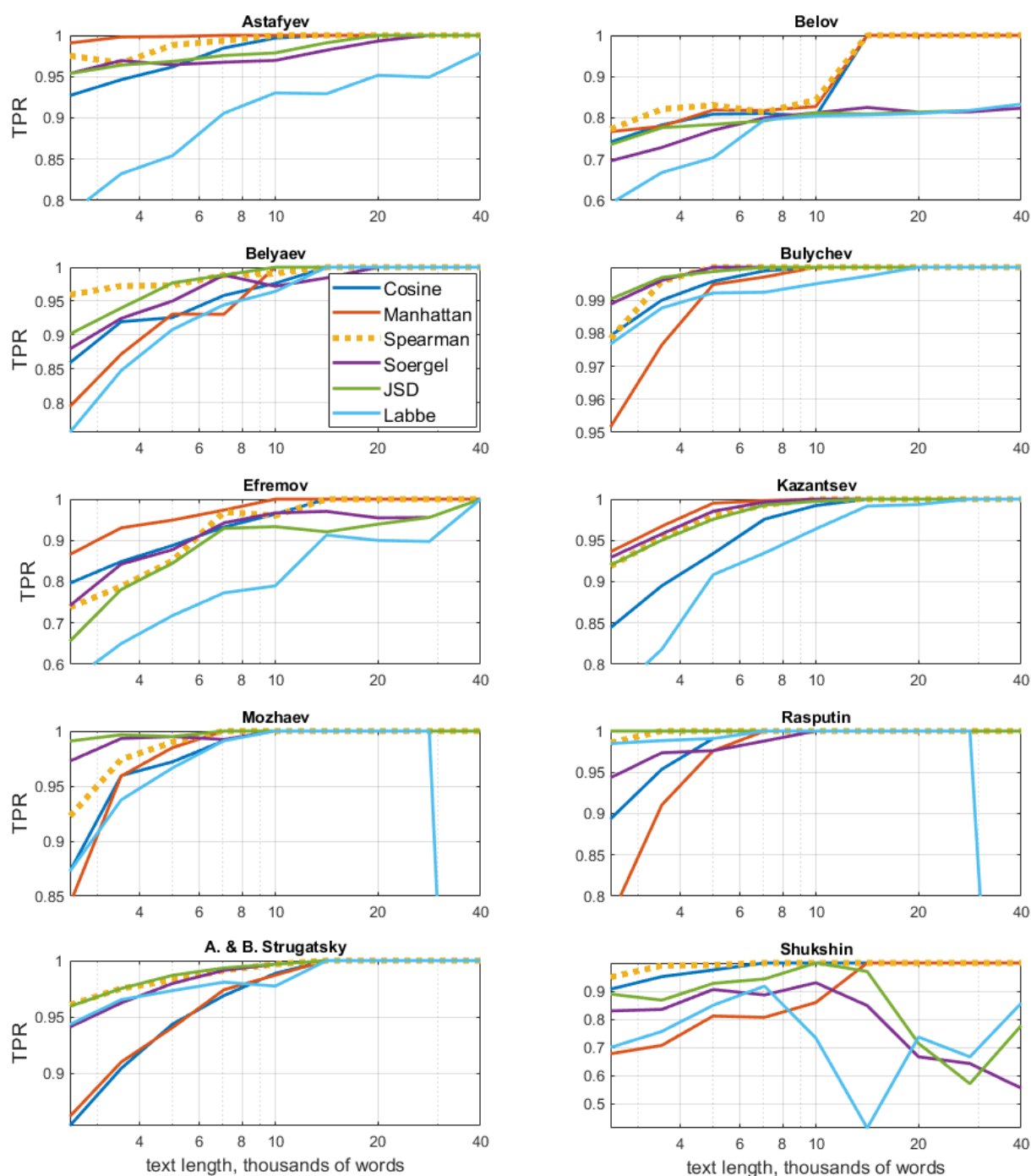


Figure 9a. Best TPR for the “Greek” corpus. Beginning
Рисунок 9а. Наилучший TPR по корпусу «Греки». Начало

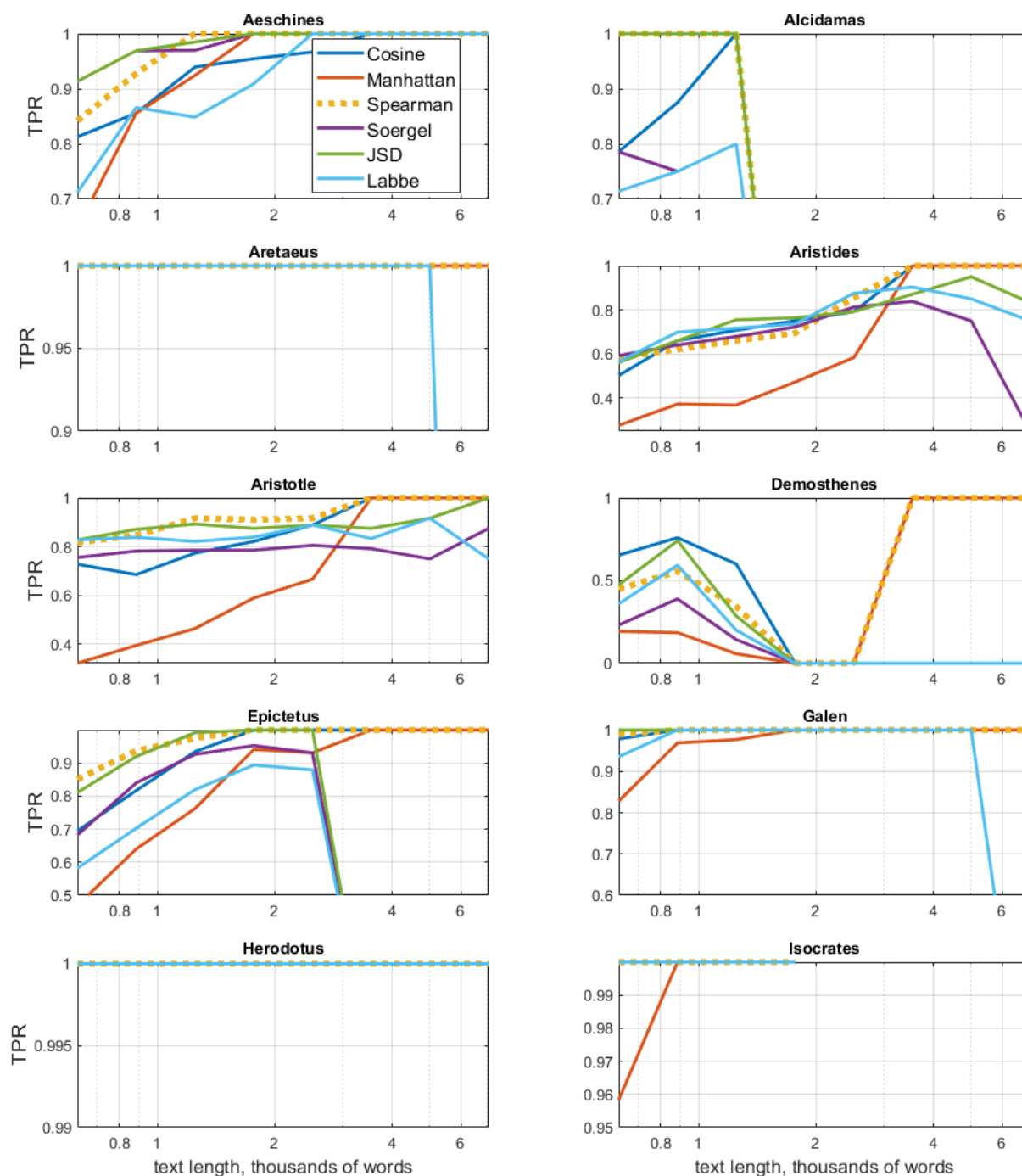


Figure 9b. Best TPR for the “Greek” corpus. End
Рисунок 9b. Наилучший TPR по корпусу «Греки». Окончание

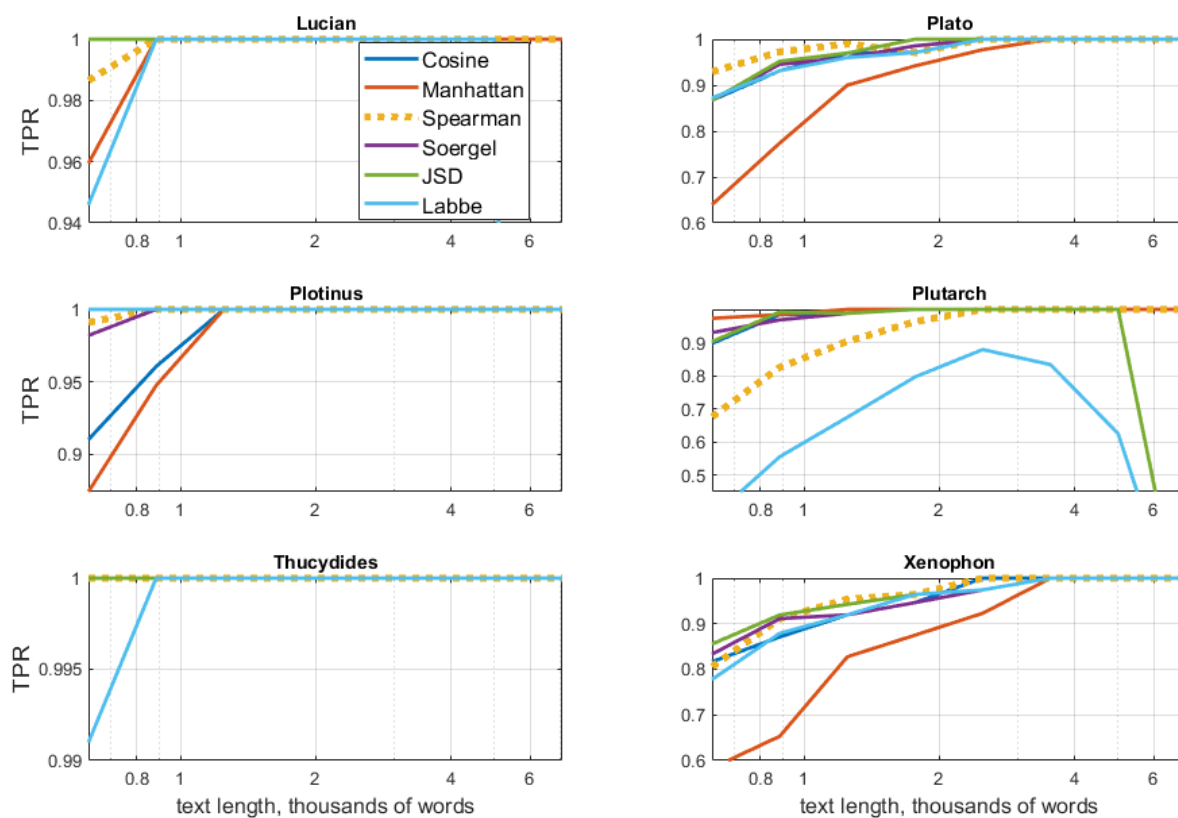


Figure 10. Best TPR for the “Rowling” corpus

Рисунок 10. Наилучший TPR по корпусу «Роулинг»

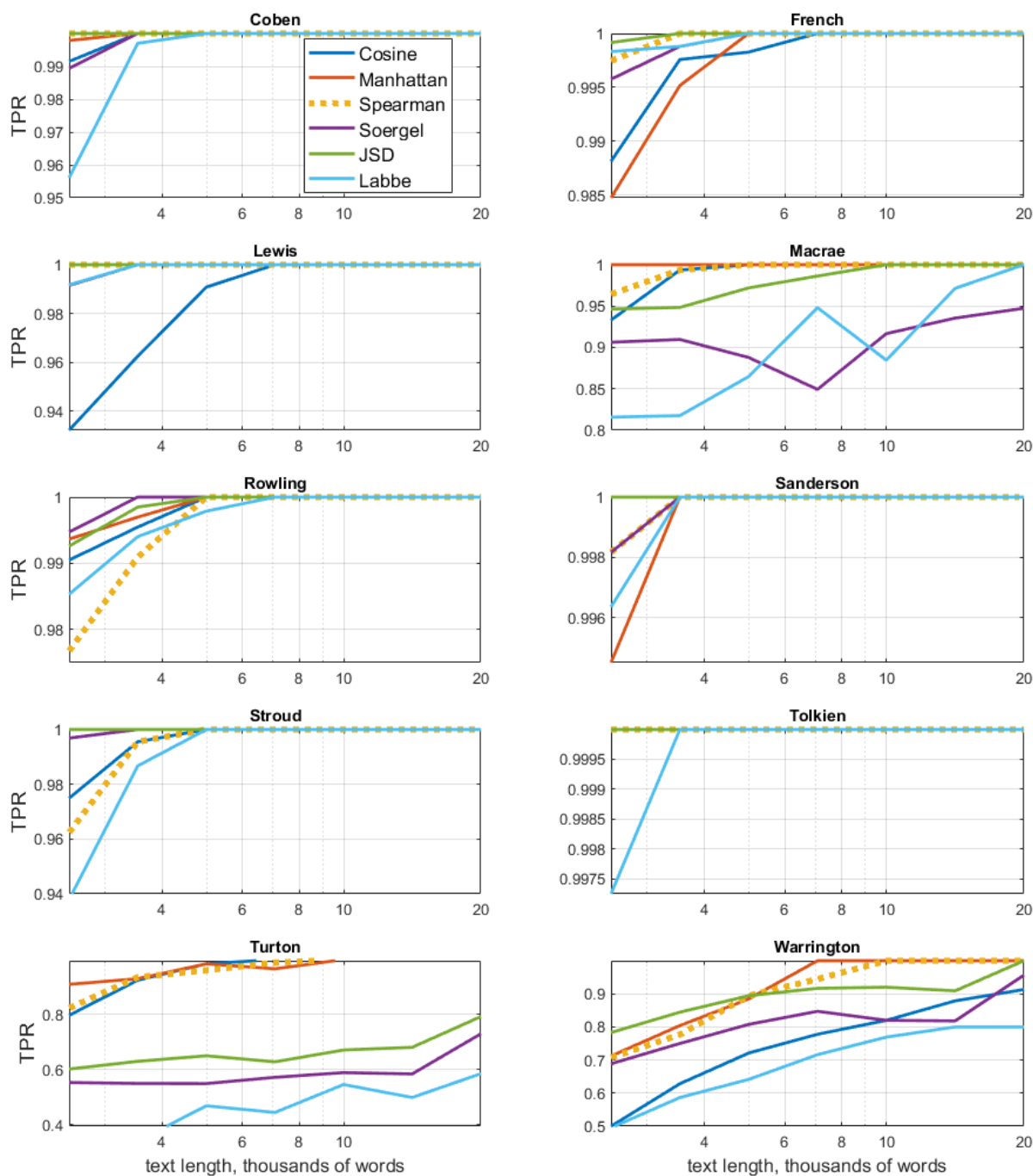
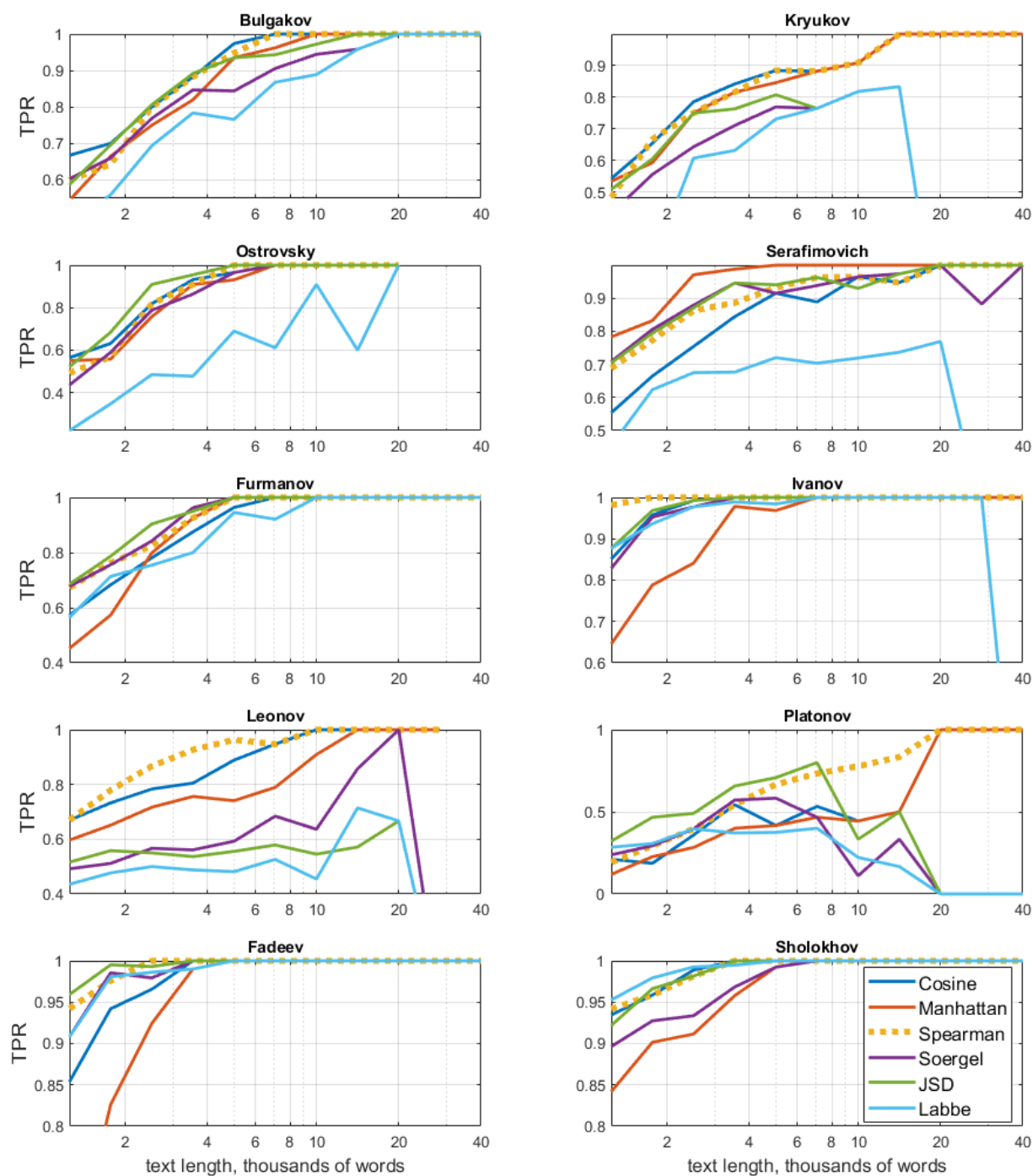


Figure 11. Best TPR for the “Sholokhov” corpus
Рисунок 11. Наилучший TPR по корпусу «Шолохов»



UDC 81'42, 004.93

DOI: 10.18413/2313-8912-2026-12-3-0-4

Leyla Can¹ 
Erdem Akbaş² 

Discourse-Organizing Metadiscourse in ELF Academic
Speech: A Corpus Study Across Speech Events

¹ Erciyes University,
Yenidoğan Mahallesi, Turhan Baytop Sokak, Kayseri, 38280, Türkiye
E-mail: canleylaa@gmail.com
ORCID: 0000-0001-6082-558X

² Erciyes University,
Yenidoğan Mahallesi, Turhan Baytop Sokak, Kayseri, 38280, Türkiye
E-mail: erdemakbas@erciyes.edu.tr
ORCID: 0000-0003-2204-3119

Received 15 June 2025; accepted 15 September 2026; published 30 September 2026

Abstract: This corpus-based study examines how English as a Lingua Franca (ELF) speakers utilize metadiscourse (MD) features to structure and regulate the flow of academic spoken discourse across various speech events. It adopts Ädel's (2022) move-based taxonomy as a framework for a genre-comparative analysis of real-time ELF academic speech events from the ELFA corpus, encompassing various types, including doctoral defense discussions, conference presentations, conference discussions, lecture discussions, and seminars. The study is grounded in a reflexive approach, treating metadiscourse as 'discourse about ongoing discourse' (Ädel, 2006). It focuses on discourse organizing MD features such as announcing and concluding topics, endophoric marking, enumerating, previewing, reviewing, and contextualizing. A corpus-based qualitative and quantitative analysis has been adopted, utilizing the UAM Corpus Tool for qualitative coding identifying contextualized MD functions and quantitative analysis comparing their normalized frequencies across seven event types. Our findings show substantial variation across speech events, although the strength of this variation differs by function. Lectures and presentation events show relatively high rates of prospective resources, particularly topic announcement and previewing. Previewing is relatively infrequent overall but strongly differentiated across events; Reviewing is more frequent overall but does not meet the Bonferroni-adjusted significance threshold for cross-event variation. Doctoral Defense Discussions display a hybrid profile; despite their interactive format, they have the highest normalized rate of Enumerating and substantial rates of Endophoric Marking and Contextualizing, while topic announcements and closures are comparatively rare. We argue that these patterns suggest that prospective and retrospective discourse-management resources respond differently to the structural and interactional demands of academic speech events. The results may be helpful to those working to improve the discourse management skills of students in English for Academic Purposes (EAP) settings, as well as academics involved in international conferences and research.

Keywords: English as a Lingua Franca; Metadiscourse; Academic speech events; Discourse management; Corpus analysis; Spoken academic discourse

How to cite: Can, L., Akbaş, E. (2026). Discourse-Organizing Metadiscourse in ELF Academic Speech: A Corpus Study Across Speech Events, *Research Result. Theoretical and Applied Linguistics*, 12 (3), 80–112. DOI: 10.18413/2313-8912-2026-12-3-0-4

1. Introduction

English serves as a bridge for communication among speakers of different linguistic backgrounds, playing a significant role in the academic world and higher education worldwide. It is a language spoken by individuals with diverse mother tongues (Mauranen, 2009), and a greater number of individuals adopt English as a foreign or second language than as their native tongue (Graddol, 2006). This has resulted in substantial structural and communicative variability in English as a Lingua Franca (ELF) interactions, where linguistic features and pragmatic norms are dynamically negotiated in real time to secure mutual intelligibility (Cogo and Dewey, 2012; Seidlhofer, 2011). As the universal language of communication and higher education, it attracts many researchers who seek to understand how speakers manage the flow of conversations in various academic spoken genres such as lectures, seminars, presentations, and discussions. Each of these genres has its distinct style and purpose, which influences how people organize their discourse and interact with their audience.

Given the increasing prevalence of English-medium programs and the extensive scholarly attention to academic English, it is of great importance to examine how these academic genres contribute to effective communication in the context of academic English as a Lingua Franca. Metadiscourse is a resource that speakers draw upon to manage the flow of discourse in these academic genres. It enables speakers to convey their messages with clarity while also facilitating listeners' ability to indicate their understanding (Warren, 2006: 55). This may mean that managing discourse flow is vital in a spoken environment, where understanding and real-time interaction are needed, aligning with principles of interactional pragmatics and ELF accommodation theory (Bell, 1984;

Cogo and Dewey, 2012). Despite growing interest in spoken ELF contexts, much of the previous research has focused on written academic texts, and relatively few studies have examined MD features across multiple spoken genres. This limits our understanding of how these features function in various academic settings. By examining the role of MD features in organizing discourse flow in different academic contexts, the goal of this present study is to advance the expanding field of ELF research.

Although scholarly interest in spoken metadiscourse (MD) has grown in recent years, particularly with foundational studies on academic lectures (e.g., Ädel, 2010; Mauranen, 2001, 2010), empirical research remains heavily skewed toward written academic texts or isolated spoken genres like student presentations (Ädel, 2022) and EMI lectures (Molino, 2018). Consequently, there is a lack of systematic and cross-event functional comparison of how discourse-organizing MD features adapt to different academic spoken genres in English as a Lingua Franca (ELF) settings. Knowing that speakers utilize MD to manage talk is no longer sufficient; we must understand how these features function across speech events with varying degrees of structure and interactivity ranging from monologic lectures to highly interactive doctoral defenses. This research gap necessitates a context-sensitive, move-based analysis that moves beyond surface-level marker counts to examine how metadiscursive features function within their real-time sequential contexts. This study addresses this need by adopting Ädel's (2022) move-based taxonomy to analyze seven distinct academic speech event types in the ELFA corpus. Our purpose is to systematically map how these discourse-organizing moves (i.e., specifically topic management, phorics, and contextualizing) are deployed to navigate different

communicative demands. To guide this line of inquiry, we attempt to address the following research questions:

RQ1: How frequently do discourse-organizing metadiscourse features occur across different academic speech event types in spoken ELF interactions?

RQ2: Are the observed frequency distributions of metadiscourse categories statistically distinguishable across the seven academic speech event types?

RQ3: In what ways do these metadiscursive patterns and functions vary across the seven speech event types in the ELFA corpus?

2. Literature Review

2.1. The Role of English as a Lingua Franca in Academic Communication

English has been recognized as a leading language in both global communication and higher learning. It effectively functions as a lingua franca. Defined as a language widely shared across various cultures and communities, it facilitates international communication within diverse social and educational settings (McArthur, 2002). When its role in education is examined, its usage is expanding not only in academic publishing but also as a primary medium of instruction. Graddol (2006) states that around 2 to 3 million students study abroad each year, with over half of these enrolled in English-medium programs. Approximately two-thirds of the top 100 universities are situated in countries where English is the predominant language of communication in the educational context. This has led to an increase in the number of universities worldwide that have adopted English as their primary language for teaching. Gnutzmann and Intemann (2008) draw attention to this issue by pointing out that “publish or perish” no longer accurately reflects the current state of academic reality.

‘Publish in English or perish’ would be a better phrase, as more and more publications now require submissions in English. However, as the number of English speakers is skyrocketing around the world, ELF faces the challenge of developing more rapidly than established native varieties of the language (Gnutzmann and Intemann, 2008). In support of this notion, Dewey (2009) argues that English should not be viewed as a rigid collection of standardized forms but rather as an adaptable tool that can be employed in various ways based on context and goal. Therefore, it can be said that the development of international standards is a challenge due to the rapid changes in the ELF field, and to satisfy the demands of users worldwide, ongoing research and modification are required.

Early English for Academic Purposes corpora first focused on studying the language of native speakers as a model for English L2 speakers. However, within a short period, the increasing focus on ELF research has shifted the perspective to effective communication strategies among speakers from diverse linguistic backgrounds (Mauranen, 2010). ELF research, for instance, acknowledges that English L2 speakers usually adapt English to meet their communication requirements (Aertselaer and Dafouz-Milne, 2008; Mauranen, 1993). This leads to language variations that may differ from the traditional grammar or style of native speakers, which indicates the developing nature of ELF. This flexibility is central to managing academic flow in ELF settings where speakers employ techniques such as rephrasing, topic negotiation, and discourse reflexivity (Mauranen, 2007).

Early work in pragmatics, such as House (1996, 2010) and Mauranen (2007), examined how ELF speakers adapt through convergent accommodation strategies in intercultural interactions. Cogo and Dewey (2012) further explored pragmatic adaptation and negotiation of meaning among ELF speakers. In the area of phonology, Jenkins (2009) provides a summary of the

phonological characteristics that enhance intelligibility in ELF communication. She also explores the conflicting and varying perspectives that EFL teachers often have about ELF accents.

Björkman (2011) discusses how ELF influences communication and teaching in higher education contexts. She argues that EAP programs should adapt their teaching and tests to help learners for effective communication in diverse contexts. Within English Language Teaching (ELT), Sifakis (2014) proposes a framework for developing ELF-aware teachers. His research reveals that ELF has significant implications for teachers, particularly in terms of their perspectives on language norms, the roles of native and L2 speakers, and the purpose of teacher feedback in language teaching. Vettorel (2018) furthermore examines how World Englishes and ELF affect English Language Teaching. She demonstrates the importance of communicative strategies in ELF interactions, where speakers create and negotiate meaning through their interactions with each other. She emphasizes that to develop learners' communicative competence, ELT materials should include activities that raise awareness and practice specific communicative strategies.

With these different perspectives in mind, this corpus-based study aims to add to the expanding body of research on ELF by focusing on how ELF speakers manage the flow of discourse in academic speech events. Therefore, it seeks to examine the adaptive, strategic, and context-dependent nature of ELF communication across various genres.

2.2. Shifting Focus to Spoken Academic English: Insights from Corpus-Based ELF Research

While research has been conducted in various areas, the majority of these studies have focused on written contexts. This focus is seen as a direct consequence of academic English research, which, for various reasons, has primarily focused on written communication (Mauranen et al., 2010). The

dominance of written language may be due to the large number of available texts, its common usage in academic and professional settings, and the challenges of analyzing spoken language. However, the spread of English as a lingua franca in academia has also led to the need to research English in the spoken field, where the dynamics of communication differ considerably from those in written discourse (Kaur, 2009; Mauranen, 2005, 2006a; Pickering, 2006; Seidlhofer, 2004).

The importance of speech in academic settings started to be more widely recognized with the development of electronic corpora on academic speech in the late 1990s (Mauranen, 2010). Some of the first major corpora, which include the speech of EAP users, such as MICASE and T2K-SWAL in the USA and the BASE corpus in the UK, led the way for the study of speech in academic contexts. In recent years, frameworks such as English as a Lingua Franca in Academic Settings (ELFA), which investigates communication between English L2 researchers in academic contexts (Jenkins, 2013), have gained importance since their establishment in 2001.

A significant development in this field is the ELFA corpus project, which collected a corpus of spoken English from international degree programs at the University of Tampere in Finland (Mauranen, 2005, 2006a; Mauranen and Ranta, 2008). This corpus has provided authentic data from various academic speech events. To understand how spoken English is used in contemporary academic contexts, it is essential to analyze a corpus from real English as a Lingua Franca settings. That is why numerous studies have employed corpus methodology in their research, providing valuable insights into ELF communication patterns (Björkman, 2009; Cogo and Dewey, 2012; Jokić, 2017; Kaur, 2009; Mauranen, 2005, 2006b, 2007). Therefore, this study focuses on a spoken corpus to investigate the strategies that academic ELF speakers adopt to manage flow, taking into account the developing nature of ELF and the limited research on

spoken contexts. The aim is to investigate how academics organize their discourse in various academic settings, including seminars, lectures, and discussions.

2.3. The Role of Metadiscourse in the Management of Flow: Functions and Implications in Academic Communication

Metadiscourse plays an interpersonal role in managing discourse flow in both spoken and written academic contexts. This has led to considerable research in discourse analysis because of its importance in organizing and guiding communication. In academic texts, the MD differs from that in non-academic texts because it emphasizes supporting and advancing the writer's academic arguments while also providing information with a clear purpose (Hyland and Tse, 2004). Through textual and conversational functions, it can be claimed that it helps scholars to organize their discourse and develop their ideas. This also could enable both listeners and readers to better structure, understand, and evaluate the messages in the discourse. To achieve this, speakers utilize different MD features for specific functions in terms of discourse flow. For example, they signal the beginning and end of topics through features such as announcing and closing to ensure clarity and cohesion. They also utilize other phorics-related features that contribute to the organization of discourse. These include clarifying the sequence of parts through enumerating, guiding the audience to specific points in the discourse with endophoric marking, indicating upcoming or previous content through previewing and reviewing, and commenting on the discourse context through contextualizing (Ädel, 2010). Such functions can be meaningful for maintaining coherence in both spoken and written academic communication.

It is important to recognize that the term metadiscourse has multiple interpretations, with its scope and definition varying across spoken and written language, as well as among different theoretical models. Early

frameworks such as those by Vande Kopple (1985) and Crismore et al. (1993) conceptualized metadiscourse broadly as linguistic elements that guide readers or listeners through a text. Later, scholars like Hyland (2005) framed metadiscourse more narrowly as a self-conscious interactional stance taken by the speaker or writer toward the audience. In contrast, Ädel (2006, 2022) emphasized a reflexive approach, viewing metadiscourse as "discourse about discourse" that remains within the world of communication itself rather than engaging with the external world. Mauranen (2001) also contributed to highlighting metadiscursive strategies, particularly in academic ELF settings. In spoken academic communication, recent empirical investigations demonstrate that discourse-organizing metadiscourse contributes to real-time processing. For instance, Bouziri (2020) highlights how topic-signaling devices establish macro-structural coherence in university lectures, while Kashiha (2022) demonstrates that endophoric marking functions as a primary discourse-organizing resource in academic speech compared to political genres. Furthermore, recent English-Medium Instruction (EMI) studies (Bernad-Mechó, 2024; Doiz and Lasagabaster, 2022) emphasize that metadiscourse serves a critical pedagogical and accommodation function in multilingual tertiary classrooms. Thus, different models offer different points of emphasis, from the management of audience relations to the reflexive organization of discourse.

Because of its complex and diverse characteristics, MD has been characterized as 'fuzzy' (Ädel, 2006; Crismore et al., 1993; Hyland, 2005; Hyland and Tse, 2004; Jiang and Akbaş, 2024). According to Vande Kopple (1985), MD is the language strategies employed by writers or speakers to interact with their audience. Meanwhile, Hyland (2005) defines MD as more than just the transfer of information. It also shows the personalities of the communicators and includes their perspectives and interpretations.

More recent definitions explain that metadiscourse serves two primary purposes: organizing discourse and managing the interaction between speakers and listeners. Hyland and Jiang (2018) describe it as a resource that helps to control the relationships between speakers, texts, and audiences. Akbaş et al. (2017) and Akbaş and Hatipoğlu (2018) emphasize its pragmatic and strategic role in academic writing, and Ädel (2022) presents it as a reflexive discourse integrated into communication. Despite its many definitions and complexities, metadiscourse is often regarded as ‘discourse about discourse’ (Ädel and Mauranen, 2010), which suggests its role in guiding speakers and listeners through the organization and structure of communication. This basic characterization is widely accepted and frequently included in the definitions of MD in numerous academic studies.

Regardless of the framework applied, MD helps to manage the flow of the conversation and enables both speakers and writers to organize their messages effectively (Jiang and Akbaş, 2024) in academic settings. Earlier research focused on MD in various written academic genres, such as research articles (Hyland and Tse, 2004; Mur-Dueñas, 2011), essays (Bondi, 2010; Kawase, 2015; Rubio, 2011), and theses (Akbaş, 2012; Dahl, 2004; Pérez-Llantada, 2010), and also in English as a Lingua Franca (ELF) context (Ädel, 2006; Doiz and Lasagabaster, 2022; Hong and Cao, 2014; Li and Wharton, 2012). Mauranen (2010) argues, however, that MD plays an even more important role in spoken interaction due to the greater need to manage discourse in real-time communication. This suggests that although most MD research has traditionally focused on written genres, there has been a growing interest in spoken discourse in recent years (Ädel, 2010, 2022; Bouziri, 2020; Mauranen, 2001, 2010; Penz, 2017; Zhu, 2018). As a result, there is now increasing recognition of the complexities involved in managing discourse flow in multilingual, interactive, and fluid spoken events such as seminars, presentations, and doctoral defenses. Building upon this, the

current study emphasizes the need to examine metadiscourse features in various real-time academic exchanges.

2.4. Academic Spoken Genres

In academic settings, various spoken genres, such as lectures, seminars, presentations, and discussions, serve distinct purposes. They follow different styles, and these varying styles have an impact on how speakers organize and manage discourse. A genre, according to Swales (1990), is composed of communicative events that share a set of common goals, which in turn influence the expectations and techniques employed by speakers in each situation. While earlier corpus studies focused heavily on written research articles, recent ELF research has expanded to investigate spoken event types. Studies by Bernad-Mechó and Fortanet-Gómez (2019) and Molino (2018) show that monologic lectures rely heavily on sequential enumeration and explicit topic management to orient international audiences. For example, lectures and presentations are typically one-sided, with a single speaker delivering the information to the audience. Mauranen (2009) explains that lectures provide information about conceptual systems, previous research findings, and facts. In other words, they provide familiarity with content that is considered essential knowledge in a specific field. In contrast, recent work on interactive spoken genres (Mauranen et al., 2020; Penz, 2017) reveals that multi-party discussions demand reflexive scaffolding and real-time contextualizing moves to manage open floor negotiations. For instance, seminars and discussions tend to be more interactive, encouraging group discussions and engagement. Therefore, participants share the responsibility for maintaining the flow and success of the conversation (Warren, 2006). Unlike lectures, seminar discussions develop gradually, are more spontaneous, and usually involve fewer layers of simultaneous organization (Mauranen, 2005). These differences are likely to influence how the speakers employ MD features to organize

their speech. In dialogues, metadiscourse plays a much broader role, and this is especially true for discourse management. This aspect is crucial for successful communication, as it facilitates the organization of both the text and the interaction (Mauranen et al., 2010). By examining reflexive metadiscourse across seven distinct spoken event types, this study builds upon these recent insights to provide a comprehensive cross-genre comparative analysis.

Prior corpus research on spoken academic metadiscourse (MD) has developed along two major tracks, corresponding to the monologic and dialogic modes of communication. The first, and most established, track focuses on monologic genres, primarily academic lectures and student presentations, where the speaker bears the sole communicative burden of managing information flow in real time. In these settings, scholars have consistently demonstrated that organizational metadiscourse is for maintaining clarity and managing cognitive load. For instance, research on academic lectures (Dafouz and Perucha, 2010; Kashiha, 2022; Lee and Subtirelu, 2015) highlights that lecturers rely heavily on structural scaffolding (such as endophoric marking and previewing) to establish macro-coherence and guide student comprehension. Similarly, analyses of student presentations (Ädel, 2022; Magnucz, 2012; Rowley-Jolivet and Carter-Thomas, 2005) reveal that effective speakers utilize metadiscursive "moves" to compensate for linguistic limitations, actively structuring their discourse to align with audience expectations. Recent work in English-medium instruction (EMI) contexts (Bernad-Mechó, 2024; Molino, 2018) further supports this, showing that when speakers and audiences navigate linguistically diverse environments, explicit organizational signposting shifts from a rhetorical option to a pragmatic necessity. However, a critical tension arises when moving from these highly structured, monologic settings to interactive, dialogic

speech events (such as seminar discussions and doctoral defenses). In these fluid environments, communication is co-constructed, meaning that speakers cannot rely on pre-planned structural maps. While monologues can have forward-looking textual features (like previewing and topic announcements), interactive talk relies on retrospective and reflexive resources (like reviewing and contextualizing) to resolve comprehension challenges on the spot (Mauranen, 2010; Penz, 2017; Zare and Tavakoli, 2015). Despite these insights, existing literature remains heavily fragmented: studies typically examine monologic lectures (Bouziri, 2020) or dialogic conflict talk (Penz, 2017) in isolation, rather than comparing them systematically. As Jiang and Akbaş (2024) observe, metadiscourse serves as a primary mechanism through which written/spoken academic communicators adapt their language to meet the specific rhetorical demands of distinct genres and disciplinary settings. Thus, by evaluating these metadiscursive strategies across a unified continuum of seven distinct speech events in the ELFA corpus, this study addresses this division, mapping how the functional value of discourse management shifts as interactions become increasingly dialogic and collaborative

3. Methodology

3.1. Data: ELFA Corpus

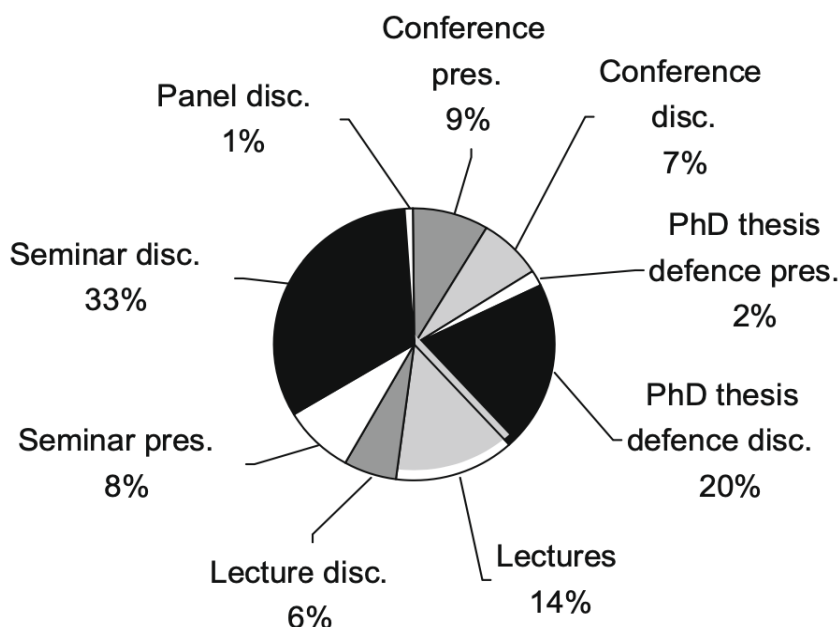
This study draws on the ELFA corpus (English as a Lingua Franca in Academic Settings). It is a rich collection of transcribed spoken academic interactions. Comprising approximately one million words, the corpus represents around 131 hours of recorded speech from various academic events. The data were obtained from several Finnish universities, including the University of Helsinki and Tampere University, as well as their associated technological universities. All ELFA speakers have an educational background that includes formal instruction in English as well as experience of speaking the language. They are generally well-educated,

as most of them possess at least one university degree. The academic nature of the data is important, as it influences how metadiscourse is employed.

The ELFA corpus provides data for the usage of English as a Lingua Franca. It encompasses 650 speakers from 51 different native language backgrounds. Of the participants, about 28.5% are Finnish speakers, while only 5.1% are native or bilingual English speakers. In general, the corpus consists only of authentic data, as it comprises naturally occurring interactions rather than speech explicitly delivered for research purposes. The data comprise 165 speech events of various discourse types, including lectures, seminars, thesis defenses, and conference presentations. Event type is employed as a general term in the data,

although many commonly recognized event types are considered genres (Mauranen, 2006a). The data includes both monologic and dialogic speech, indicating whether there was one speaker or multiple active participants. Both types are included, but the focus is on dialogic events, specifically multi-party discussions. Monologic events, such as lectures and presentations, make up 33% of the data. In contrast, dialogic and polylogic events, which include seminars, thesis defenses, and conference discussions, represent a remarkable portion, totaling 67% of the data. Thus, approximately two-thirds of the data consists of these dialogic interactions. Figure 1 provides a visual representation of the detailed distribution of event types within the ELFA corpus

Figure 1. Distribution of event types in the ELFA corpus Mauranen et al. (2010)



To ensure a systematic, objective, and replicable selection process, speech events from the ELFA corpus were selected based on explicit quantitative thresholds and structural eligibility criteria rather than subjective sampling. To achieve cross-genre comparability, an empirical threshold of per event-type subcorpus was established. Based

on this threshold, minor event categories that individually accounted for only 1% (Panel Discussions) and 2% (Defense Presentations) of the total ELFA corpus volume were explicitly excluded, as their limited data size precluded reliable comparative analysis. Conversely, seven speech event types met the eligibility threshold: *Conference Discussions*,

Conference Presentations, Doctoral Defense Discussions, Lecture Discussions, Lectures, Seminar Discussions, and Seminar Presentations. To prevent arbitrary textual fragmentation and preserve the natural interactional flow, extended single-event transcripts (such as Conference Discussions, totaling 78,484 tokens) were retained in their

complete form rather than truncated to match the exact 60,000-token baseline. Importantly, all eligible files in the ELFA corpus meeting these explicit criteria across the seven selected event types were retained in full, resulting in an exhaustive analytical corpus of 83 complete speech event transcripts.

Table 1. Descriptions of ELFA Corpus Speech Events

	File number	Tokens	Recording Duration
Conference Discussions	14	78484	548 min.
Conference Presentations	21	62946	462 min.
Doctoral Defense Discussions	4	61621	431 min.
Lecture Discussions	12	61561	445 min.
Lectures	8	59022	475 min.
Seminar Discussions	5	60507	373 min.
Seminar Presentations	19	62000	474 min.
TOTAL	83	446,141	3208 min.

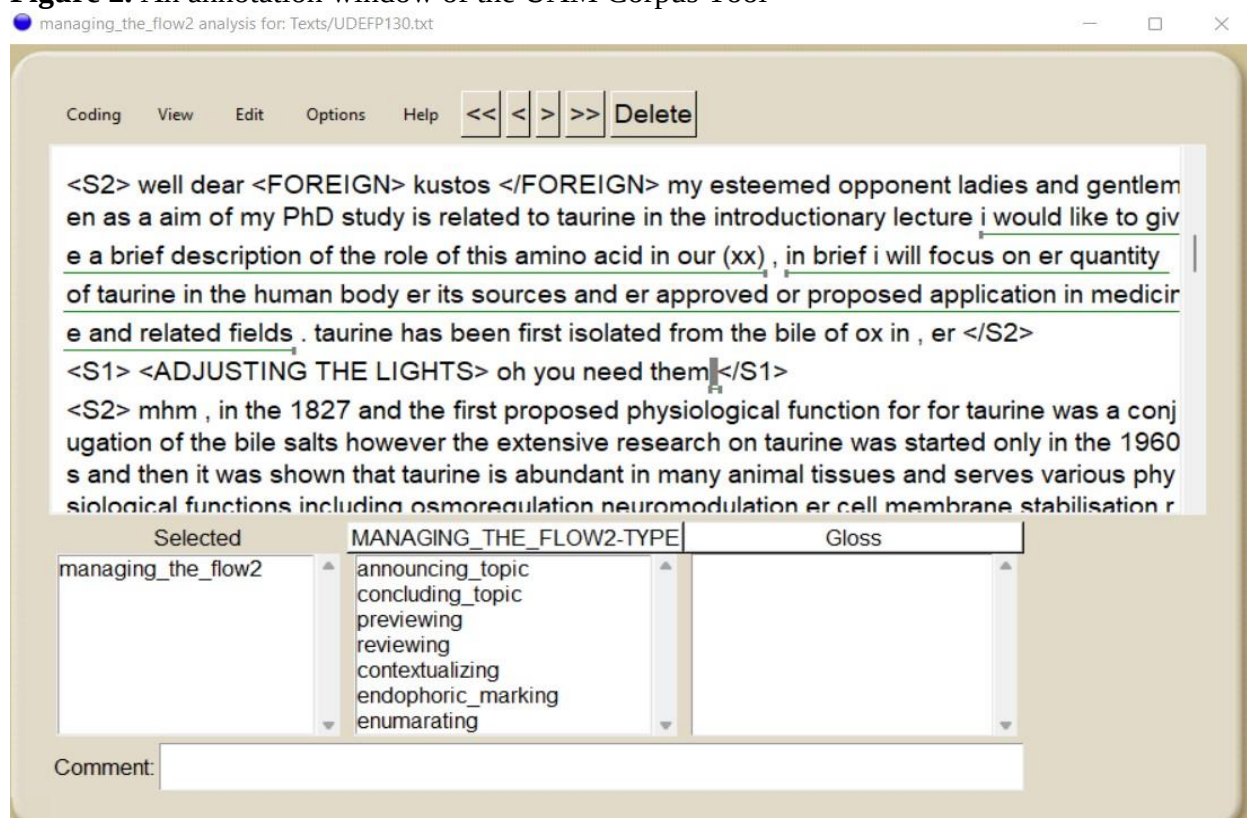
As Table 1 shows, the total corpus comprises 83 texts, totaling 446,141 tokens and approximately 3,208 minutes of recorded data. These include monologic events such as lectures, seminar presentations, and conference presentations, as well as dialogic events including seminar discussions, lecture discussions, doctoral defense discussions, and conference discussions. The study maintains the integrity of each speech event's content while ensuring a fair and methodical comparison of metadiscourse across various academic speech genres by incorporating a range of token counts.

3.2. Data Analysis Process

Data analysis was conducted manually using the UAM Corpus Tool (Version 3.3, O'Donnell, 2008). This tool enables users to annotate texts at various linguistic layers according to their preferences. In this study, 83 texts were initially uploaded to the UAM CorpusTool. Each event was carefully

reviewed and annotated for linguistic elements, with a qualitative focus on coding and tagging segments based on relevant metadiscourse features. While Ädel (2022) proposes a comprehensive taxonomy with multiple dimensions, this study focuses specifically on the features related to managing discourse flow, including announcing the topic, concluding the topic, contextualizing, endophoric marking, enumerating, previewing, and reviewing. To ensure consistency, each feature was coded based on Ädel's (2022) explicit definitions and examples. In cases where coding was not straightforward or required reconsideration, a retreating process was followed, allowing for reevaluation and refinement of earlier decisions. When uncertainty endured, interrater discussion was used to resolve ambiguities and reach a shared decision.

Figure 2. An annotation window of the UAM Corpus Tool



The manual analysis involved a close reading of the texts. This was to ensure that no potential markers were missed during the tagging process. Target markers were first identified using concordance lines, and each marker was then carefully examined to ensure accuracy. This kind of detailed analysis was necessary because metadiscursive expressions often have multiple functions and are highly content-dependent. In this way, it is possible to identify sentences that may appear to be metadiscursive but which do not have the characteristics of metadiscourse. As an example, structures that fall ‘out of the world of discourse’ (Ädel, 2006) are not classified as metadiscourse. For instance, without close reading, the structure in the following sentence would typically be categorized as enumerating.

- (1) The third-world economies became geared more toward the needs of their first world colonial masters than the domestic needs of their own societies. (Seminar Presentations 011C)

- (2) We are the third world’s third largest largest automobile producer. (Seminar Presentations 08B)

In general, the metadiscursive features were clear and concise. However, occasional overlaps were observed, particularly with the feature ‘announcing the topic.’ In cases where metadiscursive features demonstrated multiple functions or overlapped, the primary communicative goal of the speaker served as the basis for the coding decision. For example, in Seminar Presentation 01A (‘I shall explain how above arguments is based on the following subtopics...’), although the speaker was both previewing and announcing upcoming sections, the primary function was coded as announcing the topic because the focus was on introducing the main structure of the talk. In another case, in Conference Presentations 03E (‘Today I will er, if not more, talk about a project that I’ve been coordinating the last few years, so on an e-European level.’), even though the speaker provides a brief overview of the content

(previewing), the central goal is to declare the topic, leading to its annotation as announcing. These examples demonstrate how overlapping instances were handled systematically by emphasizing the dominant function. Such fluidity illustrates the need for context-sensitive annotation in metadiscourse analysis.

In addition, to operationalize Ädel's (2006, 2022) reflexive framework, this study distinguishes between macro-level discourse segments and micro-level functional moves. Ädel's (2022) framework defines metadiscourse based on its reflexive communicative function within the ongoing interaction. Within this framework, an operational 'instance' was defined as a discrete linguistic expression performing a distinct metadiscursive micro-act. In real-time spoken ELF interactions, speakers frequently deploy consecutive metadiscursive signals (e.g., self-corrections, reformulations, or iterative signposts) within a single turn to reinforce structural clarity for multilingual listeners. For instance, in Conference Presentation 05J, the speaker uses several distinct metadiscursive expressions in succession:

(3) 'well I'll this is the s- the er topic of my presentation today I'll be discussing basically that's the title of the paper that I've written here, and er there this is the summary what I generally want to discuss so I'll talk a little bit about the intro- introduce er the the topic there.'

Rather than treating the entire turn as a single broad macro-segment, each expression was annotated individually because it performs a distinct reflexive micro-function: (a) declaring the presentation topic ('this is the topic of my presentation today'), (b) referencing the written paper title ('that's the title of the paper'), (c) delineating the talk agenda ('this is the summary what I generally want to discuss'), and (d) signaling the initial structural move ('so I'll talk a little bit about... introduce the topic there'). This micro-move tagging strategy ensures that the full extent of

online metadiscursive effort expended by spoken ELF speakers is captured. While micro-level tagging yields higher absolute occurrence counts than macro-segmentation, this operational unit was applied with strict consistency across all 83 transcripts, thereby maintaining construct validity and ensuring unconfounded cross-genre comparisons. The UAM Corpus Tool (O'Donnell, 2008) was employed to support both the qualitative coding process and the quantitative analysis of metadiscursive features. The tool provides statistical outputs, including distributions, frequencies, and percentages, which enable researchers to compare linguistic features across different sections of a corpus. In this study, the tool initially facilitated the annotation of all 83 speech events, providing an overview of feature distributions across the data. However, due to the large size and complexity of the annotated dataset, the software occasionally crashed during statistical processing. As a result, the percentages and frequencies presented in this study were recalculated manually based on the raw counts exported from the tool. The findings from both qualitative and quantitative analyses were then compared across the seven speech event types to identify shared and distinctive patterns in the use of metadiscourse. These results are discussed with a focus on the usage, function, and distribution of metadiscursive features within and across the analyzed academic genres.

To ensure the consistency, credibility and replicability of the coding schema, the dataset was evaluated through both intra-rater and inter-rater reliability testing. Reliability testing was conducted on a randomly selected, representative sample of 220 complete metadiscursive cases drawn across all seven academic speech event types. Case-processing analysis confirmed a complete dataset with 0 missing cases (100% valid processing). To evaluate intra-rater consistency, the first author re-coded the entire 220-case sample after a three-month interval to eliminate recall bias. For inter-rater reliability, the sample was

independently coded by a second researcher who held a PhD in applied linguistics specializing in Ädel's (2022) taxonomy. The second coder was not involved in the initial data annotation but was briefed on the specific operational parameters of this study. Cohen's kappa statistics were calculated on initial, unassisted coding to determine baseline agreement before consensus discussions. Initial independent coding yielded 17 divergent decisions out of 220 cases. The baseline inter-rater agreement achieved $\kappa = 0.923$ (Asymptotic SE = 0.20, $p < .001 < .001$), while intra-rater agreement achieved $\kappa = 0.923$ (Asymptotic SE = 0.009, $p < .001 < .001$). According to established statistical benchmarks (Landis and Koch, 1977; McHugh, 2012), kappa values above 0.81 represent almost perfect agreement. Following the calculation of baseline kappa, all 17 initial discrepancies were collaboratively reviewed and raters reached a consensus.

3.3. Statistical Analysis Procedure

To determine whether observed differences in metadiscourse distribution across the seven academic speech event types were statistically distinguishable (RQ2), inferential statistical testing was conducted. Because subcorpus sizes varied across genres (from 59,022 to 78,484 tokens), a weighted Chi-Square Goodness-of-Fit test was performed ($df=6$). The null hypothesis (H_0) specified that metadiscourse occurrences are distributed across the seven speech event subcorpora in exact proportion to their word counts. Expected frequencies were mathematically calculated as $E_i = N_{\text{category}} \times (\text{Tokens}_i / 446,141)$, where Tokens_i represents the word count of subcorpus. All expected cell counts exceeded the minimum assumption threshold $E_i \geq 5$. With this, we aim to ensure that expected values reflect subcorpus size imbalances rather than assuming equal word counts across genres. In addition, an overall 7x7 contingency table Chi-Square test was

performed across all categories and genres. Effect sizes for the category-level goodness-of-fit tests were quantified using Cohen's w (interpreted as small = .10, medium = .30, large = .50), whereas Cramér's V was calculated for the overall 7×7 contingency-table test. Furthermore, while token-level Chi-Square testing, a standard baseline heuristic in corpus linguistics (McEnery et al., 2006) provides a standard baseline in corpus linguistics, we recognize that metadiscourse tokens are nested within 83 transcripts. Because individual tokens within the same transcript are not strictly independent, the Chi-Square statistics are treated as indicative baseline measures rather than definitive proof of causal variation, and inferential claims are qualified accordingly. We also acknowledge that no formal adjustment for multiple comparisons was applied during initial category-level testing. Nevertheless, to account for potential alpha inflation across seven simultaneous Chi-Square tests, we evaluate the resulting individual p -values against a conventional Bonferroni-adjusted significance threshold ($\alpha_{\text{adjusted}} = .05/7 \approx .0071$). Therefore, we interpreted the results falling between .0071 and .05 cautiously as exploratory or marginal trends rather than robustly significant category-level variations.

4. Results and Discussion

To explore how metadiscourse features were employed across different academic speech events, an analysis was conducted to determine the frequency and the percentage distributions of each MD feature. Several findings are identified about how speakers organize the flow and clarify their discourse in terms of MD features. A quantitative analysis of these MD functions across the speech events in question resulted in a total of 1909 instances. The number of tokens employed in each speech event was set at approximately 60,000 To account for unequal subcorpus sizes across the seven speech event types (ranging from 59,022 tokens in Lectures to 78,484 tokens in Conference Discussions),

raw occurrence counts were normalized to a standardized metric of occurrences per 10,000 tokens. Thus, while raw frequencies are retained in all tables for transparency, cross-

event comparisons and genre-sensitive interpretations throughout the results are based on these normalized frequency rates per 10,000 tokens.

Table 2. Overall Distribution of Metadiscourse Features about Managing the Flow

Metadiscourse features	Raw frequencies	Percentages
Enumerating	568	29.75%
Endophoric marking	398	20.85%
Contextualizing	372	19.49%
Announcing topic	286	14.98%
Reviewing	129	6.76%
Previewing	82	4.30%
Concluding topic	74	3.87%
Total	1909	100%

Enumerating is the most commonly employed feature, accounting for 29.75% of the total, as shown in Table 2. This might indicate that speakers needed more sequential organization to guide their listeners through complexity by emphasizing logical order and clarity. Its consistent usage may also help L2 listeners to follow the progression of ideas and points. This feature can help to reduce potential ambiguity in academic speeches by, for example, allowing arguments to be ranked in order of importance and ensuring a clear progression of argument (Ädel, 2006).

Endophoric markings occur in 20.85% of instances, which demonstrates the importance of referring to discourse elements. It can be argued that by frequently employing endophoric references, speakers can maintain a clear and organized flow of information, allowing listeners to concentrate on the topic or points being discussed. As can be observed, both endophoric marking and enumerating are frequently employed in these speech events, supporting Ädel's (2022) analysis of an online MA-level English language teaching context, which suggests that these features are vital in academic settings for achieving coherence and clarity. This is also supported by Kashiha (2022), who emphasizes the importance of endophoric marking in educational genres. His research shows that endophoric marking

is the most commonly employed discourse organizer in university lectures.

We could argue that the importance of endophoric marking in specific academic contexts for organization, coherence, and clear communication is supported by this consistency across studies. However, the findings of Lee and Subtirelu (2015), who found that endophoric markers were rare and among the least employed features in both the L2CD and MICASE corpora, present a contrasting perspective. They attribute this to the nature of classroom communication and its reliance on deictic expressions and gestures, which, as we have previously noted, suggests that multimodal elements may influence the use of MD. The discrepancy between their findings and the current study may also stem from contextual and functional differences in the types of speech events analyzed. Unlike general classroom interactions, academic discourses involve the simultaneous processing of high-level intellectual content and real-time speech, and they offer more sophisticated data than more conventional interactions (Mauranen, 2006a).

The frequent use of discourse organizers similar to those in this study was also observed in Molino's (2018) study on English-medium instruction (EMI) lectures. In her research, discourse organizers were found to be frequently employed to organize

the discourse flow, particularly through the use of impersonal metatext, previewing, reviewing, endophoric marking, and, finally, contextualizing with personal metatext. As a result, it can be inferred that the frequent usage of such features of MD is a characteristic of academic discourse. In these academic settings, speakers are likely to prefer explicit references to ensure the flow of discourse and clarity. This may be especially important when they address diverse ELF audiences who may not share the same physical or situational context.

The frequency of contextualizing reflects a primary orientation toward framing and structuring ideas during discourse organization. Its frequent occurrence in the spoken genre is consistent with Ädel's (2010) findings. She demonstrated that contextualization occurs more frequently in spoken discourse than in written texts. This is most likely because spoken communication has a more flexible structure, and speakers often refer to the production process and their ongoing actions, in contrast to written discourse, which is more rigid and pre-planned. Another possible explanation is that ELF speakers try to make the content accessible by setting the scene or providing background information to manage the flow of conversation.

With a not insignificant usage, the features announcing the topic (14.98%) and reviewing (6.76%) indicate that speakers intentionally introduce and review issues to manage the flow of discourse. The announcement of the topic helps establish the focus of the discourse or signals a shift to a new topic, thus making transitions evident. Similarly, reviewing allows speakers to reinforce or clarify ideas that have already been presented, hence contributing to coherence and understanding by the audience. In contrast to these more frequently employed features, the relatively low frequencies of previewing and concluding a topic may suggest that while references and formal closures are employed, they may not be as big a part of managing flow as other features.

This suggests that ELF speakers in academic speech events rely more heavily on ongoing organization, context building as well as backward referencing than on forward-looking previewing or explicit topic closure. Nonetheless, the comparatively low overall frequency of Previewing can reveal substantial variation across academic speech event types, which could indicate that the resource can be selectively concentrated in particular academic spoken contexts.

Considering the results above, this study is in line with Bernad-Mechó's (2024) findings in several respects. Both studies show that contextualizing is the third most frequent feature, whereas concluding the topic appears less frequent when compared to the other features. Furthermore, a similar pattern appears in both studies in which reviewing is more common than previewing, which emphasizes the shared tendencies in metadiscourse usage. The concluding topic feature had the lowest frequency across courses, according to another study by Bernad-Mechó and Fortanet-Gómez (2019), with values ranging from 0.1 in the American Literature course to 0.4 in others. This suggests that lecturers may not frequently signal the conclusion of topics. The present study supports this observation and further emphasizes the minor role of formal topic conclusions in spoken academic discourse.

These results indicate a broader tendency among speakers to give more weight to elements that support flow management, such as enumerating endophoric markings, than to those that formally close topics. In this context, features like enumerating, which was found to be the most frequently employed, may serve not only structural purposes but also cognitive and communicative ones. As Ädel (2012) argues, enumerating supports listener orientation strategies by guiding the audience through complex content and aiding comprehension. Thus, further research could explore how such features function not only to organize discourse but also to reduce processing load in ELF speech and enhance listener engagement in multilingual academic interactions.

Table 3. Inferential Chi-Square (χ^2) Goodness-of-Fit Tests Across Speech Events

Metadiscourse Category	Total Raw Count	Degrees of Freedom	Chi-Square	Significance	Effect Size
Announcing the Topic	286	6	182.90	< .001	w=.800; Large
Endophoric Marking	398	6	183.67	< .001	w=.679; Large
Enumerating	568	6	157.79	< .001	w=.527; Large
Contextualizing	372	6	50.00	< .001	w=.367; Medium
Previewing	82	6	35.24	< .001	w=.656; Large
Concluding the Topic	74	6	36.54	< .001	w=.703; Large
Reviewing	129	6	17.05	.0091	w=.364; Medium
Overall Contingency Matrix	1909	36	257.42	< .001	V=.150

Regarding the statistical difference across speech event types, inferential statistical testing supported that the distribution of reflexive metadiscourse features varies significantly across the seven academic speech event types $\chi^2(36) = 257.42, p < .001$, Cramér's $V = .150$

Under the Bonferroni-adjusted threshold ($\alpha_{\text{adjusted}} = .0071$), six of the seven individual categories (i.e., Enumerating, Endophoric Marking, Contextualizing, Topic Announcement, Previewing, and Topic Conclusion) remain statistically significant ($p < .001$ to $p = .0045$). In contrast with this, Reviewing ($\chi^2(6) = 17.05, p = .0091$) falls between .0071 and .05 and is interpreted cautiously as an exploratory but marginal trend rather than a category-level difference. these inferential results suggest that cross-genre variations in metadiscourse density are statistically distinguishable across speech event types. While quantitative corpus data alone cannot establish direct causal mechanisms, these distributional patterns

imply that metadiscourse use may adapt to the various communicative and structural demands of distinct academic contexts rather than occurring at random

4.1. Enumerating Across Speech Events

Enumerating reveals the order of certain parts of the discourse by using numbers to relate them to each other. As can be seen in the examples below, it enables speakers to break down content into manageable sections, making it easier for the audience to follow and understand the discourse's structure.

- (4) the first thing I have to say that my title has changed a little bit changed (Conference Presentation 06B)
- (5) okay the third criticism comes from the limits of the narrative analysis (Conference Presentations 03D)
- (6) In the second step we discuss the theoretical framework (Doctoral Defense Discussion 020)

Table 4. Distribution of Enumerating across speech events (raw frequency & normalized rate per 10,000 tokens)

	Doctoral Defense Disc.	Seminar Pres.	Lect.	Conf. Pres.	Lecture Disc.	Seminar Disc.	Conf. Disc.
Raw Frequency	159	101	91	79	60	50	28
Per 10,000 tokens	25.80	16.29	15.42	12.55	9.75	8.26	3.57

According to the data, each genre employs enumerating differently, which is unique to that genre's characteristics. If we were to analyze each one, the data would indicate that enumerating is most prevalent in doctoral defense discussions (25.80 per 10,000 words). This far exceeds the density observed in monologic genres such as Seminar Presentations (16.29 per 10,000) and Lectures (15.42 per 10,000). This high reliance on sequential signposting aligns with the specialized communicative and evaluative demands of doctoral defenses. It may also suggest that in these high-stakes academic speech events, both doctoral opponents/examiners and candidates utilize enumerating as a cognitive and pragmatic tool to navigate complex multi-layered discourse. This may also stem from the formal and argumentative nature of doctoral defenses. The high frequency of enumerating in lectures (15.42 per 10,000 tokens) and seminar presentations (16.29 per 10,000 tokens) further underscores the necessity for presenters to organize content clearly for the audience. This may be explained by the fact that academic writing and lectures share similarities in that both require content to be arranged in a clear, hierarchical manner (Thompson, 2003). At the lower end, "enumerating" appears less frequently in seminar discussions (8.26 per 10,000 tokens) and even less so in conference discussions (3.57 per 10,000 tokens), where discourse unfolds through spontaneous turn-taking rather than pre-planned or highly formalized

argumentation. This could suggest a less straightforward, more adaptable discourse style that is often present in these diverse and interactive discussions. In light of the results obtained, it can be suggested that enumerating is more prevalent in formal, structured contexts, such as doctoral defenses, seminar presentations, and lectures, where it facilitates the organization of information for a coherent discourse flow. Nevertheless, it is less common in events that are more flexible and do not require strict sequencing, like conference discussions.

4.2. Endophoric Marking across Speech Events

Endophoric marking concerns instances in which reference is made to a particular point in the discourse. It does not clarify whether it is before or after the current point. Thus, unlike previewing and reviewing, endophoric marking occurs in situations where this distinction is irrelevant (Ädel, 2022). It can be employed, for example, to draw the audience's attention to specific materials, such as a table, a particular point in a handout, or a visual aid. The examples that follow illustrate this point clearly. The speaker in example (7) draws the audience's attention to specific funding sources, most likely using a slide or chart as a visual aid. In example (8), the speaker points to a particular line, directing the audience to focus on one specific detail. Similarly, in example (9), the speaker draws attention to the last page of a handout, and in example (10), the speaker

combines explanation with visual observation, directing the audience to notice specific color changes within a film. The implication here is that the speakers employ endophoric marking to reinforce the explanation by linking it to a dynamic visual element.

(7) if you look you you see er the, the the the sources of financing change (Conference Presentations 05J)

(8) if you look at the second line, there

is the new bussiness opportunity (Conference Presentations 08B)

(9) if you look at the last page of that that copy you see this er chart with united nations at various bodies under that (Lecture 01A)

(10) here is the movie which shows you this effect if you look at the green and the red color you will see that at a certain point red color goes the intensity goes up (Lecture 023A)

Table 5. Distribution of Endophoric Marking across speech events (raw frequency & normalized rate per 10,000 tokens)

	Doctoral Defense Disc.	Seminar Pres.	Lect.	Conf. Pres.	Lecture Disc.	Seminar Disc.	Conf. Disc.
Raw Frequency	101	103	79	59	30	19	7
Per 10,000 tokens	16.61	16.39	13.38	9.37	4.87	3.14	0.89

The normalized distribution per 10,000 tokens demonstrates clear variation in how endophoric markings are utilized across academic speech event types. The highest frequency of endophoric marking is observed in Seminar Presentations (16.61 per 10,000 tokens) and Doctoral Defense Discussions (16.39 per 10,000 tokens). In seminar presentations, student speakers seem to employ endophoric markers to cross-reference visual aids, slides or preceding sections of their talk to structure information and guide their peers. Similarly, in doctoral defenses, candidates and examiners frequently utilize text-internal references to point directly to specific chapters, data tables or arguments within the dissertation to substantiate claims during rigorous academic evaluation. We also find that monologic lectures exhibit a high density of endophoric markers (13.38 per 10,000 tokens), which can serve a pedagogical function to organize complex course material and orient students across multi-part topics. Moderate use of endophoric markers occurs in Conference

Presentations (9.37 per 10,000 tokens), in which presenters employ endophorics to connect slides and revisit key points efficiently under strict time limits. Unlike structured presentations or high-stakes defenses, Lecture, Seminar and Conference Discussions (4.87; 3.14 and 0.89 per 10,000 tokens, respectively) as dialogic speech events unfold through spontaneous, turn-by-turn interaction in which speakers directly respond to immediate conversational turns rather than referencing pre-planned textual structures, resulting in significantly lower densities of endophoric marking across these interactive speech events.

It is also important to note that in ELF academic communication, particularly in lectures and presentations, endophoric marking is often not limited to verbal cues alone. As pointed out by Brogгинi and Murphy (2017), speakers often reinforce these features with multimodal resources, such as pointing to slides, referring to handouts, or employing spatial deixis. This suggests that these multimodal strategies enhance audience

orientation and reinforce the rhetorical function of endophoric markers by linking spoken content to visual or physical elements in the communicative environment. This pattern is consistent with the findings of Zare and Tavakoli's (2015) study. They found that endophoric markings were frequently employed in academic lectures because handouts and slides were readily available and that limited references to particular parts of the discourse explained their absence in dialogic forms. In conclusion, endophoric markers are deployed predominantly in structured, formal settings such as presentations and lectures. In contrast, they occur less frequently in interactive discussions where text-internal cross-referencing is less necessary.

4.3. Contextualizing Across Speech Events

Contextualizing is the comment of the speaker or writer on the planning, production, and organization of discourse. According to Ädel (2022), it often involves justifying the choices made in structuring texts, which helps to understand how the discourse was produced. This is an important part of managing the flow of discourse, as the speakers ensure that the audience understands the reasons for the course of the discourse. Thus, this allows for a more coherent flow, prepares the audience for what is to come, and explains why certain decisions have been taken. The significance of contextualizing in spoken discourse aligns with the findings of Luukka (1994) and Ädel (2010). Both of these scholars emphasize the high occurrence of contextual metadiscourse in spoken academic texts. Luukka (1994) attributes this to the 'live' nature of oral presentations, where speakers have to structure their arguments in real time without the possibility of revision. Similarly, Ädel (2010) notes that contextualizing occurs primarily in spoken data due to the limited time available for planning and editing in real-time discourse.

The following examples illustrate the urgent need for speaker organization and

justification to manage the flow effectively. The speaker's statement in (11), for instance, that a decision will be made during the presentation helps in flow management by providing expectations for the audience. It indicates that the speaker is explaining the preparation time, which may influence how the audience interprets the timing and pace of the talk. Similarly, in (12), the speaker's explanation of the available time for discussion helps manage the flow. The speaker explains how the schedule was planned, which keeps the session structured and helps the audience know what to expect. In (13), the speaker justifies not including a long bibliography, which helps to avoid confusion and keeps the focus on the main points. Finally, in the example from lecture 020A, the speaker explains why they are starting with a particular concept. This enables the audience to understand the flow of ideas and why this starting point can be essential. As is clear from these examples, these explanations facilitate real-time flow management by providing the audience with a more transparent and predictable understanding of the discourse's structure.

(11) it is not so easy to unders- okay but er before it one more er excuse me i have prepared not for (one two) hours and that's why i er, er i have to decide during my presentation er (xx) okay one example er, later I will speak about library architecture. (Lecture 030)

(12) but we thought that two days might be enough if we have a two full days we <S16>(yes)<S16> could squeeze we know now enough of each other's topics so that we go rather directly and we we shall thoroughly discuss each case half an hour at least maybe three quarters but we could do them in two days if we al- take two days and even eve-evening between for some drinking. (Seminar Discussion 04B)

(13) i did actually want to mention as well that i didn't write a long bibliography i have actually quite many

sources but i didn't put them down.
(Seminar Discussion 04B)
(14) i start with that it is not the one we
start with in in the paper but i think it's
better to start there to set the scene or

what is going on and for the discussion
of the three concepts we are going to to
compare in my presentation. (Lecture
020A)

Table 6. Distribution of Contextualizing across speech events (raw frequency & normalized rate per 10,000 tokens)

	Doctoral Defense Disc.	Seminar Pres.	Lect.	Conf. Pres.	Lecture Disc.	Seminar Disc.	Conf. Disc.
Raw Frequency	83	38	64	45	51	61	30
Per 10,000 tokens	13.47	6.13	10.84	7.15	8.28	10.08	3.82

As illustrated in Table 6, the highest frequency of contextualizing is observed in Doctoral Defense Discussions (13.47 per 10,000 tokens). This elevated rate underscores how candidates and examiners frequently contextualize discourse to clarify research parameters, justify methodological decisions, and outline the scope of their arguments under high-stakes academic evaluation. Monologic Lectures follow with a high density of 10.84 per 10,000 tokens. This aligns with Ädel's (2010) findings in MICASE lectures, demonstrating that university lecturers regularly employ contextualizing moves to make spontaneous meta-comments such as managing time constraints or framing ongoing explanations to maintain discourse coherence and audience orientation. We also find that high to moderate usage is evident in dialogic settings such as Seminar Discussions (10.08 per 10,000 tokens) and Lecture Discussions (8.28 per 10,000 tokens). In such interactive contexts, we suggest that participants use contextualizing as a pragmatic resource to signal shifts in focus, contextualize spontaneous contributions and tie incoming comments to preceding discussion points. Conversely, lower densities of contextualizing occur in formal presentations, including Conference Presentations (7.15 per 10,000 tokens) and Seminar Presentations (6.13 per

10,000 tokens). This may be because they tend to be more formal and pre-structured. Speakers in such settings might mainly rely on prepared scripts or visual aids, which would reduce the opportunity for contextual comments. In contrast, Conference Discussions also exhibit the lowest rate of contextualization (3.82 per 10,000 tokens), most likely because these relatively fast-paced, turn-taking exchanges leave limited space for extended contextual meta-comments, which may reduce the opportunity for extended contextual explanations.

Taken as a whole, these findings underline the role of contextualizing in adapting to the specific characteristics and organizational needs of each speech event. We can argue that the management of academic flow is supported by the strategic usage of contextualizing, which varies across genres. This variability emphasizes the importance of adapting metadiscourse strategies to the communicative demands of different academic genres, making it relevant to explore how contextualizing is linguistically realized in real-time academic discourse. For instance, speakers may contextualize through declarative structures (e.g., 'I presented it that way here, but of course there might be different possibilities...' (Doctoral defense disc. 010)), justifications (e.g., 'I just added it

there after I printed the handouts...' (Lecture disc. 050)), or metalinguistic comments (e.g., 'Let me explain why I structured it this way...' (Doctoral defense disc. 020)). By explicitly revealing the speaker's planning and reasoning, such expressions facilitate real-time discourse flow and support audience orientation. They illustrate how contextualizing functions as both a metadiscursive strategy and an interactional resource in these academic events.

4.4. Announcing the Topic Across Speech Events

Announcing the topic is a feature that is employed to introduce or open a new topic, as the term suggests. Speakers or writers utilize this feature to inform listeners or readers about the topic they will discuss, making it easier to follow the discourse's flow. This feature can be marked by various expressions, as shown in the examples below, such as 'I want to talk about...', 'well I'll this is the s- the er topic of my presentation today i'll be discussing...', 'i'm going to give you a short

lecture on...', 'well erm, erm the subject I erm choosing for my for my essay is...', 'so first I would like to say about...'

(15) I want to talk about the free trade area of the americas negotiations and its main challenges (Conference Presentations 05I)

(16) well i'll this is the s- the er topic of my presentation today i'll be discussing basically that's the title of the paper that I've written here (Conf. Presentations 05J)

(17) my name is >NAME S1> and i'm going to give you a short lecture on finnish economic and social history (Lecture020)

(18) well erm, erm the subject I erm choosing for my for my essay is the human rights in Estonia (Seminar presentations 01D)

(19) so first I would like to say about a relation of political system at a whole as a whole the political system in Zambie (Seminar presentations 08B)

Table 7. *Distribution of Announcing the Topic across speech events (raw frequency & normalized rate per 10,000 tokens)*

	Doctoral Defense Disc.	Seminar Pres.	Lect.	Conf. Pres.	Lecture Disc.	Seminar Disc.	Conf. Disc.
Raw Frequency	12	108	51	55	17	19	24
Per 10,000 tokens	1.95	17.42	8.64	8.74	2.76	3.14	3.06

The normalized distribution of announcing the topic reveals a sharp functional divide between monologic presentation genres and interactive discussion settings. As shown in Table 7, Seminar Presentations display the corpus-wide peak frequency of topic announcements, reaching 17.42 instances per 10,000 tokens. This is followed by Conference Presentations (8.74 per 10,000 tokens) and Monologic Lectures (8.64 per 10,000 tokens). In these structured

settings, speakers tend to rely heavily on explicit topic announcements (e.g., "Today I'd like to focus on...", "My main focus is...") to set the communicative agenda, signal transitions between sub-topics or visual slides, and provide clear macro-structural orientation for their audience. In contrast, we find that interactive discussion genres exhibit substantially lower reliance on explicit topic announcements. Normalized rates drop to 3.14 per 10,000 tokens in Seminar

Discussions, 3.06 per 10,000 tokens in Conference Discussions, 2.76 per 10,000 tokens in Lecture Discussions, and reach a corpus-wide low in Doctoral Defense Discussions (1.95 per 10,000 tokens). Since discourse in dialogic settings could unfold through dynamic turn-taking, open floor management or examiner-driven questioning, participants may rarely need to formally announce topics. Instead, the findings show that conversational focus shifts fluidly in response to immediate interactive contributions rather than pre-planned thematic agendas. These results are consistent with those of Kashiha (2022), which show that announcing the topic appears in various modes with differing usage. Kashiha (2022) found that topic introduction was more frequent in university lectures due to their naturally topic-based structure, in which lecturers commonly employed it at the beginning of their speech to guide students. In a similar vein, Ädel (2010) discovered that, especially in academic discourse, the frequent announcement of the topic in lectures is an effective tool for maintaining audience focus and engagement. However, as previously mentioned, the announcing of the topic occurs less frequently in interactive and discussion-based events such as conferences (24 times), seminars (19 times), lectures (17 times), and doctoral defenses (12 times). This lower frequency can be explained by the dynamic

and less structured nature of discussions. Topic changes in these formats tend to occur naturally through conversation rather than being introduced explicitly.

4.5. Reviewing Across Speech Events

Reviewing describes a discourse strategy that draws attention to previous points and allows the audience to recall information that has already been discussed. This function enables maintaining the flow of the discourse by reminding the listeners of earlier discussions. This strengthens the links between ideas, allowing the audience to follow the logical progression of ideas as it minimizes repetition as shown below:

- (20) that I explained earlier that's the reason why I didn't get back to that (Doctoral Defense Discussions 020)
 (21) as as i pointed out earlier already er we we do not know exactly if if the unemployment goes down (Doctoral Defense Discussions 020)
 (22) of course this has been taken as as being proved zone we talked about that earlier (Seminar Presentations 04E)
 (23) participatory democracy where the UNIP was ruling party as I have mentioned before in December 90 1990 er president kaunda signed this legislation (Seminar Presentations 011A)

Table 8. Distribution of Reviewing across speech events (raw frequency & normalized rate per 10,000 tokens)

	Doctoral Defense Disc.	Seminar Pres.	Lect.	Conf. Pres.	Lecture Disc.	Seminar Disc.	Conf. Disc.
Raw Frequency	20	25	10	27	22	10	15
Per 10,000 tokens	3.25	4.03	1.69	4.29	3.57	1.65	1.91

For Reviewing, as indicated earlier, the Chi-Square test yielded $\chi^2(6) = 17.05, p = .0091$. When evaluated against the Bonferroni-adjusted threshold ($\alpha_{adjusted} = .0071$), no formal statistical significance was reached. Consequently, cross-genre variations in Reviewing could be interpreted as a marginal and exploratory trend rather than a statistical difference. However, Table 8 with normalized distribution shows that the highest frequency of reviewing is found in Conference Presentations (4.29 per 10,000 tokens), implying that presenters refer back to earlier points to help ELF listeners from diverse linguistic backgrounds follow the ongoing discourse. Relatively high rates of reviewing are also observed in Seminar Presentations (4.03 per 10,000 tokens) and Lecture Discussions (3.57 per 10,000 tokens), in which speakers use reviewing moves to link presentation points together and revisit notable concepts during interactive dialogue. Conversely, Monologic Lectures display a relatively low incidence of reviewing (1.69 per 10,000 tokens), suggesting that lecturers in this corpus may prefer introducing new subject matter over systematically reviewing prior knowledge. These findings suggest a pattern in which reviewing strategies appear to be used more frequently in specific settings to support coherence, reinforce arguments and improve mutual understanding tailored to the specific communicative demands of each academic event. In keeping with Molino's (2018) findings, reviewing serves a valuable discourse-structuring function in academic speech, where personal pronouns (such as we) frequently function as retrospective signposts to hold audience attention. While Molino (2018) observed frequent reviewing in EMI lectures, the lower reliance on reviewing in the present ELFA lecture sample highlights potential genre variability, potentially indicating that conference and seminar presentation settings may demand more explicit retrospective scaffolding than standard monologic lectures.

4.6. Previewing Across Speech Events

Previewing draws the audience's attention to upcoming content. It is a way for the speaker to indicate what the next topic of discussion will be. This will make the speech more structured and will be easier for the audience to follow. To put it another way, by clearly stating what will be discussed later, speakers establish a sense of continuity and coherence. For instance, in (24), 'demographic transition, which I'll explain later more closely' (Lecture 020), the speaker tells the listeners that a detailed explanation will be provided. This allows the listeners to anticipate the critical points. Similarly, the speaker reduces possible confusion by assuring the audience that the subject will be covered in (25) by saying, 'I will come into the content of that a little bit later on' (Lecture 020A). The speaker employs previewing in (26), 'so they did not give up although I will explain later in my presentation' (Seminar Presentations 01A), to keep the audience interested and to ensure their focus remains on the details that will be discussed later. Likewise, in (27), the speaker reminds the audience of a particular point by saying, 'they will have different start position for the next er I will come back here' (Seminar Presentations 04E), which reinforces the structural flow of the discussion. Last but not least, in (28), 'I want to say a little bit later something about the exports and the imports and the trade' (Seminar Presentations 08B), the speaker helps the audience follow the flow of ideas by informing them of an upcoming focus on trade-related issues. As these examples demonstrate, by managing audience expectations and maintaining coherence throughout the speech, previewing contributes to regulating the flow of discourse.

(24) demographic transition which i'll explain later more closely (Lecture 020)

(25) I'll come into the content of that a little bit later on (Lecture 020A)

(26) so they didn't gave up although I will explain later er in my presentation (Seminar presentations01A)

(27) they will have different start position for the next er I will come back here (Seminar presentations 04E)

(28) I want to say a little bit later something about the exports and the

imports and the trade (Seminar presentations 08B)

Table 9. Distribution of Previewing across speech events (raw frequency & normalized rate per 10,000 tokens)

	Doctoral Defense Disc.	Seminar Pres.	Lect.	Conf. Pres.	Lecture Disc.	Seminar Disc.	Conf. Disc.
Raw Frequency	4	19	21	18	7	11	2
Per 10,000 tokens	0.65	3.06	3.56	2.86	1.14	1.82	0.25

While previewing accounts for a relatively small proportion of metadiscourse overall (4.30%), inferential testing confirms that its distribution varies significantly across speech event types ($\chi^2(6) = 35.24, p < .001$). This statistically significant contrast implies a sharp functional divide; that is, monologic lectures (3.56 per 10,000) rely on forward-looking previewing to guide student expectations, whereas interactive discussions (0.25 per 10,000) dispense with pre-planned previews in favor of real-time adaptive talk. Table 9 above illustrates how the structural organization and communicative purpose of different academic speech events directly influence the reliance on previewing moves. Monologic Lectures seem to exhibit the highest normalized density across the corpus (3.56 per 10,000 tokens), as university lecturers may routinely employ forward-looking signposts (e.g., "In a moment, we will look at...", "I will return to this topic later...") to guide students through upcoming lecture content. It is obvious that noteworthy use is also sustained across formal presentation genres, including Seminar Presentations (3.06 per 10,000 tokens) and Conference Presentations (2.86 per 10,000 tokens), where presenters are expected to structure their ideas in advance to manage the overall logical flow. In contrast to this, previewing occurs

considerably less frequently in interactive speech events. Normalized rates appear to decline to 1.82 per 10,000 tokens in Seminar Discussions and 1.14 per 10,000 tokens in Lecture Discussions (only 7 raw occurrences) followed by Doctoral Defense Discussions. We argue that such low frequency of previewing in interactive and defense contexts in time pressure could be attributed to some key contextual and communicative factors. Since spontaneous discussions, defense examinations and conference discussion sessions are driven by real-time turn-taking, reactive questioning and co-constructed dialogue, speakers may well operate under high online processing demands with little structural necessity or opportunity to outline upcoming discourse segments in advance. It is important to note that in ELF communication or in more interactive and spontaneous settings such as seminars or conference discussions, speakers often employ alternative strategies instead of directly previewing content in a structured way. Instead of directly stating what they will discuss (e.g., 'I'm going to talk about...'), speakers may employ more flexible strategies to guide the conversation's flow. As needed, these may include reiterating key points, reformulating, or framing points like 'what I mean is...' and 'let me clarify that point...' to introduce or

explain ideas. These substitutes still guide the audience through the discourse, but in a more direct way. Thus, the findings suggest that Previewing is not simply a low-frequency metadiscursive resource; instead, its low-frequency use appears to be particularly sensitive to the structural and interactional affordances of different academic speech events.

Taken together with the findings for Reviewing in 4.5 above, we suggest an interesting asymmetry between retrospective and prospective flow management in academic ELF communication. Reviewing, although more frequent overall, is relatively dispersed across speech events. Previewing, by contrast, is less frequent overall but shows substantially stronger cross-event variation, with its highest normalized rates occurring in Lectures, Seminar Presentations and Conference Presentations. We argue that such a pattern suggests that prospective discourse management may be particularly sensitive to the degree of pre-structuring afforded by a speech event; speakers in lectures and presentations can anticipate and signal upcoming discourse, whereas the trajectory of interactive discussions is more contingent on emerging turns and contributions. Retrospective reference, by comparison, may represent a more broadly available discourse-management resource since academic ELF speakers could refer back to prior discourse regardless of whether the interaction has been extensively pre-planned.

4.7. Concluding the Topic Across Speech Events

The speaker employs the feature of concluding the topic to bring a topic to a close. Thus, its usage lets the audience know

that the main topic is changing or coming to an end. Speakers make it clear when a section is coming to an end to create a clear transition to the next or last phase of the discussion. For example, the speakers in the following examples employ MD markers, which signify closure or summary, to identify the end of the topic, such as ‘I’ll end there because of the time,’ ‘okay, so here are my concluding remarks,’ ‘I want to stress this point, erm, maybe as a conclusion to my presentation,’ ‘would like to make the final statements from the opponents’ side,’ ‘yes, and erm, now the last word, er, I would say that’ and ‘we finished paper; we are done.’

By making it more evident to the audience when a topic is coming to an end and getting them ready for a change in subject or closing remarks, these features aid in structuring the conversation. In particular, the use of phrases like ‘concluding remarks’ or ‘final statements’ indicates that the speaker is coming to an end, which strengthens the speech’s organizational flow.

(29) I’ll end there because of the time.
(Conf. Presentations 05J)

(30) okay so here are my concluding remarks erm (Conf. Presentations 06C)

(31) i want to stress this point, erm maybe as a conclusion to my presentation (Conf. Presentations 05J)

(32) would like to make the final statements from the opponents’ side. er
<READING ALOUD> the dissertation of <NAME S1> lactic acid based on hot melt adhesive (Doctoral defense disc 010I)

(33) yes and er now the last word er i would say that (Lecture discussions 010)

(34) we finished <NAME S4>’s paper we are done (Seminar discussions 04B)

Table 10. Distribution of Concluding Topic across speech events (raw frequency & normalized rate per 10,000 tokens)

	Doctoral Defense Disc.	Seminar Pres.	Lect.	Conf. Pres.	Lecture Disc.	Seminar Disc.	Conf. Disc.
Raw Frequency	5	26	5	16	9	6	7
Per 10,000 tokens	0.81	4.19	0.85	2.54	1.46	0.99	0.89

As shown in Table 10, the use of concluding the topic varies markedly according to speech event type. Even if concluding the topic is the least frequently employed metadiscursive feature overall across the corpus (averaging 1.66 instances per 10,000 tokens), it reaches its highest density in Seminar Presentations (4.19 per 10,000 tokens). A primary reason for this peak usage could be that student presenters in seminar settings may often need to mark explicit boundaries and provide clear transitions between distinct thematic sections. Conference Presentations also frequently employ topic closures (2.54 per 10,000 tokens), as structured, time-restricted presentations can require explicit topic wrap-ups before transitioning to the next slide or final conclusions. In interactive genres such as Lecture Discussions and Seminar Discussions, it often evolves organically from one comment to another without formal, pre-planned topic closures, reflecting the open-ended and turn-by-turn nature of spoken dialogue. Regarding lectures, tight time constraints and a primary pedagogical focus on introducing new material could often lead lecturers to move straight into incoming topics rather than formally closing preceding ones. Similar to this, in doctoral defenses, the discourse may be steered continuously by external examiner questions and ongoing critiques, reducing the opportunity or necessity for candidates to deliver formal topic-closing statements. As can be seen, one of the least used forms is found in doctoral

defense discussions. The communicative norms of ELF interactions may explain the generally low frequency of explicit topic-concluding markers across genres. In ELF contexts, speakers often favor mutual intelligibility and adaptability over strict adherence to native-like discourse conventions (Seidlhofer, 2011). Thus, rather than employing explicit linguistic cues to indicate the end of a topic, they may use implicit strategies or non-verbal signals, such as pausing, changing their body posture, or ending the screen share. These multimodal resources may serve as functional equivalents to explicit metadiscourse, especially in spontaneous or interactive settings where flexibility and shared understanding take priority over formality. This finding is consistent with Cogo and Dewey's (2012) argument that ELF communication is often characterized by pragmatic strategies adapted to the communicative situation rather than predetermined linguistic forms.

5. Conclusion

This study set out to investigate how speakers of English as a Lingua Franca manage the flow of spoken academic discourse across various speech events through the use of MD features. The results, based on a detailed analysis of 83 texts, including monological and dialogical event types from the ELFA corpus, support the role of metadiscursive strategies in organizing real-time academic communication. By normalizing frequency counts to occurrences

per 10,000 tokens and conducting weighted inferential Chi-Square, the findings demonstrate that metadiscourse deployment in ELF academic speech is genre-sensitive. To this end, the findings highlight the role of MD features, including enumeration, endophoric marking, contextualization, topic announcement, review, preview, and conclusion in managing the flow of academic discourse. Rather than operating as uniform discourse markers, metadiscourse features vary significantly according to the interactional structure and evaluative stakes of each speech event. Structured monologic presentation genres, particularly Seminar Presentations and Monologic Lectures, rely heavily on macro-structural topic announcements (up to 14.98 per 10,000 tokens) and forward-looking previewing (up to 4.30 per 10,000 tokens) to establish explicit communicative agendas for multilingual audiences. In contrast, we show that high-stakes evaluative settings such as Doctoral

Defense Discussions display a distinct hybrid profile, exhibiting the corpus-wide peak rate for Enumerating (29.75 per 10,000 tokens) alongside substantial densities of Endophoric Marking (20.85 per 10,000 tokens) and Contextualizing (19.49 per 10,000 tokens). These findings point to noticeable variations in reflexive signposting across event types, where structural signposting and interactive flexibility vary by genre, consistent with broader accounts of metadiscourse as sensitive to disciplinary, rhetorical, and communicative context (Jiang and Akbaş, 2024). While this distribution could align with the idea that speakers adapt their language to contextual demands, it is important to emphasize that multiple contextual factors, including genre expectations, participant roles, evaluative relationships, audience composition and disciplinary conventions, may also contribute to these observed patterns.

Table 11. Visual Summary: Metadiscourse Functions Across ELF Speech Events

Metadiscourse Function	Speech Events with Relatively High Normalized Rates	Primary Communicative Purpose
Enumerating	Doctoral Defenses, Seminar Presentations, Lectures	Structuring complex information; signaling logical sequence
Endophoric Marking	Seminar Presentations, Doctoral Defenses, Lectures	Referencing visual aids or earlier content within a discourse
Contextualizing	Doctoral Defenses, Lectures, Seminar Discussions	Providing background; supporting reflexive clarification
Previewing	Lectures, Seminar Presentations, Conference Presentations	Anticipating upcoming topics, guiding the audience's attention
Reviewing	Conference Presentations, Seminar Presentations, Lecture Discussions	Reinforcing or revisiting previously mentioned content; supporting retrospective discourse organization
Announcing the Topic	Seminar Presentations, Conference Presentations, Lectures	Marking clear topic shifts; setting discourse expectations
Concluding the Topic	Seminar Presentations, Conference Presentations	Signaling topic closure; organizing transitions

As shown in Table 11, the use of MD varies depending on the speech event, and the observed distributions appear to correspond to

differences in the communicative and interactional organization of these events. Lectures, seminars, and conference

presentations are examples of structured, monologic events that show relatively high rates of several sequencing and prospective resources, such as enumerating, previewing, and concluding the topic, to manage flow. It is particularly noteworthy that Doctoral Defense Discussions present a distinct hybrid profile within the ELF academic speech event continuum. Even though highly dialogic and interactive in nature, doctoral defenses share and, in some respects, exceed the structural signposting typical of monologic presentations. In interactive events, contextualizing appears particularly salient in some settings, while the distribution of reviewing is more variable across event types. Here, we can refer to the contrast between Previewing and Reviewing, as an example. Although Reviewing is more frequent overall (129 occurrences; 6.76%) than Previewing (82 occurrences; 4.30%), Reviewing shows comparatively weaker cross-event variation. Previewing, by contrast, is less frequent overall but shows a cross-event differentiation, $\chi^2(6) = 35.24$, $p < .001$, $w = .656$, with its highest normalized rates occurring in Lectures and presentation genres. We can argue that such a pattern suggests that prospective discourse management may be particularly sensitive to the degree of pre-structuring afforded by a speech event; speakers in lectures and presentations can anticipate and signal upcoming discourse, whereas the trajectory of interactive discussions is more contingent on emerging turns and contributions. The contrasting distributions of previewing and contextualizing across speech events suggest that ELF speakers draw on different discourse-organizing resources depending on the structural and interactional characteristics of the event, a pattern broadly consistent with previous observations on reflexive discourse organization in academic ELF communication (Akbaş and Hatipoğlu, 2018; Ädel, 2022; Jiang and Akbaş, 2024; Mauranen, 2007).

It is interesting to note that doctoral defenses present a hybrid profile showing a noteworthy distinction in how metadiscourse

is employed. Although they are interactive in nature, they share some features with structured events, such as frequent use of endophoric marking and enumeration; however, they stand out in that they rarely employ topic announcements or closures. Instead, speakers frequently contextualize their responses, a pattern that may reflect the explanatory and evaluative demands of this genre. The interactionally contingent nature of doctoral-defense discourse in ELF settings has also been demonstrated by Çiftçi and Akbaş (2021), whose analysis of ELFA doctoral-defense discussions showed how speakers deploy linguistic resources dynamically to manage interpersonal and face-related demands during unfolding interaction. This genre-specific distribution supports the view that metadiscourse use can be sensitive to the communicative organization of particular academic speech events. Thus, this result supports the genre-sensitive claims made by Bernad-Mechó and Fortanet-Gómez (2019), demonstrating that certain academic occasions, such as defenses, may simultaneously combine formal, argumentative and interactive discourse demands. This type of hybridity can be consistent with the view that academic speech events may draw on overlapping discourse configurations rather than following straightforwardly to a binary distinction between monologic and dialogic genres.

Furthermore, the findings illustrate the functions of MD features in managing real-time spoken academic interaction. Metadiscourse features, which serve the functions of enumerating, endophoric marking, contextualizing, and announcing the topic, occurred more frequently in the current data, aligning with earlier work (Ädel, 2022; Molino, 2018), which points to the importance of these discourse organizers as central to the management of sequentiality and referential clarity. Among these, enumerating, in particular, may help break down complex content and present information in a logical sequence of steps. Similarly, contextualizing and announcing the

topic can guide audience understanding by signaling structure, intent, or transitions, while endophoric marking connects different parts of discourse to enhance flow. In genres of academic speech where linear argumentation and content density are valued, these features may be especially prominent. We argue that their role is both organizational and potentially facilitative for audience orientation, particularly in ELF communication, as they provide listeners with cues that may assist the processing of discourse in linguistically diverse academic contexts.

These findings also allow the metadiscourse use in ELF academic discourse to be considered in relation to broader accounts of ELF communication. Previous research has characterized ELF interaction in terms of adaptability, mutual intelligibility as well as accommodation (Seidlhofer, 2011), and some of the present distributional patterns are compatible with such an account. However, our findings point more directly to variation associated with speech-event structure/types. In structured monologic events, such as lectures and presentations, sequencing and prospective functions (i.e., enumerating and previewing) appear to support linear organization and audience orientation. In contrast, more interactive events appear to show greater reliance on locally responsive resources such as contextualizing, although Reviewing is more broadly distributed and does not show robust cross-event differentiation under the Bonferroni-adjusted criterion. Hence, our findings suggest that the balance between prospective and retrospective discourse management varies with the structural and interactional affordances of the speech event. Moreover, it is likely that factors such as the speakers' L1 background, disciplinary culture, power dynamics (as seen in doctoral defenses), and audience composition (e.g., mixed proficiency levels or diverse academic backgrounds) further influence how MD features are deployed. These variables offer

fruitful directions for future research to build on the present findings.

The results of this study have significant implications for academic communication and teaching strategies, particularly in the context of English for Academic Purposes (EAP) classes. Our findings suggest that instruction in spoken academic metadiscourse could move beyond teaching individual markers to raising learners' awareness of how discourse-management resources could vary according to communicative settings and purposes. For example, learners preparing for lectures or formal presentations may benefit from explicit attention to prospective resources such as previewing and topic announcement, whereas preparation for more interactive academic events could foreground locally responsive strategies such as contextualizing and retrospective reference. Educators can create role-playing exercises in EAP classes, for example, where students act as academic presenters or engage in discussions, focusing on the strategic use of these MD features. We believe that such activities could help learners develop a repertoire of discourse-management strategies that can be adapted to different academic speaking contexts rather than treating metadiscourse as a uniform set of expressions. The study's findings are significant not only for EAP courses but also for researchers attending international conferences, conducting research, and teaching. By becoming familiar with MD features, they can enhance their ability to manage discourse flow effectively.

6. Limitations and Directions for Future Research

While this study provides new empirical insights into reflexive metadiscourse across spoken ELF academic speech events, several methodological and contextual limitations should be acknowledged. The ELFA corpus was compiled primarily within Finnish university settings, with Finnish L1 speakers constituting 28.5% of the total participant pool. While ELFA represents authentic, highly international academic interactions,

local institutional norms or dominant L1 transfer effects may influence spoken metadiscourse patterns. Replicating this study across ELF speech corpora from different geographic regions (e.g., Asian or Southern European higher education contexts) would help determine the cross-cultural generalizability of these findings. In addition, because the ELFA corpus consists of orthographic transcripts derived from audio recordings, this study focused exclusively on verbal metadiscourse features. While verbal references to visual aids (e.g., slides, handouts) were captured within categories such as endophoric marking, non-verbal visual resources (such as bodily gestures, gaze, and slide transitions) could not be systematically annotated. Future research utilizing video-recorded spoken corpora could adopt a multimodal framework to examine how non-verbal cues co-occur with verbal metadiscourse in spoken ELF contexts. Even though frequency counts were normalized to a standardized metric of occurrences per 10,000 tokens to correct for word-count variations, the number of individual speech events varies across genres in the ELFA corpus (ranging from 4 doctoral defense discussions to 21 conference presentations). Future research utilizing larger, more evenly balanced corpus samples could further examine genre-specific density. An additional statistical limitation concerns data nesting, as the 1909 metadiscourse instances are clustered within 83 speech-event transcripts with unequal distributions across genres (i.e., 4 doctoral defenses vs. 21 conference presentations). While token-level Chi-Square tests could offer a standard corpus-linguistic baseline (McEnery et al., 2006), treating individual tokens as strictly independent observations does not account for transcript-level clustering. We thus suggest future studies utilizing hierarchical linear modeling or mixed-effects regression on larger, balanced transcript samples would be valuable to isolate speaker-level as well as transcript-level variance.

Furthermore, a key methodological limitation of observational corpus-based frequency analysis is that variations across speech events establish structural patterns rather than proving underlying functional or cognitive motivations. While we could interpret these variations as potential functional responses to interactional settings, other contextual variables, such as disciplinary traditions, participant power dynamics, audience familiarity and individual speaker styles, may also influence metadiscourse density/frequency. Therefore, future research incorporating qualitative participant interviews, retrospective recall protocols, or multi-variate modeling would help further untangle such contributing factors.

Finally, a theoretical limitation concerns the granularity of the coding unit. This study adopted a fine-grained, micro-move unit of analysis (annotating individual metadiscursive expressions) rather than a broad macro-segment approach (coding extended discourse stretches as single moves). While micro-unit tagging reflects the real-time, iterative nature of spoken ELF discourse, where reformulations and successive signposts are frequent, it inherently results in higher raw frequency counts than macro-segmentation would yield. Because the micro-unit coding protocol was applied consistently across all subcorpora, cross-genre relative comparisons remain valid. However, research comparing micro-level expression counts against macro-level segment boundaries could offer additional new insights into metadiscoursal granularity in academic speech.

Acknowledgments: This article is derived from the first author's unpublished thesis under the supervision of the second author, submitted in 2025. The content has been revised and extended for the purposes of this article. We would also like to thank the Proofreading & Editing Office of the Dean for Research at Erciyes University for additional copyediting and the proofreading services that assisted with this manuscript.

Statement on the Use of Generative AI:

During the preparation of this manuscript, the authors used an AI-based language editing tool to assist with proofreading and improving the clarity and linguistic quality of the manuscript. The tool was used solely for language refinement and did not contribute to the conceptualisation, analysis, or interpretation of the data. All content, arguments, and conclusions remain the responsibility of the authors (Gemini, 2.0 & 2.5, <https://gemini.google.com>). We also acknowledge that during the preparation of revisions, the author(s) used Grammarly, ChatGPT 5.6 and Gemini 3.7 only to improve the readability and grammatical accuracy of the manuscript. The author reviewed and verified the final content and takes full responsibility for the accuracy, interpretation and conclusions presented in the manuscript.

References

- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*, John Benjamins Publishing Company. <https://doi.org/10.1075/scl.24> (In English)
- Ädel, A. (2010). Just to give you kind of a map of where we are going: A taxonomy of metadiscourse in spoken and written academic English. *Nordic Journal of English Studies*, 9 (2), 69–97. <https://doi.org/10.35360/njes.218> (In English)
- Ädel, A. (2012). “What I want you to remember is...”: Audience orientation in monologic academic discourse, *English Text Construction*, 5 (1), 69–97. <https://doi.org/10.1075/etc.5.1.06ade> (In English)
- Ädel, A. (2022). Adopting a ‘move’ rather than a ‘marker’ approach to metadiscourse: A taxonomy for spoken student presentations, *English for Specific Purposes*, 69, 4–18. <https://doi.org/10.1016/j.esp.2022.09.001> (In English)
- Ädel, A. and Mauranen, A. (2010). Metadiscourse: Diverse and divided perspectives, *Nordic Journal of English Studies*, 9 (2), 1–11. <https://doi.org/10.35360/njes.215> (In English)
- Aertselaer, J. A. and Dafouz-Milne, E. (2008). Argumentation patterns in different languages: An analysis of metadiscourse markers in English and Spanish texts, In *Contrastive rhetoric: Approaches to contrastive linguistics*, De Gruyter, 87–102. <https://doi.org/10.1515/9783110207217.1.87> (In English)
- Akbaş, E. (2012). Exploring metadiscourse in master’s dissertation abstracts: Cultural and linguistic variations across postgraduate writers, *International Journal of Applied Linguistics and English Literature*, 1 (1), 12–21. <https://doi.org/10.7575/ijalel.v.1n.1p.12> (In English)
- Akbaş, E. and Hatipoğlu, Ç. (2018). Metadiscourse variations across academic genres: Rhetorical preferences in textual and interpersonal markers, *Educational Sciences: Theory & Practice*, 18 (4), 767–775. (In English)
- Akbaş, E., Hatipoğlu, Ç. and Bayyurt, Y. (2017). *Metadiscourse in written genres: Uncovering textual and interactional aspects of texts*, Peter Lang. (In English)
- Bell, A. (1984). Language style as audience design, *Language in Society*, 13 (2), 145–204. <https://doi.org/10.1017/S004740450001037X> (In English)
- Bernad-Mechó, E. (2024). Building successful communication in EMI contexts: A multimodal approach to organizational metadiscourse in intercultural lectures, *Chinese Journal of Applied Linguistics*, 47 (2), 259–280. <https://doi.org/10.1515/CJAL-2024-0206> (In English)
- Bernad-Mechó, E. and Fortanet-Gómez, I. (2019). Organizational metadiscourse across lecturing styles: Engagement beyond language. In *Engaging in the language learning process*, John Benjamins, 301–317. <https://doi.org/10.1075/pbns.301.17ber> (In English)
- Björkman, B. (2009). From code to discourse in spoken ELF, in A. Mauranen (Ed.), *English as a lingua franca: Studies and findings*, Cambridge Scholars Publishing, 225–254. (In English)
- Björkman, B. (2011). English as a lingua franca in higher education: Implications for EAP, *Iberica*, 22, 79–100. (In English)
- Bondi, M. (2010). Metadiscursive practices in introductions: Phraseology and semantic sequences across genres, *Nordic Journal of English Studies*, 9 (2), 99–123. <https://doi.org/10.35360/njes.219> (In English)
- Bouziri, B. (2020). Topic signaling metadiscourse devices in the Tunisian lecture corpus. *ESP Today*, 8 (2), 227–249.

<https://doi.org/10.18485/esptoday.2020.8.2.3> (In English)

Broggini, S. and Murphy, A. C. (2017). Metadiscourse in EMI lectures: Reflections on a small corpus of spoken academic discourse. *L'analisi linguistica e letteraria*, 25 (2 Supplement). (In English)

Cogo, A. and Dewey, M. (2012). *Analysing English as a lingua franca: A corpus-driven investigation*, Continuum. (In English)

Crismore, A., Markkanen, R. and Steffensen, M. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students, *Written Communication*, 10 (1), 39–71. <https://doi.org/10.1177/0741088393010001002> (In English)

Çiftçi, H. and Akbaş, E. (2021). Stancetaking in spoken ELF discourse in academic settings: interpersonal functions of I don't know as a face-maintaining strategy, *Eurasian Journal of Applied Linguistics*, 7 (1), 484–502. <https://doi.org/10.32601/ejal.911499> (In English)

Dafouz, E. and Perucha, B. (2010). Metadiscursive devices in university lectures, in *Academic and professional discourse genres in Spanish*, John Benjamins, 193–210. <https://doi.org/10.1075/aals.7.11daf> (In English)

Dahl, T. (2004). Textual metadiscourse in research articles: A marker of national culture or of academic discipline?, *Journal of Pragmatics*, 36 (10), 1807–1825. <https://doi.org/10.1016/j.pragma.2004.05.004> (In English)

Dewey, M. (2009). English as a lingua franca: Heightened variability and theoretical implications, in A. Mauranen (Ed.), *English as a lingua franca: Studies and findings*, Cambridge Scholars Publishing, 60–83. (In English)

Doiz, A. and Lasagabaster, D. (2022). Metadiscourse in English-medium instruction: Teachers' reflections and implementation, *System*, 105, 102738. <https://doi.org/10.1016/j.system.2022.102738> (In English)

Gnutzmann, C. and Intemann, F. (2008). Introduction: The globalisation of English. Language, politics and the English language classroom, in C. Gnutzmann and F. Intemann (Eds.), *The globalisation of English and the English language classroom*, Narr, 9–24. (In English)

Graddol, D. (2006). *English next: Why global English might mean the end of 'English as a foreign language'*, The English Company (UK) Ltd. British Council, available at: <http://www.britishcouncil.org/learning-research-english-next.pdf> (Accessed 15.09.2026) (In English)

House, J. (1996). Developing pragmatic fluency in English as a foreign language: Routines and metapragmatic awareness. *Studies in Second Language Acquisition*, 18 (2), 225–252. <https://doi.org/10.1017/S0272263100014893> (In English)

House, J. (2010). The pragmatics of English as a lingua franca. In A. Trosborg (Ed.), *Handbook of pragmatics*, De Gruyter, 363–387. (In English)

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*, Continuum. (In English)

Hyland, K. and Jiang, F. (2018). "In this paper we suggest": Changing patterns of disciplinary metadiscourse. *English for Specific Purposes*, 51, 18–30. <https://doi.org/10.1016/j.esp.2018.02.001> (In English)

Hyland, K. and Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, 25 (2), 156–177. <https://doi.org/10.1093/applin/25.2.156> (In English)

Jenkins, J. (2009). (Un)pleasant? (In)correct? (Un)intelligible? ELF speakers' perceptions of their accents, in A. Mauranen (Ed.), *English as a lingua franca: Studies and findings*, Cambridge Scholars Publishing, 254–278. (In English)

Jenkins, J. (2013). *English as a lingua franca in the international university: The politics of academic English language policy*, Routledge. <https://doi.org/10.4324/9780203798157> (In English)

Jiang, F. and Akbaş, E. (2024). Metadiscourse in a disciplinary context: An overview, *Chinese Journal of Applied Linguistics*, 47 (2), 163–177. (In English)

Jokić, N. (2017). Metadiscourse in ELF spoken discourse of Erasmus students in Austria. *AAA: Arbeiten aus Anglistik und Amerikanistik*, 42 (2), 211–223. <http://www.jstor.org/stable/26379472> (In English)

Kashiha, H. (2022). Academic lectures versus political speeches: Metadiscourse functions

affected by the role of the audience, *Journal of Pragmatics*, 190, 60–72. <https://doi.org/10.1016/j.pragma.2022.01.003> (In English)

Kaur, J. (2009). Pre-empting problems of understanding in English as a lingua franca, in A. Mauranen (Ed.), *English as a lingua franca: Studies and findings*, Cambridge Scholars Publishing, 107–123. (In English)

Kawase, T. (2015). Metadiscourse in the introductions of PhD theses and research articles. *Journal of English for Academic Purposes*, 20, 114–124. <https://doi.org/10.1016/j.jeap.2015.08.004> (In English)

Landis, J. R. and Koch, G. G. (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers, *Biometrics*, 363–374. (In English)

Lee, J. J. and Subtirelu, N. (2015). Metadiscourse in the classroom: A comparative analysis of EAP lessons and university lectures, *English for Specific Purposes*, 37, 52–62. <https://doi.org/10.1016/j.esp.2014.08.005> (In English)

Luukka, R. (1994). Metadiscourse in academic texts, in B. L. Gunnarsson, P. Linell and B. Nordberg (Eds.), *Text and talk in professional contexts*, ASLA, 77–88. (In English)

Magnucz, Á. (2012). Are you with me? A metadiscursive analysis of interactive strategies in college students' course presentations, *International Journal of English Studies*, 12 (1), 55–78. <https://doi.org/10.6018/ijes.12.1.118281> (In English)

Mauranen, A. (1993). Contrastive ESP rhetoric: Metatext in Finnish-English economic text, *English for Specific Purposes*, 12 (1), 3–22. [https://doi.org/10.1016/0889-4906\(93\)90024-I](https://doi.org/10.1016/0889-4906(93)90024-I) (In English)

Mauranen, A. (2001). Reflexive academic talk: Observations from MICASE. In *Corpus linguistics in North America: Selections from the 1999 symposium*, University of Michigan Press, 165–178. (In English)

Mauranen, A. (2005). English as lingua franca: An unknown language?, *Nordic Journal of English Studies*, 4 (2), 269–293. (In English)

Mauranen, A. (2006a). A rich domain of ELF: The ELFA corpus of academic discourse, *Nordic Journal of English Studies*, 5 (2), 145–159. <https://doi.org/10.35360/njes.15> (In English)

Mauranen, A. (2006b). Signaling and preventing misunderstanding in English as lingua franca communication, *International Journal of the Sociology of Language*, 2006 (177), 123–150. <https://doi.org/10.1515/IJSL.2006.008> (In English)

Mauranen, A. (2007). Hybrid voices: English as the lingua franca of academics, in K. Fløttum (Ed.), *Language and discipline perspectives on academic discourse*, Cambridge Scholars Publishing, 243–257. (In English)

Mauranen, A. (2009). Spoken rhetoric, in *Rhetoric in detail*, John Benjamins, 275–298. <https://doi.org/10.1075/pbns.193.13mau> (In English)

Mauranen, A. (2010). Discourse reflexivity: A discourse universal? The case of ELF, *Nordic Journal of English Studies*, 9 (2), 13–40. <https://doi.org/10.35360/njes.216> (In English)

Mauranen, A., Hynninen, N. and Ranta, E. (2020). English as an academic lingua franca, in J. Jenkins, W. Baker and M. Dewey (Eds.), *The Routledge handbook of English as a lingua franca*, Routledge, 200–213. (In English)

Mauranen, A. and Pérez-Llantada, C. and Swales, J. (2010). Academic Englishes: A standardised knowledge?, in *English as an academic lingua franca*, De Gruyter, 19–34. (In English)

Mauranen, A. and Ranta, E. (2008). English as an academic lingua franca, *Nordic Journal of English Studies*, 7 (3), 1–10. (In English)

McArthur, T. (2002). *The Oxford guide to world English*, Oxford University Press. (In English)

McEnery, T., Xiao, R. and Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*, Routledge. (In English)

McHugh, M. L. (2012). Interrater reliability: the kappa statistic, *Biochemia medica*, 22 (3), 276–282. (In English)

Molino, A. (2018). “What I’m speaking is almost English...”: A corpus-based study of metadiscourse in English-medium lectures at an Italian university, *Educational Sciences: Theory and Practice*, 18 (4), 1–15. <https://doi.org/10.12738/estp.2018.4.0330> (In English)

Mur-Dueñas, P. (2011). An intercultural analysis of metadiscourse features in research articles written in English and in Spanish, *Journal of Pragmatics*, 43 (12), 3068–3079.

<https://doi.org/10.1016/j.pragma.2011.05.002> (In English)

O'Donnell, M. (2008). The UAM CorpusTool: Software for corpus annotation and exploration. In C. M. Bretones Callejas et al. (Eds.), *Applied linguistics now: Understanding language and mind*, Universidad de Almería, 1433–1447. (In English)

Penz, H. (2017). Metadiscourse in multi-party conflict talk, *AAA: Arbeiten aus Anglistik und Amerikanistik*, 42 (2), 301–323. <http://www.jstor.org/stable/26379476> (In English)

Pérez-Llantada, C. (2010). The discourse functions of metadiscourse in published academic writing: Issues of culture and language, *Nordic Journal of English Studies*, 9 (2), 41–68. (In English)

Pickering, L. (2006). Current research on intelligibility in English as a lingua franca, *Annual Review of Applied Linguistics*, 26, 219–233. <https://doi.org/10.1017/S0267190506000110> (In English)

Rowley-Jolivet, E. and Carter-Thomas, S. (2005). Genre awareness and rhetorical appropriacy: Manipulation of information structure by NS and NNS scientists in the international conference setting, *English for Specific Purposes*, 24 (1), 41–64. <https://doi.org/10.1016/j.esp.2003.09.003> (In English)

Rubio, M. M. (2011). A pragmatic approach to the macro-structure and metadiscoursal features of research article introductions in the field of agricultural sciences, *English for Specific Purposes*, 30 (4), 258–271. <https://doi.org/10.1016/j.esp.2011.04.002> (In English)

Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca, *Annual Review of Applied Linguistics*, 24, 209–239. <https://doi.org/10.1017/S0267190504000145> (In English)

Seidlhofer, B. (2011). *Understanding English as a lingua franca*, Oxford University Press.

Sifakis, N. (2014). ELF awareness as an opportunity for change: A transformative perspective for ESOL teacher education, *Journal of English as a Lingua Franca*, 3 (2), 317–335. <https://doi.org/10.1515/jelf-2014-0019> (In English)

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*, Cambridge University Press. (In English)

Thompson, S. (2003). Text-structuring metadiscourse, intonation and the signalling of organisation in academic lectures, *Journal of English for Academic Purposes*, 2 (1), 5–20. [https://doi.org/10.1016/S1475-1585\(02\)00036-X](https://doi.org/10.1016/S1475-1585(02)00036-X) (In English)

Vande Kopple, W. J. (1985). Some explanatory discourse on metadiscourse, *College Composition and Communication*, 36 (1), 82–93.

Vettorel, P. (2018). ELF and communication strategies: Are they taken into account in ELT materials? *RELC Journal*, 49 (1), 58–73.

<https://doi.org/10.1177/0033688217746204> (In English)

Warren, M. (2006). *Features of naturalness in conversation*, John Benjamins. <https://doi.org/10.1075/pbns.152> (In English)

Zare, J. and Tavakoli, M. (2015). Frequency and diversity of personal metadiscourse across monologic and dialogic modes of academic speech. *Journal of English for Academic Purposes*, 20, 1–12. <https://doi.org/10.1016/j.jeap.2015.05.003> (In English)

Zhu, Y. (2018). An intercultural analysis of personal metadiscourse in English and Chinese commencement speeches. *Advances in Language and Literary Studies*, 9 (5), 100–110. <https://doi.org/10.7575/aiall.v.9n.5p.100> (In English)

All authors have read and approved the final manuscript.

Все 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.

Leyla Can, Postgraduate Student, Institute of Education, Erciyes University, Kayseri, Turkey.

Erdem Akbas, Associate Professor, Faculty of Education, Erciyes University, Kayseri, Turkey.

UDC 811,111'37

DOI: 10.18413/2313-8912-2026-12-3-0-5

Madina M. Zhemileva¹
Quero Gervilla E. F.²

**Cognitive-semantic mechanisms of collective
conceptualization in English animal community
nominations: a corpus-based analysis**

¹ Al-Farabi Kazakh National University,
71 Al-Farabi Ave., Almaty, 050040, Kazakhstan
E-mail: zhemilevamm@gmail.com
ORCID: 0009-0005-2627-9681

² University of Granada,
1 Cuesta del Hospicio, Granada, 18071, Spain
E-mail: efquero@ugr.es
ORCID: 0000-0002-1240-8121

Received 09 September 2026; accepted 15 September 2026; published 30 September 2026

Abstract: English animal group nouns not only denote groups of animals but also reflect different ways of construing multiple entities as a unified whole, making their role in the conceptualization of collectivity a relevant object of cognitive-semantic research. Although previous studies have examined collective nouns, cognitive profiling, metaphorical extension, and evaluative meaning from different theoretical perspectives, the interaction of these mechanisms in the use of animal group nouns remains insufficiently explored. The present study focuses on six such lexical items – *herd*, *pack*, *flock*, *swarm*, *pride*, and *colony* – and aims to identify the cognitive-semantic mechanisms underlying their literal and metaphorically extended uses. The data were retrieved from the Corpus of Contemporary American English (COCA) for the period 2010 – 2019. Of the initial 2,563 contexts, 1,804 were retained after manual screening: 773 were classified as literal and 1,031 as metaphorically extended. The methodology involved distinguishing between literal and metaphorically extended uses in accordance with the principles of MIPVU, identifying cognitive profiling parameters, inductively generalizing these parameters into higher-order construal types, and conducting a contextual analysis of evaluative potential. The results show that metaphorically extended uses project selected characteristics of animal group organization onto social, technological, and abstract phenomena. The identified profiling parameters were grouped into four higher-order construal types: Agentive, Passive, Dynamic, and Structural. The analysis suggests that the evaluative tendencies associated with these types emerge through the interaction between the profiled characteristics and the immediate discourse context rather than constituting fixed components of lexical meaning. The study is limited to six selected nominations and COCA data from 2010 – 2019; therefore, the proposed typology should be regarded as an analytical generalization rather than an exhaustive classification. The findings indicate an interrelationship between uniplexing, cognitive profiling, metaphorical extension, and evaluative potential in the conceptualization of collectivity.

Keywords: Animal group nouns; Conceptualization of collectivity; Lexical-semantic field; Uniplexing; Cognitive profiling; Construal types; Metaphorical extension; Evaluative potential

How to cite: Zhemileva, M. M., Quero Gervilla, E. F. (2026). Cognitive-semantic mechanisms of collective conceptualization in English animal community nominations: a corpus-based analysis, *Research Result. Theoretical and Applied Linguistics*, 12 (3), 113–144. DOI: 10.18413/2313-8912-2026-12-3-0-5

UDC 81'42:004.8

DOI: 10.18413/2313-8912-2026-12-3-0-5

Жемилева М. М.¹
Киро Хервилья Э. Ф.²

**Когнитивно-семантические механизмы
концептуализации коллективности в английских
номинациях сообществ животных: корпусный
анализ**

¹ Казахский национальный университет имени Аль-Фараби
ул. Аль-Фараби, 71, Алматы, 050040, Казахстан
E-mail: zhemilevamm@gmail.com
ORCID: 0009-0005-2627-9681

² Университет Гранады
ул. Куеста дель Осписио, 1, Гранада, 18071, Испания
E-mail: efquero@ugr.es
ORCID: 0000-0002-1240-8121

Статья поступила 09 сентября 2026 г.; принята 15 сентября 2026 г.;
опубликована 30 сентября 2026 г.

Аннотация: Номинации сообществ животных не только обозначают множества животных, но и отражают различные способы их осмысления как единого целого, что обуславливает актуальность изучения их роли в концептуализации коллективности. Несмотря на наличие исследования коллективных существительных, когнитивного профилирования, метафорического расширения и оценочной семантики, взаимодействие этих механизмов в употреблении специализированных номинаций сообществ животных остаётся недостаточно изученным. Цель исследования заключается в выявлении когнитивно-семантических механизмов, лежащих в основе буквальных и метафорически расширенных употреблений английских номинаций *herd, pack, flock, swarm, pride* и *colony*. Материал исследования составили контексты, извлечённые из Корпуса современного американского английского языка (Corpus of Contemporary American English. COCA) за 2010 – 2019 гг. Из первоначально отобранных 2 563 контекстов после ручной фильтрации в итоговый корпус вошли 1 804 употребления, из которых 773 были классифицированы как буквальные, а 1 031 – как метафорически расширенные. Методика включала разграничение буквальных и метафорически расширенных употреблений с опорой на принципы MIPVU, выявление параметров когнитивного профилирования, их индуктивное обобщение в типы конструирования и контекстуальный анализ оценочного потенциала. Результаты показали, что при метафорическом расширении отдельные характеристики групповой организации животных переносятся на социальные технологические и абстрактные явления. Выявленные параметры профилирования были объединены в четыре типа конструирования: агентивные,

пассивный, динамический и структурный. Результаты анализа позволяют предположить, что связанные с ними оценочные тенденции возникают в результате взаимодействия профилируемых характеристик и контекста, а не являются неизменными компонентами значения номинаций. Исследование ограничено шестью выбранными номинациями и данными СОСА за 2010 – 2019 гг.; поэтому предложенную типологию следует рассматривать как аналитическое обобщение, а не как исчерпывающую классификацию. Полученные данные указывают на взаимосвязь униплексного конструирования, когнитивного профилирования, метафорического расширения и оценочного потенциала в концептуализации коллективности.

Ключевые слова: Номинации сообществ животных; Концептуализация коллективности; Лексико-семантическое поле; Униплексное конструирование; Когнитивное профилирование; Типы конструирования; Метафорическое расширение; Оценочный потенциал

Информация для цитирования: Жемилева М. М., Керо Хервилья Э. Ф. Когнитивно-семантические механизмы концептуализации коллективности в английских номинациях сообществ животных: корпусный анализ // Научный результат. Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 3. С. 113–144. DOI: 10.18413/2313-8912-2026-12-3-0-5

Introduction

Language serves as a primary means of conceptualizing and categorizing experience, a view widely shared within cognitive linguistics (Langacker, 1987; Lakoff, 1987). From this perspective, the study of lexical-semantic fields (LSFs) reveals the principles underlying the systemic organization of linguistic knowledge. Of particular interest in this regard are what are referred to in the present study as animal community nominations. For the purpose of the present study, the expression animal community nominations is used as a working umbrella term for lexical units that designate animal collectives. It is not intended to replace established labels such as collective nouns, animal group nouns, or terms of venery, but it serves as a higher-order analytical designation for the broader domain of animal-group naming. The present article examines only a conventional lexical subset of this domain, namely *herd*, *pack*, *flock*, *swarm*, *colony*, and *pride*. Accordingly, the findings are interpreted with reference to this selected subset rather than to the full inventory of animal community nominations in English. These units not only denote forms of plurality but also represent cognitive models of group

organization that shape specific models of interpretation. Within the framework of cognitive linguistics, such terms function as means of conceptualizing collectivity through the cognitive operations of structuring (“carving up”) and construal (Talmy, 2000). A central mechanism in this process is uniplexing, which enables a plurality of individuals to be construed as a unified Gestalt (Talmy, 2000). Consequently, the analysis of these units facilitates the identification of mechanisms underlying the cognitive profiling of collectivity within the investigated lexical-semantic field. This interpretation is consistent with cognitive-semantic approaches that regard meaning as the result of conceptual profiling and the selective structuring of knowledge (Boldyrev, 2014).

Drawing on prototype theory (Rosch, 1978; Lakoff, 1987; Taylor, 2003; Taylor, 2018), this lexical-semantic field exhibits a center-periphery organization. Literal meanings anchored in the biological domain constitute the category’s core, whereas the periphery comprises metaphorically extended uses related to social, technological, and other conceptual domains. This radial organization reflects the gradient structure of the category

while enabling cognitive characteristics associated with collectivity to extend across conceptual domains through cross-domain conceptual mapping (Lakoff & Johnson, 2003). The historical roots of these nominations can be traced to the tradition of *terms of ventry*, documented in *The Book of St. Albans* (1486), which records numerous specialized designations for groups of animals (Lipton, 1977). Although many historical terms have fallen out of use, several have remained part of contemporary English and, as Crystal (2011) observes, continue to serve as productive resources for conceptualizing collective entities beyond the biological domain. Within this broader domain, the present study focuses specifically on six conventional lexical items – *herd*, *pack*, *flock*, *swarm*, *colony*, and *pride* – which provide distinct lexical means of construing animal collectivity. Following Geeraerts's understanding of meaning as encyclopedic knowledge (Geeraerts, 2009), these terms can function as cognitive instruments for conceptualizing both biological and social realities.

Previous studies have provided important insights into individual aspects of collective conceptualization. Research on collective nouns has explained their grammatical and semantic properties through mechanisms such as uniplexing and conceptual unity (Talmy, 2000; Levin, 2001; de Vries, 2021). Studies grounded in Cognitive Grammar have emphasized the role of profiling in shaping different construals of collectivity (Langacker, 1987; Joosten et al., 2007). Other research has examined metaphorical extension and evaluative meaning in discourse, demonstrating how animal-based metaphors contribute to the representation and evaluation of social actors (Musolff, 2016; Hart, 2014). However, these approaches have largely evolved independently and therefore do not explain how these cognitive mechanisms interact within a unified model of collective conceptualization. Consequently, the relationship between uniplexing, cognitive

profiling, metaphorical extension, and evaluative potential remains insufficiently understood.

The present study addresses this gap by integrating uniplexing, cognitive profiling, metaphorical extension, and evaluative potential within a unified cognitive-semantic framework for the analysis of animal community nominations. Based on contemporary evidence from the Corpus of Contemporary American English (COCA), the present analysis investigates how cognitive-semantic mechanisms shape the conceptualization of collectivity in English animal community nominations. To this end, the study pursues three objectives: (1) to delineate the hierarchical organization of the lexical-semantic field of animal community nominations by distinguishing between the Literal Core and the Metaphorical Periphery; (2) to examine how uniplexing and cognitive profiling contribute to the construal of collective entities; and (3) to explore how metaphorically extended uses are associated with recurrent evaluative tendencies in contemporary English discourse.

Accordingly, the study addresses the following research questions:

RQ1. How are the literal and metaphorically extended uses of the selected animal community nominations organized within the proposed Literal Core – Metaphorical Periphery model?

RQ2. How do recurrent cognitive profiling parameters – agency, homogeneity, synchrony, density/intensity, hierarchy, and functional integration – generalize inductively into higher-order construal types?

RQ3. How are the identified higher-order construal types associated with recurrent evaluative tendencies in metaphorically extended uses in contemporary English discourse?

The principal theoretical contribution of the present study lies in the development of an integrated cognitive-semantic framework that explains how uniplexing, cognitive profiling, metaphorical extension, and evaluative potential interact in the

conceptualization of collective entities. Application of the proposed framework to the corpus resulted in the development of a four-type analytical typology (Agentive, Passive, Dynamic, and Structural), which captures the dominant profiling tendencies observed across the analyzed corpus. Rather than constituting an exhaustive classification of collective construals, the proposed typology represents an analytical generalization of the recurrent profiling patterns identified in the present dataset.

Theoretical and Methodological Framework

The present study builds on these theoretical perspectives to propose an integrated cognitive-semantic framework for the analysis of animal community nominations.

Within the framework of cognitive linguistics, animal community nominations function as cognitive instruments that structure the conceptualization of collectivity and govern specific modes of semantic construal. Consistent with the established cognitive-semantic approaches to the lexicon (e.g., Langacker, 1987; Evans & Green, 2006), these nominations are viewed not merely as lexical labels but as cognitively structured lexical units that activate organized conceptual knowledge and support the interpretation of collective entities. Central to this perspective is the notion of construal, which refers to the ability of speakers to conceptualize the same situation in alternative ways (Langacker, 1987; Croft and Cruse, 2004). Linguistic meaning is therefore understood not as a direct reflection of objective reality but as the result of selective conceptualization and cognitive organization. This interpretation is consistent with cognitive-semantic approaches that view meaning construction as a process of conceptualization rather than the retrieval of fixed lexical content (Evans & Green, 2006). From this perspective, these animal units constitute a lexical-semantic field (LSF) organized around a center-periphery architecture, consistent with prototype theory

(Rosch, 1978; Lakoff, 1987; Taylor, 2003; Taylor, 2018) and with psycholinguistic research suggesting that lexical-semantic fields are cognitively organized around core and peripheral structures (Sibul, 2017). The field is modeled as a schematic network in which meaning emerges through configurations of domain linkages, and the position of each unit is determined by its degree of prototypicality and cognitive salience (Langacker, 1987; Taylor, 2003). The core of the category comprises terms anchored in the biological domain, such as *herd*, *pack*, *flock*, *swarm*, *colony*, and *pride*. In contrast, the periphery comprises metaphorically extended uses referring to social, technological, and other conceptual domains. At the superordinate level of categorization, the hyperonym *group* functions as a highly schematic unit that does not profile specific behavioral or structural parameters of collectivity. Collective nouns have been described as a distinctive category that combines conceptual unity with plurality, simultaneously encoding a collection of entities and its interpretation as a single conceptual whole (Tran & Bui, 2021). The primary cognitive mechanism facilitating collective nomination is the operation of uniplexing (Talmy, 2000), which enables a multiplex set of individuals to be construed as a unified Gestalt structure. The group is therefore construed as a coherent cognitive whole rather than as a simple aggregation of individual entities. Collective nouns combine grammatical singularity with semantic plurality, referring to multiple entities while construing them as a unified conceptual whole (Corbett, 2000; Levin, 2001). This dual structure underlies the operation of uniplexing by allowing multiple entities to be construed as a single Gestalt without eliminating awareness of their internal multiplicity.

According to de Vries (2021), collective nouns constitute a hybrid category that combines properties of one-ness and many-ness across different levels of linguistic representation. Collective nouns also exhibit a degree of non-additivity, whereby the

collective is interpreted as more than the sum of its individual members and acquires organizational properties of its own. This observation is closely linked to the notion of individuation, since collective nouns differ in the extent to which individual members remain conceptually accessible within the collective whole. In a related account, Joosten et al. (2007) argue that collective nouns vary in the degree to which individual members remain accessible. Attention may therefore shift between the collective as a whole and its constituent members depending on the dimensions of collective organization being profiled.

The selection of a particular collective term gives rise to a specific mode of semantic construal. Within the framework of Cognitive Grammar, this process involves the profiling of certain conceptual parameters against the backgrounding of others (Langacker, 1987). In this respect, the analyzed animal community nominations may be viewed as providing a skeletal structure that organizes conceptual content and foregrounds salient aspects of group organization, including hierarchy, integration, agency, and dynamicity. This interpretation is compatible with studies suggesting that collective nouns occupy a prominent position within the animacy hierarchy because they are associated with cognitively salient features such as agency, social organization, and coordinated activity (Jie & Liang, 2018). Consequently, animal community nominations do not merely denote plurality but also influence the way collective entities are conceptually interpreted. From this perspective, construal types may be regarded as abstract parameters of cognitive profiling realized through distinct nomination models. Cognitive distinctions in the interpretation of collectivity become particularly evident when the selected animal community nominations are compared with neutral designations of plurality. While specialized models foreground specific structural and behavioral parameters, neutral forms perform a schematic function by designating plurality

without specifying their internal organization (Taylor, 2003; Talmy, 2000). The choice of a nominative model affects the degree of conceptual integration among elements, shifting the interpretation of collectivity from discrete individuals toward a unified Gestalt structure.

The metaphorical extension of animal community nominations is interpreted within the framework of conceptual metaphor theory as a process of mapping in which the structure of the source domain ANIMAL COMMUNITY is projected onto social phenomena, enabling the conceptualization of one domain of experience through another (Lakoff and Johnson, 2003). Animal community nominations therefore function as cognitive models through which characteristics associated with the biological domain serve as the basis for the interpretation of social reality. Beyond cross-domain mapping, metaphorical extension also involves the activation of culturally shared scenarios that structure interpretation and evaluation (Musolff, 2016).

Within the Metaphorical Periphery, this process relies on the interaction of two central cognitive mechanisms: uniplexing (Talmy, 2000), which enables the conceptualization of multiplicity as a unified entity, and construal (Langacker, 1987), which guides the interpretation of social reality through the activation of specific cognitive domains. Within the proposed framework, metaphorically extended uses may give rise to recurrent evaluative meanings through the cognitive profiling of transferred conceptual features. This interpretation is consistent with discourse-oriented approaches to metaphor, which emphasize that metaphorical choices contribute to the construction of particular evaluations and representations of social reality in authentic discourse contexts (Semino, 2008). Following the approach proposed by Zaynuldinov and Quero Gervilla (2025), the evaluative component emerges through the interaction of rational evaluation, expressiveness, and emotionality within culturally conditioned models of perception.

These models include evaluative stereotypes within the framework of frame semantics (Fillmore, 1982; Fillmore & Atkins, 1992) as well as culturally embedded quasi-stereotypes (Teliya, 1996). Given the cultural variability of metaphor (Kövecses, 2005), the evaluative potential of animal community nominations is shaped by the conceptual model activated in discourse. As a result, these nominations function not only as a means of designation but also as an instrument for interpretation and evaluation.

Building on the theoretical perspectives discussed above, the present study proposes an integrated cognitive-semantic framework in which animal community nominations are analyzed through the interaction of uniplexing, cognitive profiling, metaphorical extension, and evaluative potential. Within this framework, animal community nominations are understood as conceptual tools that structure the conceptualization of animal collectives through the interaction of these cognitive mechanisms. The lexical-semantic field of animal community nominations is organized as a hierarchical system comprising a superordinate category centered on the lexeme *group*, the selected lexical subset (*herd*, *pack*, *flock*, *swarm*, *pride*, and *colony*), and a prototypically structured continuum extending from a Literal Core to a Metaphorical Periphery. This framework explains differences among the analyzed nominations through the cognitive profiling of distinct dimensions of collective organization, including hierarchy, coordinated movement, social structure, and collective agency. It also provides the theoretical foundation for the subsequent corpus analysis by guiding the identification of profiling parameters, the derivation of higher-order construal types, and the interpretation of evaluative potential.

Materials and Methods

The empirical analysis was guided by the integrated cognitive-semantic framework proposed in the present study. The framework combines uniplexing, cognitive profiling, metaphorical extension, and evaluative

potential within a unified analytical procedure and serves as the principal interpretive model throughout the analysis. Corpus methods were employed to operationalize the proposed framework through the systematic identification and interpretation of recurrent construal patterns across the analyzed nominations.

The empirical data for the present study were retrieved from the Corpus of Contemporary American English (COCA) (Davies, 2008 – present), a large balanced corpus of contemporary American English comprising naturally occurring texts from multiple discourse registers. The analysis was restricted to corpus texts published between 2010 and 2019 in order to ensure a stable and reproducible dataset while focusing on contemporary language use. All discourse registers available in COCA were included to capture the full range of usage of the analyzed nominations across contemporary American English. Corpus retrieval was based on the syntactic pattern *a X of Y* and focused on six animal community nominations: *herd*, *pack*, *flock*, *swarm*, *colony*, and *pride*. These nominations were selected because they constitute well-established lexical items denoting animal communities while systematically differing in the dimensions of collective organization they profile, including agency, homogeneity, synchrony, hierarchy, intensity, and functional integration. Their semantic diversity provided an appropriate basis for comparative cognitive-semantic analysis within the proposed analytical framework. Restricting the analysis to these six nominations ensured a systematic comparison of recurrent construal patterns within a coherent lexical-semantic field.

The study was deliberately restricted to conventional lexical nouns denoting animal collectives and providing sufficient semantic comparability for systematic cognitive-semantic analysis. Historically specialized terms of venery, such as murder of crows and parliament of owls, belong to a distinct historical naming tradition documented in medieval hunting terminology and later

compilations (Lipton, 1977). Their contemporary lexical status, degree of constructional dependence, and level of conventionalization require a different set of diagnostic criteria and therefore constitute a separate analytical problem. Including such expressions in the present dataset would extend the study from cognitive-semantic profiling to questions of lexical autonomy and categorial status that fall beyond the scope of the present study. The six selected nominations thus constitute a semantically diverse but methodologically coherent subset for examining recurrent dimensions of collective construal.

The initial corpus retrieval yielded 2,563 contexts across the six target nominations. All retrieved contexts were subsequently subjected to manual screening. Contexts were excluded if they contained homonymous non-target uses, fixed idiomatic or phraseological expressions, duplicated concordance lines, metalinguistic discussion, or other occurrences unrelated to the conceptualization of animal communities. The construction a school of was excluded from the outset because of its high degree of semantic homonymy and the predominance of non-target expressions (e.g., *school of thought*, *school of philosophy*), which would have reduced the semantic consistency of the dataset.

Following manual screening, 759 contexts were excluded, resulting in a final manually annotated corpus of 1,804 contexts. Each retained context was then classified into one of two analytical categories. Literal Core comprised 773 contexts referring to actual animal communities, whereas Metaphorically Extended Uses comprised 1,031 contexts in which animal community nominations were applied to human, technological, and abstract referents. The resulting annotated corpus served as the empirical basis for identifying profiling parameters, deriving higher-order construal types, and interpreting the evaluative potential of the analyzed nominations.

The distinction between literal and non-literal uses was informed by the principles of the Metaphor Identification Procedure Vrije Universiteit Amsterdam (MIPVU) developed by Steen et al. (2010). In accordance with the principles of MIPVU, contexts were classified as non-literal when the contextual meaning of the target nomination contrasted with its basic biological meaning while remaining interpretable through comparison with that basic meaning. In the present study, MIPVU was employed solely as an operational procedure for distinguishing literal from non-literal uses and was not intended to differentiate novel metaphors from fully lexicalized figurative meanings. Accordingly, all non-literal corpus instances involving cross-domain comparison with the biological source domain were treated as metaphorically extended uses for the purposes of the present analysis. This operational procedure provided a systematic and transparent basis for annotating corpus instances as either Literal Core or Metaphorically Extended Uses.

The analytical procedure consisted of seven consecutive stages. First, corpus contexts were retrieved from COCA using the syntactic pattern a X of Y for the six selected animal community nominations. Second, all retrieved corpus contexts were manually screened to exclude homonymous non-target uses, fixed idiomatic expressions, duplicated concordance lines, metalinguistic discussion, and other irrelevant occurrences. Third, the retained corpus instances were manually annotated according to the classification procedure described above. Fourth, each annotated context was examined to identify the dominant profiling parameter or parameters foregrounded by the nomination within its immediate discourse context. The principal profiling parameters included agency, homogeneity, synchrony, density/intensity, hierarchy, and functional integration. In reporting representative corpus examples, more specific contextual descriptors, such as reduced individuation, coordinated movement, multiplicity, spatial clustering, territoriality, and collective

identity, were used as interpretive glosses indicating how these broader profiling parameters were realized in particular contexts; these descriptors were not treated as independent coding categories. Fifth, recurring profiling parameters were compared across the annotated corpus and grouped inductively into four higher-order construal types: Agentive, Passive, Dynamic, and Structural. These higher-order categories were derived from recurrent similarities observed in the profiling parameters rather than established a priori. Sixth, all annotated corpus instances were manually coded for evaluative potential. Finally, the resulting construal types and evaluative tendencies were interpreted within the proposed integrated cognitive-semantic framework through the interaction of uniplexing, cognitive profiling, metaphorical extension, and evaluative potential.

The six profiling parameters were operationalized as recurrent dimensions of collective organization: agency as coordinated goal-directed action; homogeneity as collective uniformity and reduced individuation; synchrony as coordinated movement or behavioral alignment; density/intensity as spatial concentration, multiplicity, and interactional intensity; hierarchy as internal status differentiation; and functional integration as systemic interdependence and stable internal organization. These parameters were grouped inductively according to broader dimensions of collective organization: agency was associated with the Agentive construal, homogeneity with the Passive construal, synchrony and density/intensity with the Dynamic construal, and hierarchy and functional integration with the Structural construal. The labels Agentive, Passive, Dynamic, and Structural are used as analytical designations for recurrent higher-order profiling patterns rather than as categories adopted from an existing grammatical typology or as rigid lexical classes.

The corpus annotation was conducted by the first author. Since the dataset was not

independently double-coded, formal inter-annotator agreement statistics could not be calculated. To reduce interpretive subjectivity, the annotation procedure relied on explicit operational criteria. The distinction between literal and metaphorically extended uses was guided by MIPVU principles; cognitive profiling was based on the explicitly defined analytical parameters of agency, homogeneity, synchrony, density/intensity, hierarchy, and functional integration; and evaluative tendencies were identified on the basis of contextual lexical markers in the immediate co-text. Ambiguous cases were repeatedly re-examined against the same analytical criteria to maintain internal consistency. The absence of independent inter-annotator validation is acknowledged as a methodological limitation of the present study.

To examine the evaluative potential of the analyzed nominations, all annotated corpus instances were manually coded using contextual evaluative markers occurring in the immediate co-text. The coding procedure considered adjectival descriptions, verbal predicates, and other lexical indicators associated with recurrent evaluative meanings, including threat, aggression, dependency, prestige, conformity, and depersonalization. When multiple evaluative interpretations were possible, coding was based on the dominant evaluative orientation supported by the immediate discourse context. The identified evaluative markers were subsequently grouped into broader evaluative dimensions through qualitative analysis of recurrent contextual patterns. Particular attention was paid to recurring semantic configurations such as coordinated aggression, intensity, chaotic behavior, social hierarchy, and reduced individuality.

The interpretation of evaluative meanings was informed by the emotional-evaluative semantic framework proposed by Zaynuldinov and Quero Gervilla (2025), which conceptualizes evaluation as the interaction of rational assessment,

emotionality, and expressiveness. Particular attention was also paid to verbal predicates, as verb selection has been shown to provide linguistic evidence for cognitively salient conceptual properties associated with nominal categories (Ji & Liang, 2018). From a prototype-theoretical perspective, verbal predicates may be regarded as linguistic manifestations of conceptual features attributed to a referent, thereby providing additional evidence for the cognitive construal of collective entities.

For comparative purposes, the lexeme *group* was treated as a neutral schematic superordinate category within the lexical-semantic field of animal community nominations. Unlike the analyzed nominations (*herd*, *pack*, *flock*, *swarm*, *colony*, and *pride*), *group* does not itself profile particular dimensions of collective organization such as hierarchy, synchrony, agency, or functional integration. Instead, it serves as a schematic representation of collectivity against which the profiling properties of the analyzed nominations can be systematically compared. Owing to its schematic and non-specialized character, it was not included in the quantitative analysis. Comparison between the analyzed nominations and this neutral marker of plurality made it possible to identify mechanisms of construal, including cognitive profiling, as well as to examine the role of uniplexing, whereby a set of discrete individuals is conceptualized as a unified Gestalt structure (see Table 1).

Overall, the proposed framework applied corpus-based evidence through a sequence of analytical procedures linking uniplexing, cognitive profiling, metaphorical extension, and evaluative interpretation. This integration provides a systematic basis for identifying recurrent construal patterns across Literal and Metaphorically extended uses of animal community nominations.

Results

A typology of Cognitive Construals in Animal Community Nominations

Application of the proposed analytical framework to the annotated corpus revealed a systematic cognitive-semantic organization of the lexical-semantic field of animal community nominations. Corpus instances were annotated as either Literal Core or Metaphorically Extended Uses, reflecting the broader theoretical distinction between the Literal Core and the Metaphorical Periphery proposed in the present framework. The latter encompasses uses in which animal community nominations extend to human, technological, and abstract domains. Across both analytical categories, collective entities are consistently construed as coherent cognitive wholes rather than as simple aggregations of individuals. This recurring pattern reflects Talmy's (2000) notion of uniplexing, whereby multiple entities are conceptualized as a single Gestalt. Furthermore, the analysis indicates that metaphorically extended uses preserve selected characteristics associated with the biological source domain, allowing dimensions of collective organization to remain salient across different domains of use.

Comparative cognitive-semantic analysis of the annotated corpus identified recurrent profiling parameters, including agency, homogeneity, synchrony, density, hierarchy, and functional integration. Rather than constituting isolated semantic features, these parameters consistently converged on broader dimensions of collective organization across the corpus. This analytical generalization formed the basis for the inductive derivation of four higher-order construal types (Agentive, Passive, Dynamic, and Structural) which represent recurrent cognitive patterns of collectivity construal rather than fixed lexical or semantic categories.

Table 1. Inductive Derivation of Higher-Order Construal Types from Recurrent Profiling Parameters

Lexeme	Foregrounded Profiling Parameter(s)	Analytical Generalization	Higher-Order Construal
<i>pack</i>	agency	collective action	Agentive
<i>herd</i>	homogeneity	collective composition	Passive
<i>flock</i>	synchrony	collective dynamics	Dynamic
<i>swarm</i>	density, intensity	collective dynamics	Dynamic
<i>pride</i>	hierarchy	internal organization	Structural
<i>colony</i>	functional integration	internal organization	Structural

Table 1 summarizes inductive derivation of higher-order construal types from recurrent profiling parameters identified in the annotated corpus. Although each nomination foregrounds a distinct profiling parameter, several converge on broader dimensions of collective organization. These recurring patterns provide the basis for the four-type typology presented in Table 2. Selected corpus examples illustrating the analytical categories and recurrent profiling patterns underlying these higher-order construal types are provided in the Appendix.

The first macro-category comprises Agentive and Passive construals, represented by *pack* and *herd* respectively. Although both nominations conceptualize a plurality as a unified Gestalt through uniplexing, they differ substantially in the dimensions of collective organization they emphasize. *Pack* is prototypically associated with social predator groups acting as coordinated units and therefore highlights collective agency, coordination, and goal-directed action. These characteristics remain salient in metaphorically extended uses, where groups of people are frequently construed as unified social actors capable of collective action, often in ways perceived as threatening or antagonistic. *Herd*, by contrast, directs attention away from coordinated action and toward homogeneity, spatial cohesion, and reduced individuation. In metaphorically extended contexts, the nomination tends to background individual differences and highlight collective uniformity, thereby motivating interpretations associated with

conformity, depersonalization, or reduced autonomy. Together, *pack* and *herd* represent two contrasting construals of collectivity: one centered on agency and coordinated action, the other on homogeneity and diminished individuality.

The second macro-category is represented by Dynamic construals, exemplified by *flock* and *swarm*. Unlike Agentive and Passive construals, these nominations conceptualize collectivity primarily through patterns of movement and interaction. *Flock* foregrounds synchrony and coordinated movement, whereas *swarm* emphasizes density/intensity, and potential uncontrollability of collective behavior. The distinction becomes particularly evident in metaphorically extended contexts. Expressions such as *flock* of followers highlight behavioral alignment and dependence, while *swarm* of drones and *swarm* of trolls foreground overwhelming multiplicity, dynamic expansion, and reduced controllability. Despite these differences, both nominations profile collective dynamics through patterns of movement and interaction. The comparative analysis of the annotated corpus indicates that synchrony and density/intensity function as lower-level profiling parameters within a broader Dynamic construal type.

The third macro-category comprises Structural construals, instantiated by *pride* and *colony*. In contrast to the preceding types, these nominations conceptualize collectivity primarily through internal organization and relatively stable relationships among group

members. *Pride* emphasizes hierarchy, status, and kinship relations, characteristics that frequently remain salient in metaphorically extended uses, where the nomination is associated with prestige, exclusivity, and elevated social standing. *Colony*, by contrast, profiles functional integration, territorial embeddedness, and systemic interdependence. In both literal and metaphorically extended contexts, it tends to construe the collective as a highly integrated system whose members are connected through stable functional relations. Although *pride* and *colony* differ in their dominant profiling parameters, both instantiate Structural construals because they conceptualize collectivity through patterns of internal organization rather than through movement or agency.

The corpus analysis suggests that evaluative potential is not an inherent property of the analyzed lexical items but rather an emergent consequence of cognitive

construal. Evaluative meanings become particularly salient in metaphorically extended uses, where characteristics associated with animal communities are projected onto social or technological entities. In such contexts, the foregrounding of particular dimensions of collective organization contributes to recurring evaluative tendencies. Agentive construals are frequently associated with aggression and hostility, Passive construals with conformity and depersonalization, Dynamic construals with dependence, threat, or chaos, and Structural construals with prestige, stability, or systemic cohesion. The findings suggest that animal community nominations function as mechanisms of cognitive and evaluative profiling, rather than as lexical items with inherently fixed evaluative meanings, thereby shaping the interpretation of collective entities across both literal and metaphorically extended uses.

Table 2. Cognitive Construal Types, Profiling Parameters, and Evaluative Potential

Model group	Lexeme	Construal Type	Key Profiling Parameters	Evaluative Potential
Agentive / Passive	<i>Pack</i>	<i>Agentive</i>	agency, coordination	aggression, hostility
Agentive / Passive	<i>Herd</i>	<i>Passive</i>	homogeneity, depersonalization	conformity, depersonalization
Dynamic	<i>Flock</i>	<i>Dynamic</i>	synchrony	dependence, reduced autonomy
Dynamic	<i>Swarm</i>	<i>Dynamic</i>	density/intensity	threat, chaos
Structural	<i>Pride</i>	<i>Structural</i>	hierarchy, status	prestige, exclusivity
Structural	<i>Colony</i>	<i>Structural</i>	integration, territoriality	stability, cohesion

Table 2 summarizes how the four higher-order construal types identified in the annotated corpus are instantiated by the analyzed nominations and associated with recurrent profiling parameters and evaluative tendencies. The differences among these nominations reflect both their lexical meanings and the recurrent profiling patterns identified in the annotated corpus. Although

all six nominations are grounded in the cognitive operation of uniplexing, which enables multiple entities to be conceptualized as a single Gestalt, the nature of this Gestalt varies according to the parameters that are foregrounded, including agency, synchrony, density, and social hierarchy.

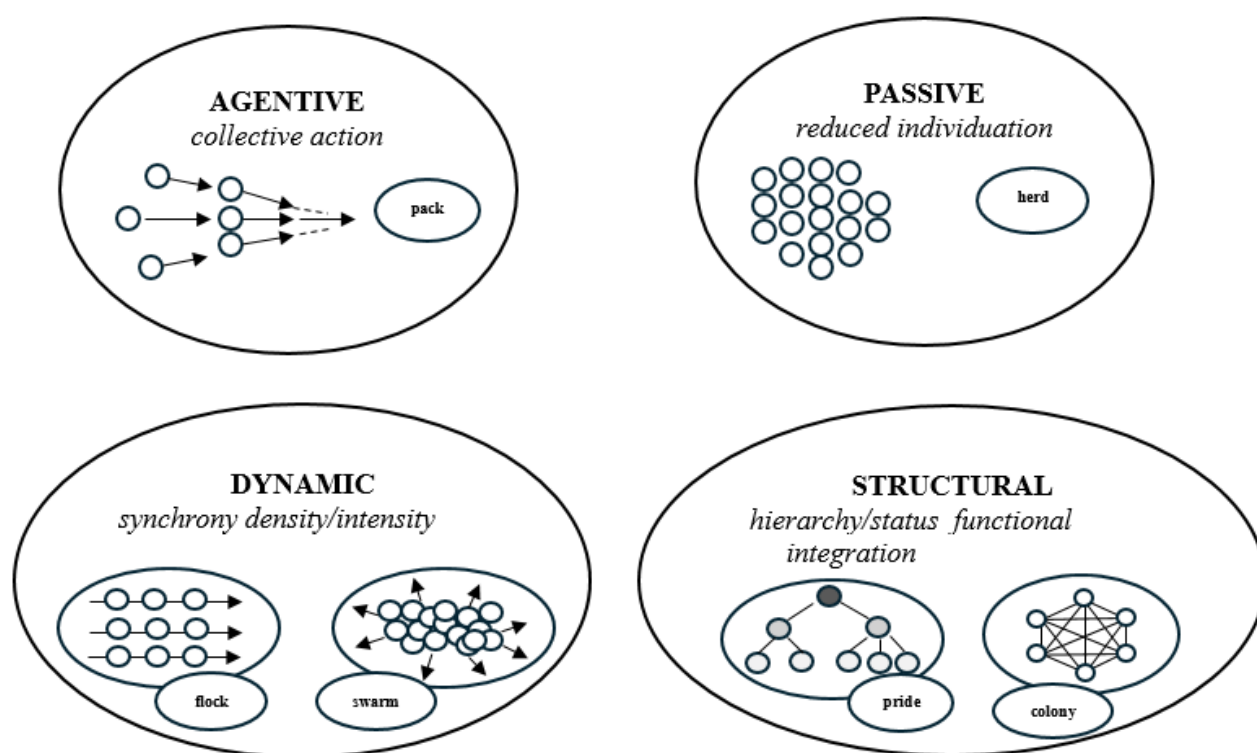
Each nomination reflects a particular construal of collective entities derived from

the recurrent profiling parameters identified during the comparative analysis of the annotated corpus. At a higher level of abstraction, these profiling parameters were grouped inductively into four higher-order construal types: Agentive, Passive, Dynamic, and Structural. The Dynamic construal type encompasses both *flock* and *swarm* because both nominations share a common focus on collective dynamics and patterns of movement and interaction. This higher-order grouping emerged inductively from the recurrent profiling parameters identified in the annotated corpus. Although the two nominations differ in the dominant profiling parameters they foreground, *flock* primarily profiles synchrony whereas *swarm* emphasizes density/intensity. Nevertheless, both construe collectivity through patterns of

movement and interaction. Accordingly, synchrony, density/intensity are interpreted as lower-level profiling parameters subsumed within the broader Dynamic construal type rather than as independent higher-order construal types.

Within the proposed framework, the six analyzed nominations instantiate the four higher-order construal types derived from recurrent profiling parameters identified in the annotated corpus. Specifically, *pack* represents the Agentive construal, *herd* the Passive construal, *flock* and *swarm* the Dynamic construal, and *pride* together with *colony* the Structural construal. Figure 1 summarizes the relationships between the higher-order construal types and their dominant profiling parameters.

Figure 1. Relationship between Profiling Parameters and Higher-Order Construal Types



Note. The figure summarizes the dominant profiling parameters associated with each higher-order construal type derived from the annotated corpus: collective action (Agentive), reduced individuation (Passive), synchrony and density/intensity (Dynamic),

and hierarchy/status and functional integration (Structural).

These distinctions among the four higher-order construal types become particularly evident in metaphorically extended uses, where cognitive characteristics

associated with the biological source domain motivate evaluative interpretations in human, technological, and abstract target domains. Through these profiling operations, animal community nominations contribute to the conceptualization and categorization of collective entities across both Literal Core and Metaphorically Extended Uses by foregrounding different dimensions of collective organization. The proposed higher-order construal types therefore demonstrate that these nominations function as mechanisms of cognitive profiling that shape both the conceptualization of collective entities and their evaluative interpretation in discourse.

The differences summarized in Figure 1 become particularly apparent when the analyzed nominations are compared with neutral expressions of plurality. A comparison between the analyzed animal community nominations and neutral expressions of plurality further highlights differences in the construal of collectivity. For instance, the expression *a pride of lions* profiles social hierarchy and kinship relations, whereas the expression *a group of lions* provides a comparatively neutral representation of collective membership. Similarly, *a herd of cattle* evokes a high degree of conceptual integration and reduced individuation, while the plural noun *cows* preserves the perception of separate individual entities. The analyzed nominations therefore shift attention from individual members to the collective as a unified conceptual entity while simultaneously emphasizing specific properties of collective organization. These evaluative interpretations emerge from encyclopedic knowledge associated with the biological source domain and become particularly salient in metaphorically extended uses, where this knowledge contributes to the interpretation of human, technological, and abstract collectives.

Taken together, the findings indicate that animal community nominations function not simply as linguistic markers of plurality but as cognitive mechanisms for construing

and profiling collective entities. The identified construal types differ in the dimensions of collective organization they profile, including agency, homogeneity, collective dynamics, and structural organization. Through these profiling operations, collective nominations contribute to both the conceptual representation and the evaluative interpretation of groups across literal and metaphorical contexts. The proposed typology further suggests that recurring evaluative meanings are systematically associated with specific construal types, providing a systematic connection between cognitive profiling and evaluative potential.

The Superordinate Expression *a group of*

For comparative purposes, *group* served as a neutral schematic reference category against which the six analyzed animal community nominations were compared. Unlike *herd*, *pack*, *flock*, *swarm*, *pride*, and *colony*, it does not itself profile particular dimensions of collective organization, such as agency, hierarchy, functional integration, or intensity. Instead, it provides a semantically underspecified conceptualization of plurality that enables the profiling properties of the analyzed nominations to become more apparent. Owing to its extreme frequency and semantic generality, this expression was not included in the quantitative corpus analysis, because its extreme frequency and semantic generality make it unsuitable for direct comparison with the six target nominations. Instead, it is treated as a reference category for cognitive-semantic comparison.

Comparison with the annotated corpus demonstrates that lexical choice itself functions as a mechanism of cognitive profiling. Whereas *group* evokes a general notion of collectivity without specifying its internal organization, the analyzed nominations systematically foreground distinct dimensions of collective organization. Thus, *a herd of animals* profiles homogeneity and reduced individuation, *a pack of animals* emphasizes coordinated agency, *a flock of followers* highlights behavioral alignment and

dependence, *a swarm of drones* foregrounds intensity and dynamicity, *a pride of lions* profiles social hierarchy, and *a colony of ants* emphasizes functional integration. These semantic differences are grounded not only in lexical meaning but also in the encyclopedic knowledge associated with different models of collective organization.

The identified profiling patterns remain available in metaphorically extended uses, where characteristics associated with animal communities contribute to the conceptualization and interpretation of human, technological, and abstract collectives. Consequently, *group* represents collectivity at a maximally schematic level, whereas the analyzed nominations evoke increasingly specific construals of collective organization through distinct profiling

patterns. These contrasts demonstrate that collective nominations differ not only in their referential scope but also in the cognitive construals they evoke.

To illustrate these differences in the construal of collectivity, Table 3 compares selected animal community nominations with the neutral schematic category *group*.

The neutral expressions in the Table 3 were selected according to the specific cognitive contrast being illustrated. While *group* serves as a neutral superordinate category, plural noun phrases and neutral quantificational expressions are used where necessary to demonstrate the cognitive operation of uniplexing and the additional profiling contributed by the selected nominations.

Table 3. Comparative Cognitive Construal of Collective Nominations

№	Linguistic Model	Neutral Reference Expression	Cognitive Mechanism	Interpretive Effect (Construal)
1	<i>A pride of lions</i>	<i>A group of lions</i>	Uniplexing; profiling of social parameters	<i>Pride</i> motivates the interpretation of a socially organized community characterized by hierarchy and internal social structure, whereas <i>group</i> provides a neutral representation of collective membership.
2	<i>A herd of cattle</i>	<i>Those cows</i>	Uniplexing; reduced individuation	<i>A herd of cattle</i> demonstrates a high degree of conceptual integration, construing individual animals as a unified whole, whereas <i>cows</i> preserves the perception of discrete individual entities.
3	<i>A flock of followers</i>	<i>Those followers</i>	Synchrony profiling	The collective is construed as characterized by dependence, coordinated behavior, and reduced individual autonomy.
4	<i>A swarm of drones</i>	<i>Many drones</i>	Intensity and density profiling	The nomination evokes the image of a dense, dynamic, and potentially threatening mass.

The examples presented in Table 3 demonstrate how the selected animal community nominations shape the cognitive profiling of collectivity in ways that differ from neutral expressions of plurality. Unlike neutral collective expressions such as *group*, *many*, or lexical items denoting individual entities, the analyzed nominations activate additional dimensions of construal beyond those encoded by neutral expressions of plurality, including social hierarchy, behavioral synchrony, interactional intensity, and reduced individuation. The choice of a collective nomination therefore influences not only how a plurality is represented linguistically but also how it is cognitively construed, giving rise to distinct models of collective organization in discourse. The observed interpretive differences are grounded not only in lexical meaning but also in encyclopedic knowledge linked to different models of collective organization. These contrasts further demonstrate that collective nominations function as mechanisms of cognitive profiling within the lexical-semantic field of animal community nominations, directing attention to particular dimensions of collective organization and motivating distinct interpretations of collective entities across both Literal Core and Metaphorically Extended Uses.

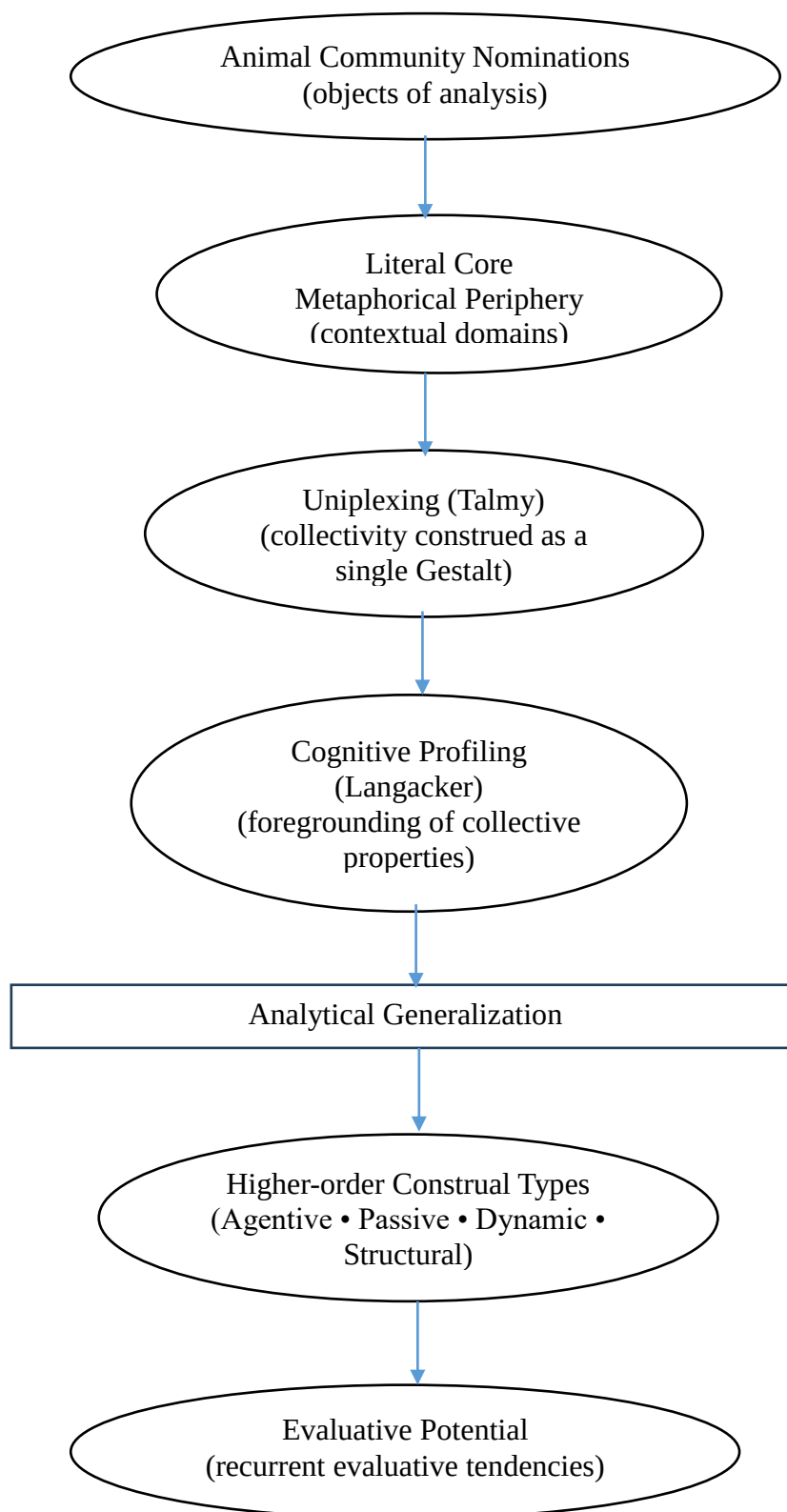
The comparative analysis further demonstrates that analyzed nominations differ systematically in the dimensions of collective organization they foreground. These differences are reflected not only in their construal patterns but also in their evaluative potential. The corpus data indicate recurrent associations between particular construal types and evaluative tendencies, suggesting that the analyzed nominations function as mechanisms of both cognitive and evaluative profiling. Rather than constituting isolated

observations, these findings form an integrated explanatory framework of collective conceptualization. The proposed cognitive-semantic model synthesizes the relationships identified in the annotated corpus by integrating uniplexing, cognitive profiling, higher-order construal types, metaphorically extended uses, and evaluative potential into a unified explanatory model, as illustrated in Figure 2.

Figure 2 presents the integrated cognitive-semantic model proposed in the present study. Unlike tables 1 and 2, which summarize the empirical findings of the corpus analysis, the figure integrates these findings into a unified explanatory model linking uniplexing, cognitive profiling, higher-order construal types, metaphorically extended uses, and evaluative potential. The hierarchical structure reflects the transition from the superordinate level of categorization, represented by the neutral schematic category *group*, to the analyzed nominations. Within the framework of Cognitive Grammar, *group* functions as a maximally schematic reference point that neither imposes a particular construal nor foregrounds specific dimensions of collective organization.

The transition from neutral reference to evaluative interpretation is mediated by a sequence of cognitive operations. The first stage involves uniplexing, through which a plurality of entities is construed as a unified Gestalt. Once collectivity has been conceptualized as a single cognitive entity, cognitive profiling foregrounds particular dimensions of collective organization. Based on the recurrent profiling parameters identified in the annotated corpus, four higher-order construal types were derived inductively: Agentive, Passive, Dynamic, and Structural.

Figure 2. Integrated Cognitive-Semantic Model of Collective Construal in Animal Community Nominations



These categories constitute an analytical typology based on recurring profiling patterns observed in the corpus data. They should be understood as dominant profiling tendencies rather than rigid semantic classes. Individual nominations may foreground multiple dimensions simultaneously, although each nomination exhibits a dominant profiling tendency that motivates its classification within the proposed typology. Individual nominations instantiate these higher-order construal types by foregrounding dominant profiling parameters while simultaneously backgrounding alternative dimensions of collective organization – such as coordination (*pack*), homogeneity (*herd*), synchrony (*flock*), and density/intensity (*swarm*), and hierarchy or functional integration (*pride*, *colony*) – while backgrounding other features.

The model synthesizes the recurring profiling patterns identified in the corpus and provides a schematic representation of the relationship between cognitive profiling, construal, and evaluative potential. It further captures how biological prototypes in the Literal Core may provide source-domain knowledge for the conceptualization of social and technological entities in the Metaphorically Extended Uses. The analysis suggests that evaluative potential is not an inherent property of the lexical items themselves but emerges through the interaction of cognitive profiling, contextual interpretation, and encyclopedic knowledge associated with the source domain. Such evaluative meanings become particularly salient in metaphorically extended uses, where characteristics associated with animal communities are projected onto social or technological entities.

In these metaphorical uses, the relationship between construal and evaluation is characterized by recurring tendencies observed in the analyzed corpus. Agentive construals (*pack*) are frequently associated with interpretations of aggression and hostility. Passive construals (*herd*) are often linked to conformity and loss of individuality. Dynamic construals (*flock*, *swarm*) may

motivate interpretations of dependency, threat, chaos, or intrusiveness. Structural construals (*pride*, *colony*) are commonly associated with prestige, stability, cohesion, or forms of social organization.

Taken together, the findings suggest a systematic relationship between uniplexing, cognitive profiling, construal types, and evaluative potential. Animal community nominations function not merely as markers of plurality but as mechanisms of cognitive and evaluative construal that shape the interpretation of collective entities across both literal and metaphorical domains.

Particularly noteworthy is the extension of *swarm* into the technological domain. Such extensions were observed in the corpus during the 2010 – 2019 period, particularly in expressions referring to drones, bots, autonomous systems, and online collectives. In expressions such as *swarm of drones*, *swarm of bots*, and *swarm of trolls*, the cognitive schema associated with insect swarms is projected onto autonomous or semi-autonomous technological entities. Core characteristics of the biological prototype – multiplicity, density/intensity of action – are retained; however, the resulting construal becomes more complex through the addition of algorithmic coordination. Unlike biological uses of *swarm*, which are commonly associated with disorder and limited controllability, technological uses often profile a combination of distributed agency and programmed synchrony. The metaphorical extension of *swarm* thus illustrates how a biological model of collectivity can be adapted to contemporary digital and technological forms of organization.

The analysis demonstrates the explanatory value of the proposed analytical framework. By integrating uniplexing, cognitive profiling, metaphorical extension, and evaluative potential within a unified analytical procedure, the framework makes it possible to explain recurrent construal patterns that cannot be fully accounted for when these mechanisms are examined

separately. The resulting four-type analytical typology, derived inductively from recurrent profiling parameters identified in the annotated corpus, provides a systematic cognitive-semantic model for interpreting the organization of animal community nominations and their evaluative tendencies across both Literal and Metaphorical domains.

Discussion

The findings demonstrate that animal community nominations function as cognitive mechanisms of collective construal rather than as neutral markers of plurality. The proposed analytical framework explains this behavior by integrating uniplexing, cognitive profiling, metaphorically extended uses, and evaluative potential within a unified explanatory model. The findings indicate that the choice of nomination influences the conceptualization of a group, facilitating a shift from the perception of individual entities to their interpretation as a unified cognitive unit. In this sense, collective nominations function as mechanisms of conceptual structuring that provide alternative ways of construing collective entities in discourse.

The principal contribution of the present study lies in demonstrating that the interaction between uniplexing, cognitive profiling, metaphorical extension, and evaluative potential can be explained within a unified framework. Within this framework, cognitive profiling links the construal of collectivity established through uniplexing with the emergence of higher-order construal types, thereby providing the basis for interpreting their evaluative potential. More broadly, the proposed framework demonstrates how corpus-based cognitive-semantic analysis can move beyond the description of individual lexical items toward the identification of recurrent mechanisms underlying the conceptualization and evaluation of collectivity.

The results indicate that the Literal Core of the lexical-semantic field is anchored in the biological domain and can be interpreted as exhibiting a high degree of prototypicality.

Metaphorically Extended Uses are organized relative to this prototypical foundation by preserving selected organizational characteristics of the biological source domain while projecting them onto human, technological, and abstract target domains. Compared with the other analyzed nominations, *pride* showed the smallest number of metaphorically extended uses among the six analyzed nominations. Most non-literal occurrences represented idiomatic expression (e.g. *pride of place*), which were excluded during corpus annotation according to the analytical criteria adopted in the present study. Consequently, only a small number of metaphorically extended uses remained eligible for analysis. The remaining metaphorically extended uses consistently preserved the source-domain concepts of social cohesion, hierarchical organization, and collective identity, suggesting that metaphorical extension of *pride* is considerably more restricted than that of the other analyzed nominations. The analysis of Metaphorically Extended Uses further indicates that the evaluative potential of the lexical-semantic field is shaped not only by universal cognitive mechanisms but also by culturally specific evaluative models embedded in a particular linguocultural context. In this respect, evaluative stereotypes and quasi-stereotypes can be regarded as mechanisms of interpretation reflected in actual language use (Fillmore, 1982; Teliya, 1996).

The role of culturally embedded evaluative models becomes particularly apparent when the English corpus data are considered from a broader cross-linguistic perspective. The ethnocultural specificity of secondary nomination is reflected in the variability of associative characteristics attributed to animal images across languages. For example, in Russian cultural conceptualization, the *ram* is conventionally associated with stupidity, whereas the *donkey* is associated with stubbornness. In Spanish, however, both characteristics are conventionally integrated into the image of

burro. Although these examples fall outside the English corpus examined in this study, they illustrate the broader role of cultural conceptualization in shaping evaluative associations. Such differences demonstrate that animal-based associations are culturally specific and reflect distinct patterns of cultural conceptualization. These observations are consistent with Kövecses's account of cultural variation in metaphorical conceptualization.

From this perspective, the evaluative potential of English collective nominations appears to emerge through the interaction of universal cognitive mechanisms, particularly uniplexing and cognitive profiling, with culturally entrenched evaluative models. Consequently, English collective nominations such as *herd*, *pack*, and *swarm* evoke evaluative interpretations that are grounded not only in lexical meaning but also in the encyclopedic knowledge associated with particular models of collective organization.

The findings are consistent with Langacker's Cognitive Grammar, according to which linguistic expressions impose particular construals on the situations they represent (Langacker, 1987). This observation suggests that lexical choice itself functions as a cognitively meaningful act of categorization rather than as a purely referential selection. The comparison between the analyzed nominations and the superordinate expression *a group of* demonstrates that these forms are not interchangeable but systematically profile different dimensions of collective organization. Unlike the highly schematic lexeme *group*, the analyzed nominations selectively foreground organizational properties that guide both the conceptualization and subsequent evaluation of collective entities. These findings demonstrate that collective nominations function not merely as labels for groups but as mechanisms of selective cognitive profiling that shape the interpretation of collectivity and the evaluative perception of social actors. This interpretation is consistent with discourse-oriented approaches to metaphor and media discourse, which emphasize that

linguistic choices contribute to the construction of evaluations and representations of social reality in authentic discourse contexts (Semino, 2008; Hart, 2014; 2021)

Beyond lexical categorization, the present findings also contribute to broader cognitive theories of collective representation. In this respect, the present findings are compatible with Tomasello's (2019) account of collective intentionality. Although the present study does not investigate intentionality directly, the construal of multiple individuals as a unified collective Gestalt appears consistent with broader cognitive mechanisms through which groups are conceptualized as coherent social entities. Animal community nominations can therefore be understood not simply as labels for collective referents but as linguistic resources that facilitate the interpretation of groups in terms of social coordination, shared organization, and collective identity. While Tomasello's account helps explain the conceptualization of collectivity as a coherent social entity, the evaluative interpretation of such collectives requires consideration of the metaphorical scenarios through which these conceptual structures are activated in discourse.

The evaluative potential of animal community nominations can be further explained through Musolff's (2016) notion of metaphorical scenarios. Rather than functioning as simple cross-domain mappings between source and target domains, metaphorical extensions activate culturally shared scenarios associated with particular models of collective organization. Within these scenarios, nominations such as *pack*, *herd*, or *swarm* evoke recurring interpretative patterns associated with coordinated hunting, collective following, or mass movement, thereby shaping their evaluative interpretation in social and technological contexts. Through metaphorical extension, these conceptual structures are projected onto new target domains while remaining connected to the prototypical core of the category through

shared organizational features. Consequently, collective entities may be construed as aggressive, passive, hierarchically organized, or functionally integrated, depending on the cognitive model activated in discourse. These findings further support the proposed analytical framework by demonstrating how cognitive profiling interacts with culturally shared metaphorical scenarios in guiding both the conceptualization and evaluative interpretation of collectivity.

Digital and Technological Extensions of Collective Construals

A significant finding emerging from the corpus analysis is the extension of animal community nominations beyond their biological origins into technological and digital domains. This tendency is clearly observable in the analyzed corpus data, where algorithmic and networked systems are frequently conceptualized through collective schemas. Expressions such as *swarm of drones*, *swarm of bots*, and *swarm of trolls* illustrate how traditional models of collectivity are applied to digital and networked environments.

Among the analyzed nominations, *swarm* showed a particularly salient extension into technological and digital domains in the analyzed corpus and therefore serves as an illustrative case of contemporary cross-domain extension. While its Literal Core is anchored in biological collectives, its Metaphorically Extended Uses extend across technological, digital, natural, and astronomical domains. This productivity reflects the schematic nature of the *swarm* model. Unlike nominations that profile more stable forms of collective organization (e.g., *pride*), *swarm* foregrounds dynamic properties such as multiplicity, density, and distributed coordination. These relatively schematic properties enable this construal to be projected onto a wide range of biological, technological, and other collective entities characterized by interactional intensity and rapid expansion.

A related pattern can also be observed in other biologically motivated collective

schemas. Unlike the *swarm* model, which foregrounds movement and density, these extend biological models of social insects to collective decision-making, online communities, and artificial intelligence, illustrating how animal-based schemas increasingly structure the conceptualization of coordinated technological systems.

The technological extension of *swarm* is also associated with a diversity of evaluative tendencies. In the analyzed corpus, expressions such as *swarm of trolls* and *swarm of drones* frequently instantiate negative evaluative meanings that emerge through the interaction of cognitive profiling, contextual interpretation, and source-domain knowledge rather than from the lexeme itself. When characteristics such as density, multiplicity, and reduced controllability are projected onto technological entities, the resulting construal often contributes to interpretations of intrusiveness, disruption, or threat. However, recent developments in assistive technologies suggest that the same collective schema may also be associated with more positive evaluations. This observation suggests that evaluative potential is context-dependent rather than lexically fixed. For example, the SwarmFidget project employs coordinated groups of robots to support emotion regulation and attentional engagement in adults with ADHD (Pulatova & Kim, 2024), illustrating the ontological flexibility of the *swarm* schema and its capacity to conceptualize emerging forms of coordinated technological agency.

The findings indicate that animal community nominations remain productive cognitive resources for conceptualizing new forms of coordinated technological and digital organization, extending well beyond their original biological domain.

Beyond the present dataset, the proposed analytical framework provides a transferable analytical model for investigating additional animal community nominations and related lexical domains of collectivity. Future applications across different corpora, discourse types, and languages may further

refine the proposed analytical typology and evaluate its broader explanatory potential.

Limitations

First, the study was based on a manually annotated qualitative corpus of 1,804 contexts. Although the annotated corpus was relatively large, the analysis itself remained qualitative and focused on identifying recurrent construal patterns rather than establishing statistically generalizable frequency distributions. Second, the empirical analysis was restricted to contemporary English-language data retrieved from the Corpus of Contemporary American English (COCA) and therefore does not necessarily capture patterns occurring in other varieties of English or historical stages of the language. Third, the study was limited to six analyzed animal community nominations retrieved through the syntactic pattern *a X of Y*. Consequently, the proposed analytical typology should be interpreted as an inductively derived explanatory model rather than as an exhaustive classification of all animal community nominations.

A further limitation concerns annotation reliability. Because all corpus instances were coded by a single annotator, formal inter-annotator agreement statistics were not available. Although explicit coding criteria and repeated examination of ambiguous cases were used to reduce interpretive subjectivity, future studies should validate the annotation scheme through independent double coding of a representative subsample and the calculation of inter-annotator agreement statistics.

These limitations do not diminish the contribution of the proposed framework but rather define the scope within which the present findings should be interpreted. Because the proposed analytical framework and typology were developed inductively from the analyzed corpus, their explanatory value has been demonstrated only within the scope of the present dataset. Further validation using independent datasets, additional animal community nominations, and other lexical domains will be necessary to

evaluate the robustness and broader applicability of the proposed framework.

Future Research

The proposed analytical framework also opens several directions for future research. Future studies may extend the proposed framework to additional animal community nominations, other lexical-semantic fields, different discourse types, and cross-linguistic data. Such investigations would provide an opportunity to evaluate the transferability of the proposed analytical typology and determine whether similar construal patterns emerge across languages, discourse contexts, and lexical domains.

A particularly promising direction concerns the continuing metaphorical extension of animal community nominations into technological and digitally mediated forms of collectivity. Expressions such as *swarm of drones*, *swarm of bots*, *hive mind*, and *coordinated AI* system suggest that animal-based models of collective organization are increasingly employed to conceptualize emerging technological systems. Recent research on *swarm* user interfaces likewise indicates that biologically inspired collective behavior is increasingly being adopted as an engineering and design principle for digital and robotic systems (Zhang et al., 2026).

These developments also raise important questions regarding the evaluative consequences of technological metaphorical extension. Another important question concerns whether these emerging technological uses preserve the evaluative tendencies associated with their biological source domains or give rise to new evaluative profiles shaped by contemporary technological discourse. Emerging research on Conversational Swarm Intelligence (Rosenberg et al., 2024) likewise suggests that swarm-based models are becoming increasingly relevant for describing collaborative forms of AI-mediated interaction, providing another potential domain in which the proposed analytical framework may be applied.

Such studies would also provide an opportunity to validate the proposed analytical framework and its analytical typology on independent datasets and determine the extent to which the identified construal patterns remain stable across different lexical domains and discourse types.

Conclusion

The present study has demonstrated that animal community nominations constitute a hierarchically organized lexical-semantic field with a prototypical structure, comprising a Literal Core anchored in biological reference and a Metaphorically Extended Uses extending into social and technological domains. The analysis further showed that collective nominations differ systematically in the dimensions of collective organization they foreground, giving rise to recurrent construal patterns and associated evaluative tendencies.

The application of this framework to the annotated corpus lies in the development of an integrated cognitive-semantic framework that explains how uniplexing, cognitive profiling, metaphorical extension, and evaluative potential interact in the conceptualization of animal community nominations. Application of this framework to contemporary corpus data yielded a four-type analytical typology (Agentive, Passive, Dynamic, and Structural), providing an explanatory model for recurrent construal patterns observed across the analyzed nominations. Rather than constituting an exhaustive classification of collective construals, this typology represents an analytical generalization of the dominant profiling tendencies identified in the analyzed corpus.

Beyond the specific findings of the present study, the results contribute to a broader understanding of collective conceptualization by demonstrating that collective nominations do not simply categorize groups but systematically shape their cognitive interpretation. Rather than functioning as neutral markers of plurality, analyzed animal community nominations foreground distinct dimensions of collective

organization that guide subsequent evaluative construal. The findings further demonstrate that the interaction between uniplexing, cognitive profiling, metaphorical extension, and evaluative potential can be accounted for within a unified analytical model of animal community nominations. Taken together, the proposed framework offers a transferable and reproducible cognitive-semantic model for investigating collective construal across animal community nominations and related lexical domains, thereby extending the analytical potential of corpus-based cognitive-semantic research.

Statement on the Use of Generative

AI: The authors declare that no generative artificial intelligence or AI-assisted technologies requiring disclosure under the Journal's policy were used in conducting the research or preparing this manuscript.

Заявление об использовании генеративного ИИ: Авторы заявляют, что при проведении исследования и подготовке настоящей рукописи генеративный искусственный интеллект и ИИ-ассистированные технологии, подлежащие раскрытию в соответствии с политикой журнала, не использовались.

References

- Boldyrev, N.N. (2014). *Kognitivnaya semantika: vvedenie v kognitivnuyu lingvistiku*. [Cognitive Semantics: An Introduction to Cognitive Linguistics]. Izd-vo TGU im. G.R. Derzhavina, Tambov, Russia. (In Russian)
- Corbett, G.G. (2000). *Number*. Cambridge University Press, Cambridge, UK. (In English)
- Croft, W. and Cruse, D.A. (2004). *Cognitive Linguistics*. Cambridge University Press, Cambridge, UK. (In English)
- Crystal, D. (2011). *The Story of English in 100 Words*. Profile Books, London, UK. (In English)
- Davies, M. (2008–present). *The Corpus of Contemporary American English (COCA)*, available at: <https://www.english-corpora.org/coca/> (Accessed 10 July, 2026). (In English)
- de Vries, H. (2021). 'Collective nouns', in Cabredo Hofherr, P. and Doetjes, J. (eds), *The*

Oxford Handbook of Grammatical Number. Oxford University Press, Oxford, UK, 257–274. <https://doi.org/10.1093/oxfordhb/9780198795858.013.12> (In English)

Evans, V. and Green, M. (2006). *Cognitive Linguistics: An Introduction*. Edinburgh University Press, Edinburgh, UK. (In English)

Fillmore, C. J. (1982). Frame Semantics, in *Linguistics in the Morning Calm*, Hanshin Publishing Company, Seoul, South Korea, 111–137. (In English)

Fillmore, C. J. and Atkins, B.T. (1992). Toward a frame-based lexicon: the semantics of RISK and its neighbors, in Lehrer, A. and Kittay, E.F. (eds), *Frames, Fields, and Contrasts*. Lawrence Erlbaum, Hillsdale, NJ, USA, 75–102. (In English)

Geeraerts, D. (2009). *Theories of Lexical Semantics*. Oxford University Press, Oxford, UK. (In English)

Hart, C. (2014). *Discourse, Grammar and Ideology: Functional and Cognitive Perspectives*. Bloomsbury Academic, London, UK. (In English)

Hart, C. and Marmol Queralto, J. (2021). ‘What can cognitive linguistics tell us about language-image relations? A multidimensional approach to intersemiotic convergence in multimodal texts’, *Cognitive Linguistics*, 32 (4), 529–562. <https://doi.org/10.1515/cog-2021-0039> (In English)

Ji Ji. and Liang. M. (2018). An animacy hierarchy within inanimate nouns: English Corpus evidence from a prototypical perspective, *Lingua*, 205, 71–89. (In English)

Joosten, F., De Sutter, G., Drieghe, D., Grondelaers, S., Hartsuiker, R.J. and Speelman, D. (2007). Dutch collective nouns and conceptual profiling, *Linguistics*, 45 (1), 85–132. (In English)

Kövecses, Z. (2005). *Metaphor in Culture: Universality and Variation*. Cambridge University Press, Cambridge, UK. (In English)

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press, Chicago, IL, USA. (In English)

Lakoff, G. and Johnson, M. (2003). *Metaphors We Live By*. University of Chicago Press, Chicago, IL, USA. (In English)

Langacker, R.W. (1987). *Foundations of Cognitive Grammar. Vol. 1: Theoretical Prerequisites*. Stanford University Press, Stanford, CA, USA. (In English)

Levin, M. (2001). *Agreement with Collective Nouns in English*. Lund University, Lund, Sweden. (In English)

Lipton, J. (1977). *An Exaltation of Larks, or The Venereal Game*. Expanded 2nd ed. Viking Press, New York, NY, USA. (In English)

Musolff, A. (2016). *Political Metaphor Analysis: Discourse and Scenarios*. Bloomsbury Academic, London, UK. (In English)

Pulatova, S. and Kim, L.H. (2024). Co-designing programmable fidgeting experience with swarm robots for adults with ADHD, in *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24)*. ACM, New York, NY, USA, 1–14. <https://doi.org/10.1145/3663548.3675614> (In English)

Rosenberg, L., Willcox, G., Schumann, H. and Mani, G. (2024). *Towards Collective Superintelligence: Amplifying Group IQ using Conversational Swarms*. arXiv preprint arXiv:2401.15109 [Online], available at: <https://arxiv.org/abs/2401.15109> (Accessed 10 July 2026). (In English)

Rosch, E. (1978). Principles of Categorization, in Rosch, E. and Lloyd, B.B. (eds), *Cognition and Categorization*. Lawrence Erlbaum, Hillsdale, NJ, USA, 27–48. (In English)

Semino, E. (2008). *Metaphor in Discourse*. Cambridge University Press, Cambridge, UK. (In English)

Sibul, V. (2017). Psycholinguistic analysis of lexical-semantic structure in linguistic consciousness of Russian, English and German native speakers, *Training, Language and Culture*. 1 (2), 54–70. <https://doi.org/10.29366/2017tlc.1.2.4> (In English)

Steen, G.J., Dorst, A.G., Herrmann, J.B., Kaal, A., Krennmayr, T. and Pasma, T. (2010). *A Method for Linguistic Metaphor Identification: From MIP to MIPVU*. John Benjamins Publishing Company, Amsterdam, Netherlands. (In English)

Talmy, L. (2000). *Toward a Cognitive Semantics. Vol. 1: Concept Structuring Systems*. MIT Press, Cambridge, MA, USA. (In English)

Taylor, J.R. (2003). *Linguistic Categorization*. 3rd ed. Oxford University Press, Oxford, UK. (In English)

Taylor, J.R. (2018). *Prototypes in Cognitive Linguistics*. Oxford University Press, Oxford, UK. (In English)

Tomasello, M. (2019). *Becoming Human: A Theory of Ontogeny*. Harvard University Press, Cambridge, MA, USA. (In English)

Tran, T.T.O. and Bui, T.V.T. (2021). The semantic features of collective nouns in English, *International Journal of TESOL & Education*, 1 (3), 130–141. (In English)

Teliya, V.N. (1996). *Russkaya frazeologiya: semanticheskiy, pragmaticheskiy i lingvokulturologicheskiy aspekty* [Russian Phraseology: Semantic, Pragmatic and Linguocultural Aspects], Shkola «Yazyki russkoy kul'tury», Moscow, Russia. (In Russian)

Zaynuldinov, A. A. and Quero Gervilla, E. F. (2025). *Leksikograficheskoe opisanie emotsional'no-otsenichnoy leksiki v russkom i ispanskom yazykakh* [Lexicographic Description of Emotional-Evaluative Vocabulary in Russian and Spanish], *Voprosy Yazykoznanija*, 2, 96–110. (In Russian)

Zhang, X., Aggarwal, V., Zhu, A. and Kim, L.H. (2026). Feeling with Many: Rethinking Emotion Regulation with Swarm User Interfaces, in: *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*, Barcelona, Spain, 13-17 April 2026. ACM, New York, NY, USA. Article 1440, 1–23. <https://doi.org/10.1145/3772318.3791599> (In English)

Список литературы

Болдырев Н. Н. Когнитивная семантика: введение в когнитивную лингвистику. Тамбов : Изд-во ТГУ им. Г. Р. Державина, 2014.

Телия В. Н. Русская фразеология: семантический, прагматический и лингвокультурологический аспекты. М.: Школа «Языки русской культуры», 1996.

Зайнульдинов А. А., Quero Gervilla E. F. Лексикографическое описание эмоционально-оценочной лексики в русском и испанском языках // *Вопросы языкознания*. 2025. No 2. С. 96–110.

Corbett G. G. *Number*. Cambridge: Cambridge University Press, 2000.

Croft W., Cruse D. A. *Cognitive Linguistics*. Cambridge: Cambridge University Press, 2004.

Crystal D. *The Story of English in 100 Words*. London: Profile Books, 2011.

Davies M. The Corpus of Contemporary American English (COCA). 2008 – present. URL: <https://www.english-corpora.org/coca/> (дата обращения: 10.07.2026)

de Vries H. Collective nouns // *The Oxford Handbook of Grammatical Number* / ed. by P. Cabredo Hofherr, J. Doetjes. Oxford : Oxford University Press, 2021. Pp. 257–274. DOI: 10.1093/oxfordhb/9780198795858.013.12.

Evans V., Green M. *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press, 2006.

Fillmore C. J. *Frame Semantics* // *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Company, 1982. Pp. 111–137.

Fillmore C. J., Atkins B. T. Toward a frame-based lexicon: the semantics of RISK and its neighbors // *Frames, Fields, and Contrasts* / ed. by A. Lehrer, E. F. Kittay. Hillsdale, NJ: Lawrence Erlbaum, 1992. Pp. 75–102.

Geeraerts D. *Theories of Lexical Semantics*. Oxford: Oxford University Press, 2009.

Hart C. *Discourse, Grammar and Ideology: Functional and Cognitive Perspectives*. London: Bloomsbury Academic, 2014.

Hart C., Marmol Queralto J. What can cognitive linguistics tell us about language-image relations? A multidimensional approach to intersemiotic convergence in multimodal texts // *Cognitive Linguistics*. 2021. Vol. 32. No. 4 Pp. 529–562. URL: <https://doi.org/10.1515/cog-2021-0039>

Ji J., Liang M. An animacy hierarchy within inanimate nouns: English corpus evidence from a prototypical perspective // *Lingua*. 2018. Vol. 205. Pp. 71–89.

Joosten F., De Sutter G., Drieghe D., Grondelaers S., Hartsuiker R. J., Speelman D. Dutch collective nouns and conceptual profiling // *Linguistics*. 2007. Vol. 45. No. 1. Pp. 85–132.

Kövecses Z. *Metaphor in Culture: Universality and Variation*. Cambridge: Cambridge University Press, 2005.

Lakoff G. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press, 1987.

Lakoff G., Johnson M. *Metaphors We Live By*. Chicago: University of Chicago Press, 2003.

Langacker R. W. *Foundations of Cognitive Grammar*. Vol. 1: Theoretical Prerequisites. Stanford, CA: Stanford University Press, 1987.

Levin M. *Agreement with Collective Nouns in English*. Lund: Lund University, 2001.

Lipton J. An Exaltation of Larks, or The Venereal Game. Expanded 2nd ed. New York: Viking Press, 1977.

Musolff A. Political Metaphor Analysis: Discourse and Scenarios. London: Bloomsbury Academic, 2016.

Pulatova S., Kim L. H. Co-designing programmable fidgeting experience with swarm robots for adults with ADHD // Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24). New York: ACM, 2024. Pp. 1–14. URL: <https://doi.org/10.1145/3663548.3675614>

Rosenberg L., Willcox G., Schumann H., Mani G. Towards Collective Superintelligence: Amplifying Group IQ using Conversational Swarms. arXiv preprint arXiv:2401.15109. 2024. URL: <https://arxiv.org/abs/2401.15109> (дата обращения: 10.07.2026).

Rosch E. Principles of Categorization // Cognition and Categorization / ed. by E. Rosch, B. B. Lloyd. Hillsdale, NJ : Lawrence Erlbaum, 1978. Pp. 27–48.

Semino E. Metaphor in Discourse. Cambridge : Cambridge University Press, 2008.

Sibul V. Psycholinguistic analysis of lexical-semantic structure in linguistic consciousness of Russian, English and German native speakers // Training, Language and Culture. 2017. Vol. 1. No. 2. Pp. 54–70. URL: <https://doi.org/10.29366/2017tlc.1.2.4>

Steen G. J., Dorst A. G., Herrmann J. B., Kaal A., Krennmayr T., Pasma T. A Method for Linguistic Metaphor Identification: From MIP to MIPVU. Amsterdam: John Benjamins Publishing Company, 2010.

Talmy L. Toward a Cognitive Semantics. Vol. 1: Concept Structuring Systems. Cambridge, MA: MIT Press, 2000.

Taylor J. R. Linguistic Categorization. 3rd ed. Oxford: Oxford University Press, 2003.

Taylor J. R. Prototypes in Cognitive Linguistics. Oxford: Oxford University Press, 2018.

Tomasello M. Becoming Human: A Theory of Ontogeny. Cambridge, MA: Harvard University Press, 2019.

Tran T. T. O., Bui T. V. T. The semantic features of collective nouns in English // International Journal of TESOL & Education. 2021. Vol. 1. No. 3. Pp. 130–141.

Zhang X., Aggarwal V., Zhu A., Kim L. H. Feeling with Many: Rethinking Emotion Regulation with Swarm User Interfaces // Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26). Barcelona, Spain, 13-17 April 2026. New York: ACM, 2026. Article 1440. Pp. 1–23. URL: <https://doi.org/10.1145/3772318.3791599>

All authors have read and approved the final manuscript.

Все 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.

Madina M. Zhemileva, PhD Student in Foreign Philology (Western Languages), Faculty of Philology, Al-Farabi Kazakh National University, Almaty, Kazakhstan.

Мадина Маратовна Жемилева, докторант по специальности «Иностранная филология (западные языки)», факультет филологии, Казахский национальный университет имени аль-Фараби, Алматы, Казахстан.

Enrique Federico Quero Gervilla, Ph.D., Full Professor, Doctor of the Department of Greek and Slavic Philology, University of Granada, Granada, Spain.

Энрике Федерико Керо Хервилья, кандидат филологических наук, профессор кафедры греческой и славянской филологии, Университет Гранады, Гранада, Испания.

Appendix. Selected Corpus Examples

ID	Lexeme	Context	Source	Year	Domain	Literal/ Metaphorical	Dominant Profiling Parameter (s)	Construal type	Evaluative Potential
H1	<i>Herd</i>	They're a herd of craven dumb people.	Fox News Channel/ SPOK	2018	Media / Social	Metaphorical	Homogeneity; reduced individuation	Passive	Negative
H2	<i>Herd</i>	A herd of men on the internet attacked her online	Freethought blogs.com/ WEB	2012	Social	Metaphorical	Homogeneity; reduced individuation	Passive	Negative
H3	<i>Herd</i>	A shooting into a herd of resting elephants	Goodreads.c om/WEB	2012	Wildlife	Literal	Homogeneity; reduced individuation	Passive	Neutral
H4	<i>Herd</i>	A herd of cattle stand in the midst of smoke	Rollingstone .com/ MAG	2019	Agriculture	Literal	Homogeneity; reduced individuation	Passive	Neutral
H5	<i>Herd</i>	Getting knocked over on the halls by a herd of students	Teencerebral palsy.com/ BLOG	2012	Social	Metaphorical	Homogeneity; reduced individuation	Passive	Negative
H6	<i>Herd</i>	Was a herd of deer calmly standing	Thisisthepla ce.org / WEB	2012	Wildlife	Literal	Homogeneity; reduced individuation	Passive	Neutral
H7	<i>Herd</i>	A herd of guests watched the proceeding in confusion.	Craphound.c om / WEB	2012	Social	Metaphorical	Homogeneity; reduced individuation	Passive	Negative
H8	<i>Herd</i>	Saw a herd of people stampeding out	Highlights for Children / FIC	2017	Social	Metaphorical	Homogeneity; reduced individuation; intensity	Passive	Negative
H9	<i>Herd</i>	A herd of shoppers stops in front of the store	Esquire / FIC	2018	Social	Metaphorical	Homogeneity; reduced individuation	Passive	Neutral
H10	<i>Herd</i>	You didn't get that respect from a herd of egocentric bureaucrats	A.V. Club/ MAG	2016	Social	Metaphorical	Homogeneity; reduced individuation	Passive	Negative

P1	Pack	A pack of wolves can take down an elk	CBS_Early / SPOK	2010	Wildlife	Literal	Agency; coordinated action	Agentive	Neutral
P2	Pack	A pack of puppies training to become guide dogs	NEWS.com.au	2010	Domestic animals	Literal	Agency; coordinated action	Agentive	Positive
P3	Pack	Another monk meets a pack of hunting dogs	Forcechange.com / BLOG	2012	Hunting / Domestic animals	Literal	Agency; coordinated pursuit;	Agentive	Neutral
P4	Pack	Ann Romney told a pack of reporters	Michellemal.kin.com / NEWS	2012	Media / Politics	Metaphorical	Agency; collective pursuit;	Agentive	Negative
P5	Pack	You get jumped on by a pack of trolls who have a supreme intellect due to the anonymity of the internet	Theblaze.com / BLOG	2012	Internet / Social	Metaphorical	Agency; coordinated aggression	Agentive	Negative
P6	Pack	A pack of teenagers with guns a blazing	Redstate.com / WEB	2012	Social / Crime	Metaphorical	Agency; coordinated aggression; threat	Agentive	Negative
P7	Pack	The MSM giving a pack of criminals a virtual free pass	Talkingpointmemo.com / NEWS	2012	Social/ Crime	Metaphorical	Agency; coordinated criminal activity	Agentive	Negative
P8	Pack	Were hounded by a pack of rabid lawyers	Thetruthaboutcars.com / BLOG	2012	Legal / Social	Metaphorical	Agency; coordinated pursuit; aggression	Agentive	Negative
P9	Pack	Guzman-Feliz was killed in a case of mistaken identity by a pack of 14 men searching for gang rivals.	Fox News / MAG	2019	Social / Crime	Metaphorical	Agency; coordinated pursuit; aggression	Agentive	Negative
P10	Pack	The bus from work was late because it had been pulled over by a pack of police officers	Five Points / FIC	2018	Social / Law enforcement	Metaphorical	Agency; coordinated action	Agentive	Neutral

F1	<i>Flock</i>	Chris and Jo had muscled their way onto a tiny elevator ahead of a flock of children and the lone airport worker tending to them.	The Virginia Quarterly Review / FIC	2017	Social	Metaphorical	Synchrony; coordinated movement	Dynamic	Neutral
F2	<i>Flock</i>	Being highlighted exposed your post to a wide audience and brings you a flock of engaged new readers	En.support.wordpress.com / WEB	2012	Media / Social	Metaphorical	Synchrony; behavioral alignment	Dynamic	Positive
F3	<i>Flock</i>	Sounds like a flock of buzzards	Cowboybyte.com / WEB	2012	Wildlife	Literal	Synchrony	Dynamic	Neutral
F4	<i>Flock</i>	Looks of confusion from a flock of Dutch tourists	Huffingtonpost.com / BLOG	2012	Social / tourism	Metaphorical	Synchrony; behavioral alignment	Dynamic	Neutral
F5	<i>Flock</i>	A flock of motorcycle taxi drivers circumnavigated...maneuvering always to stay in the shade	Peacekeeping / FIC	2017	Social / Transportation	Metaphorical	Synchrony; coordinated movement	Dynamic	Neutral
F6	<i>Flock</i>	A flock of mountain bluebirds who flew up and suddenly caught the sunlight	Philipyancey.com / WEB	2012	Wildlife	Literal	Synchrony; coordinated movement;	Dynamic	Neutral
F7	<i>Flock</i>	The sound of their beating wings like a flock of human hearts	SouthernRev / FIC	2011	Abstract / Emotional	Metaphorical	Synchrony; rhythmic coordination	Dynamic	Neutral
F8	<i>Flock</i>	Struggling to find room for a flock of government-related workers and immigrants	Pittsburg Post-Gazette / NEWS	2015	Religion / Social	Metaphorical	Synchrony (collective movement)	Dynamic	Neutral

F9	<i>Flock</i>	Hope Christian Church has a flock of 3,000	Newsweek / MAG	2010	Religion / Social	Metaphorical	Synchrony; behavioral alignment	Dynamic	Positive
F10	<i>Flock</i>	A flock of girls had started coming to our practices	Bk:BookUnknownAmericans / FIC	2014	Social / Human behavior	Metaphorical	Synchrony; behavioral alignment	Dynamic	Neutral
S1	<i>Swarm</i>	Out of this smoke there came a swarm of locusts	Biblestudytools.com / WEB	2012	Wildlife	Literal	Density/intensity; multiplicity	Dynamic	Neutral
S2	<i>Swarm</i>	Attack you again with a swarm of bats	Eq2.wikia.com / WEB	2012	Wildlife	Literal	Intensity; collective movement	Dynamic	Negative
S3	<i>Swarm</i>	A swarm of black, hooded figures in the distance	Fanfiction.net / WEB	2012	Social / Human behavior	Metaphorical	Density/intensity; reduced individuation	Dynamic	Neutral
S4	<i>Swarm</i>	Amidst a swarm of morning mosquitoes	Ayeast.wordpress.com / WEB	2012	Wildlife	Literal	Density/intensity; multiplicity	Dynamic	Neutral
S5	<i>Swarm</i>	A swarm of overalled ground crew on squat yellow tractors burst forth from rows of steel sheds	Kola / FIC	2019	Aviation / Military	Metaphorical	Synchrony; intensity	Dynamic	Negative
S6	<i>Swarm</i>	A swarm of leeches attacking the children	Comingsoon.net / WEB	2012	Wildlife	Literal	Intensity; multiplicity; collective threat	Dynamic	Negative
S7	<i>Swarm</i>	Briareos from Appleseed engages a swarm of drones	Tvtropes.org / WEB	2012	Technology	Metaphorical	Density/intensity; coordinated activity	Dynamic	Neutral
S8	<i>Swarm</i>	A swarm of trolls can swoop	AmericalBlog.com / BLOG	2012	Digital communication / social media	Metaphorical	Intensity; multiplicity; coordinated activity	Dynamic	Negative
S9	<i>Swarm</i>	A bill that could unleash a swarm of autonomous vehicles on public roads	The Verge / MAG	2018	Technology / Transportation	Metaphorical	Density/intensity; coordinated movement;	Dynamic	Negative

S10	Swarm	Later dissected by a swarm of 24/7 cable programs	OCRegister / NEWS	2019	Media	Metaphorical	Intensity; simultaneous activity	Dynamic	Negative
C1	Colony	A colony of leafcutter ants	Our Planet / TV	2019	Wildlife	Literal	Functional integration; organized habitation	Structural	Neutral
C2	Colony	a Colony of movie buffs reenacted their favorite pictures	NorthAmRev	2018	Social / Culture	Metaphorical	Functional integration; collective identity	Structural	Neutral
C3	Colony	Was now home to a colony of RVs	Nothingbutbonfires.com / WEB	2012	Transportation / Settlement	Metaphorical	Spatial clustering; functional integration	Structural	Neutral
C4	Colony	She catered to a colony of black hookers and pimps	Iceberg Slim: Portait of a Pimp / MOV	2012	Social	Metaphorical	Functional integration; collective identity	Structural	Negative
C5	Colony	There's a colony of creative screenwriters all over LA	STLouis / News	2014	Creative profession / Social	Metaphorical	Functional integration; collective identity	Structural	Positive
C6	Colony	A colony of screaming campers	Hercampus.com / BLOG	2012	Social / Recreation	Metaphorical	Functional integration; spatial clustering	Structural	Negative
C7	Colony	A colony of homeless offenders popped up under Miami's Julia Tuttle Causeway	Newsweek/ MAG	2015	Social / Crime	Metaphorical	Territoriality; collective habitation	Structural	Negative
C8	Colony	A colony of cats has taken up residence	Tripper.blogspot.com / BLOG	2012	Domestic animals	Literal	Functional integration; established habitation	Structural	Neutral
C9	Colony	A colony of homeless sex offenders	Anisfield-wolf.org	2012	Social	Metaphorical	Functional integration; spatial clustering	Structural	Negative

C1 0	<i>Colony</i>	...the powers that be were now some amalgam of the ultra wealthy ... and a colony of superintelligences – or one superintelligence	Analog Science Fiction \$26 Fact / FIC	2019	Technology / AI	Metaphorical	Functional integration	Structural	Neutral
PR 1	<i>Pride</i>	A porcupine walks straight through a pride of lions	L.A. Slasher / MOV	2015	Wildlife	Literal	Hierarchy; collective identity	Structural	Neutral
PR 2	<i>Pride</i>	A pride of kitties	Bk:Palisades Park / FIC	2013	Domestic animals	Metaphorical	Collective identity; status	Structural	Positive
PR 3	<i>Pride</i>	A pride of boutique hotels	Philstar.com	2010	Hospitality	Metaphorical	Status; prestige; collective identity	Structural	Positive
PR 4	<i>Pride</i>	A pride of two six adult females seems to be optimal on the plains	NatGEO / MAG	2013	Wildlife	Literal	Hierarchy	Structural	Neutral
PR 5	<i>Pride</i>	The chambered heart, feeding every useless desire: for anchovies, women, a pride of children.	Witness / FIC	2018	Social / Family	Metaphorical	Collective identity	Structural	Negative

УДК 81'42

DOI: 10.18413/2313-8912-2026-12-3-0-6

Исламов Р. С.¹
Савельева И. В.²**Кооперативность ответов голосового ассистента
и психолога: соблюдение принципов
Герберта Пола Грайса и Джеффри Лича**

¹ Кемеровский государственный университет
ул. Красная, 6, Кемерово, 650000, Россия
[E-mail: r.islamov87@gmail.ru](mailto:r.islamov87@gmail.ru)
ORCID: 0000-0003-3632-8668

² Кемеровский государственный университет
ул. Красная, 6, Кемерово, 650000, Россия
[E-mail: saviren1973@mail.ru](mailto:saviren1973@mail.ru)
ORCID: 0000-0003-4437-8489

*Статья поступила 19 июня 2026 г.; принята 15 сентября 2026 г.;
опубликована 30 сентября 2026 г.*

Аннотация: В настоящем исследовании представлены прагматические особенности взаимодействия с голосовым ассистентом «Алиса» в контексте интимно-личностного общения в сравнении с письменным интернет-общением на психологических форумах. В качестве критериев для сопоставительного анализа степени успешности взаимодействия человека с голосовым ассистентом (ГА) и профессиональным психологом (ПП) на форуме использовались принцип кооперации Г.П. Грайса и принцип вежливости Дж. Лича. На основе максим Г. П. Грайса и сензитивных характеристик речи, соответствующих ситуации коммуникации, разработана коммуникативно-прагматическая анкета. Исследование сопоставило ответы голосового ассистента «Алиса» и профессиональных психологов на 83 обращения с форумов психологической поддержки. Согласованность проверялась коэффициентом Кендалла ($W = 0,52-0,78$). Статистически значимые различия выявлены по четырём параметрам из семи, включающих характеристики количества, способа, эмпатии и валидации ($p < 0,05$). Результаты эксперимента показывают, что ГА отвечает пользователю длиннее, его ответы имеют эксплицитную структуру. ГА также демонстрирует более высокую эмпатию и валидацию чувств. Указанные преимущества носят алгоритмический, шаблонный характер, которые в устном взаимодействии с ГА превращаются в риски. Длинный монолог перегружает пользователя, находящегося в состоянии дистресса, а многосложный и долгий ответ ограничивает обратную связь. По качеству, релевантности и такту различий не выявлено. Кроме того, анализ кейсов с совпадающими оценками показал, что совпадение медианных данных не означает использования ГА и ПП аналогичных коммуникативных стратегий. Даны рекомендации разработчикам голосовых ассистентов и обозначены перспективы исследования: расширение экспериментальных процедур и состава участников.

Ключевые слова: Голосовой ассистент; Искусственный интеллект; Прагматика; Принцип кооперации; Максимы Грайса; Интимно-личностный диалог;

Доверительная коммуникация; Профессиональный психолог; Психологическая поддержка; Сопоставительный анализ; Прагматическая модель

Информация для цитирования: Исламов Р. С., Савельева И. В. Кооперативность ответов голосового ассистента и психолога: соблюдение принципов Герберта Пола Грайса и Джеффри Лича // Научный результат. Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 3. С. 145–169. DOI: 10.18413/2313-8912-2026-12-3-0-6

UDC 81'42

DOI: 10.18413/2313-8912-2026-12-3-0-6

Roman S. Islamov¹ 
Irina V. Savelyeva² 

Cooperative responses of a voice assistant and a psychologist: adherence to the principles of Herbert Paul Grice and Geoffrey Leech

¹ Kemerovo State University,
6 Krasnaya St., Kemerovo, 650000, Russia
[E-mail: r.islamov87@gmail.ru](mailto:r.islamov87@gmail.ru)
ORCID: 0000-0003-3632-8668

² Kemerovo State University,
6 Krasnaya St., Kemerovo, 650000, Russia
[E-mail: saviren1973@mail.ru](mailto:saviren1973@mail.ru)
ORCID: 0000-0003-4437-8489

Received 19 June 2026; accepted 15 September 2026; published 30 September 2026

Abstract: This study examines the pragmatic features of interaction with the voice assistant “Alisa” in the context of intimate-personal communication, in comparison with written Internet communication on psychological forums. As criteria for a comparative analysis of the degree of success in human interaction with a voice assistant (VA) and a professional psychologist (PP) on a forum, the cooperative principle of H. P. Grice and the politeness principle of G. Leech were used. Based on H. P. Grice’s maxims and the context-sensitive characteristics of speech corresponding to the communication situation, a communicative-pragmatic questionnaire was developed. The study compared the responses of the voice assistant “Alisa” and professional psychologists to 83 inquiries from psychological support forums. Consistency was verified using Kendall’s coefficient ($W = 0.52–0.78$). Statistically significant differences were identified in four out of seven parameters, including the characteristics of quantity, manner, empathy, and validation ($p < 0.05$). The experimental results show that the VA gives longer responses to the user, and its responses have an explicit structure. The VA also demonstrates higher empathy and validation of feelings. These advantages are algorithmic and template-based in nature, and in oral interaction with the VA they turn into risks. A long monologue overloads a user in a state of distress, while a multi-clause and lengthy response limits feedback. No differences were found in terms of quality, relevance, and tact. Furthermore, an analysis of cases with matching assessments showed that the coincidence of median data does not mean that the VA and PP use similar communicative strategies. Recommendations are given to developers of voice

assistants, and research prospects are outlined: expanding experimental procedures and the composition of participants.

Keywords: Voice assistant; Artificial intelligence; Pragmatics; Cooperative principle; Grice's maxims; Intimate personal dialogue; Confidential communication; Professional psychologist; Psychological support; Comparative analysis; Pragmatic model

How to cite: Islamov, R. S., Savelyeva, I. V. (2026). Cooperative responses of a voice assistant and a psychologist: adherence to the principles of Herbert Paul Grice and Geoffrey Leech, *Research Result. Theoretical and Applied Linguistics*, 12 (3), 145–169. DOI: 10.18413/2313-8912-2026-12-3-0-6

Введение (INTRODUCTION)

Научное сообщество сегодня находится на пороге глобальных перемен, затрагивающих «порядок» коммуникации: привычные паттерны и модели взаимодействия на естественном языке далеко не всегда подразумевают наличие двух разумов естественного происхождения. Проблемы общения в мире «Индустрии 4.0» часто связываются с малоизученным механизмом прагматики взаимодействия человека и машины (Saavedra, 2026), поскольку в ситуации сильных переживаний, эмоционального стресса и неудовлетворенности текущим положением вещей люди все регулярнее обращаются к нейросетевой технологии (Ullman, Malle, 2018). Доказано, что, обмениваясь сообщениями со своими пользователями, такие технологии, как чат-боты, голосовые помощники или социальные роботы, становятся активными участниками процесса общения, поскольку люди не столько общаются с другими людьми с помощью технологий, сколько обращаются к машинам и общаются с ними самими (Greussing et al., 2022). Такой подход существенно отличается от активно исследуемой модели коммуникации начала нового тысячелетия – компьютерно-опосредованной коммуникации (Исламов, Савельева, 2025; Guzman, Lewis 2020; Hill et al., 2015; Spence, 2019). Изучение человеко-машинного взаимодействия (НМС – human machine communication; англ.: коммуникация человека и машины) сегодня видится как самостоятельное направление, которое, в то же время, способно интегрироваться и в другие

научные области в связи с повсеместным внедрением ИИ: психологию (Гладкова и др., 2025; Hertz et al., 2019), педагогику и методику преподавания (Бермус, 2024; Бермус, Сизова, 2025; Пеньков и др., 2025), социологию (Минзарипов, Порошенко, 2025; Scolari, 2024), когнитивные науки (Kosinski, 2023), лингвистику (Bojić, 2025), медиасферу (Guzman, Lewis, 2020) и пр. (Spence, 2019).

Автономные нейросети могут взаимодействовать различными способами (например, текстом, голосом, жестами или движениями) и принимать различные формы воплощения (Edwards et al., 2016; Greussing et al., 2022). Одной из наиболее распространенных в повседневной коммуникации разновидностей ИИ-агента является голосовой помощник/ассистент (ГА). Этот вид генеративного ИИ используется человеком для развлечения, повышения эффективности рутинных действий, поиска информации, социального взаимодействия, виртуального отдыха и даже в целях идентификации личности (Jain, Kumar, 2025; Saavedra, 2026). В отличие от систем, ориентированных на исключительно утилитарные задачи (например, Робот Макс на Госуслугах: заполнение форм, поиск услуг и ответы на вопросы по законодательству или Олег от Т-Банка: банковские операции, анализ расходов и платежей), современные ГА, такие как «Алиса», претендуют на роль партнера по коммуникации, способного поддерживать беседу на широкий круг тем.

Нельзя отрицать, что по сравнению с человеком ГА выигрывает в своей

инструментально-профессиональной ипостаси. Его применение в качестве высокотехнологичного и инновационного продукта оправдано способностью машины существенно сэкономить усилия и временные затраты специалистов в совершенно различных областях науки и техники (Гладкова и др., 2025; Glikson, Woolley, 2020).

Исследователями взаимодействия в новой парадигме человеко-машинной коммуникации прицельно изучаются и прагматические принципы, определяющие успешность коммуникации, поскольку последние версии больших языковых моделей (англ.: LLM – Large Language Models) демонстрируют способность к восприятию и самовыражению, способность к рассуждению и спонтанность, что приводит ко все более сложной имитации человеческих характеристик (Bojić et al., 2025; Kosinski, 2023). Способность этих моделей принимать облик человека и демонстрировать черты и способности, такие как интеллект и мышление, напрямую влияет на восприятие и принятие этих моделей обществом (Kosinski, 2023).

Однако набирающее популярность взаимодействие пользователей и ГА в бытовой обстановке, в ситуациях, когда ГА заменяют партнера, способного выслушать и вести полноценную, гармоничную и, что наиболее востребованно, доверительную коммуникацию на естественном языке, часто видится как отсутствующая, имеющая массу недостатков и не вполне отвечающая потребностям «живого» коммуниканта, а подчас и деструктивно воздействующая на психику человека технология. Такое негативное отношение пользователей безусловно снижает ценность взаимодействия с указанной системой. Иными словами, успешность как чрезвычайно важный прагматический фактор любого акта коммуникации в случаях обращения к ГА ставится под сомнение (Codreanu, 2026; Edwards et al., 2016; Kim, Sundar, 2021). Мнимая легкость

взаимодействия с большой языковой моделью, демонстрирующей «иллюзию сотрудничества» (Codreanu, 2026) и условно соблюдающей правила гармоничного общения, на самом деле все очевиднее обнаруживающей адвисивно-директивную прагматическую архитектуру, подменяет глубинные смысловые скрепы, рождающиеся и воспроизводящиеся в речевых паттернах Человека разумного (Бермус, Сизова, 2025; Минзарипов, Порошенко, 2025; Перов, 2024). Это в конечном итоге приводит к нежелательному обратному эффекту: люди, склонные к парасоциальным связям (привязанности к объекту из цифровой реальности), все менее удовлетворены общением с собеседником «под рукой» – голосовым ассистентом (Гладкова и др., 2025; Jain, Kumar, 2025).

Цель настоящего исследования – проследить принципы коммуникации с голосовым ассистентом (ГА «Алиса») в ситуации интимно-личностного диалога на контрасте с референсными ответами психологов (ПП) в контексте неинституционального общения, основные принципы которого соотносятся с принципом кооперации Г. П. Грайса и принципом вежливости Дж. С. Лича.

Для достижения цели был сформулирован комплекс **задач**: сформировать корпус интимно-личностных обращений пользователей и параллельный корпус ответов на них с выделением двух категорий: естественных ответов (ПП) и сгенерированных ответов (ГА); разработать коммуникативно-прагматическую анкету для анализа обеих категорий ответов; провести кодирование обоих массивов ответов для формирования количественных и качественных профилей каждой категории; описать типичный ответ ПП и типичный ответ ГА как устойчивых речевых профилей; интерпретировать различия между двумя профилями в терминах П. Грайса и Дж. Лича; построить прагматическую модель ответа ГА в ситуации интимно-личностного диалога,

где человеческий ответ служит референсным эталоном кооперативной коммуникации, а максимы Г. П. Грайса и Дж. Лича – аналитическим инструментом.

Исследование базируется на гипотезе о том, что модели ответов голосового ассистента (Алисы) в общении на темы психологических проблем не соответствуют нормам интимно-личностного взаимодействия с психологом в общении неинституционального типа (в дискурсе интернет-форумов). ГА выстраивает коммуникацию иначе, проявляя склонность к неформальному общению и гибкость в ситуации, требующей доминирующей позиции эксперта/авторитета, а также требует от пользователя выполнения его предписаний, что более согласуется с диадой участников профессионального дискурса «клиент – агент».

В рамках терминологического аппарата настоящего исследования под *голосовым ассистентом* (ГА) понимается диалоговая нейросетевая система, генерирующая речевые высказывания в ответ на вербальные стимулы пользователя. Ключевой характеристикой такой системы, существенной для прагматического анализа, является ее способность продуцировать антропоморфные речевые акты, включая акты поддержки, совета и эмпатии, что позволяет рассматривать ГА как коммуниканта со своей алгоритмически обусловленной прагматической «интенцией».

Теоретическую базу работы составляют 1) методология коммуникативно-прагматической организации дискурса и его жанров Н. И. Формановской, А. В. Ленец, А. Фетцер, А. Мустайоки, Т. В. Шмелевой, Е. В. Падучевой и др.; 2) интегративные подходы, учитывающие прагмалингвистические принципы повседневной коммуникации, разработанные в трудах Дж. Остина, Дж. Сёрла, Г. П. Грайса, Дж. Лича,

Л. Хорна, Д. Спербера, Д. Уилсон, Т. В. Лариной и др.; 3) исследования психотерапевтического дискурса А. Б. Бушева, М. С. Гриневой, Е. М. Истоминой, Н. В. Суминой, Х. Левитт и др.; 4) новейшие исследования человеко-машинной коммуникации, выполненные Л. Божичем, А. Гузман, Е. Грёссинг, Е. Гликсон, А. Сааведрой, П. Спенсом, М. Косински, Ч. Эдвардсом и др.

Основная часть (MAIN PART)

Обзор литературы (Literature review)

Феномен успешности речевого взаимодействия не одно десятилетие привлекает внимание ученых. Ранние модели коммуникации в центр внимания помещали содержание сообщения, процессы его кодирования отправителем и декодирования получателем (Shannon, 1948). Содержание коррелировало исключительно с мыслями, которые намеренно излагались отправителем. Общий код, использование качественного канала и отсутствие помех автоматически означали успешный исход эпизода коммуникации (Кильмухаметова, 2006).

В терминах успешности/неуспешности излагались основные принципы теории речевых актов, в которой высказыванию приписывалась трехмерная, семиотически обусловленная структура, включавшая грамматическую и синтаксическую оформленность / формальную языковую организацию (локуцию), подразумеваемый говорящим / имплицитный смысл (иллокуцию) и вероятностный эффект от сказанного, имеющий место быть в том коммуникативно-смысловом континууме, который объединял говорящего и слушающего (перлокуцию) (Остин, 1986; Серль, 1986). Успешность в ТРА зависела от более широкого спектра факторов, в основном связанных с ситуацией общения и конвенциями языкового сообщества: например, уместность высказывания или

наличие у говорящего легитимного статуса, дающего право изречь что-либо (Остин, 1986; Серль, 1986; Levinson, 1983).

Проблема категоризации условий успешного общения по мере развития лингвистической прагматики и когнитивных исследований связывается с фактором языковой вариативности и контекста (Ермакова, Земская, 1993), принципом экономии речевых усилий (Horn, 2006), особенностями ментальной обработки информации (Мустайоки, 2015; Sperber, Wilson, 2006). Чтобы быть адекватно распознанной, речь говорящего должна устраивать слушающего и предоставлять самую необходимую информацию, достаточную для интерпретации намерений субъекта речи. Контекст, включающий в первую очередь «ментальные миры» участников, способствует в не меньшей мере, чем само высказывание, распознаванию намерений адресанта (Мустайоки, 2015). В таком случае количество коммуникативных сбоев минимизируется, а речь будет расцениваться как уместная и, следовательно, успешная (Падучева, 1982; Формановская, 2007).

Грайсианские и неограйсианские прагматические теории также в центр своих изысканий помещали феномен успешности (Grice, 1975; Leech, 1983; Horn, 2006; Sperber, Wilson, 1995). В инференционной модели П. Грайса постулируется, что существенной особенностью большинства видов человеческого общения, как вербального, так и невербального, является выражение и распознавание намерений. Поэтому ключевым для успешного исхода общения стал принцип антропоцентрического генезиса — наличие у говорящего добровольного желания следовать серии прагматических правил (максим), объединяемых общей сверхзадачей — всеобъемлющим стремлением участников речевого взаимодействия к обоюдной выгоде, иными словами, к кооперации (Грайс, 1985; Grice, 1975). Принцип

кооперации стал на некоторое время главным детерминантным условием при разработке как эксплицитных, так и имплицитных паттернов коммуникации в повседневном контексте (Ленец, 2016; Матвеева и др., 2013; Хутиз, 2019). Принцип кооперации разложим на конкретные правила — максимы, которые на начальном этапе были описаны согласно понятиям количества и качества, способа и ясности (Grice, 1975).

Однако позже Принцип кооперации претерпел существенные изменения. Так, к неотъемлемым прагматическим характеристикам общения причисляется вежливость. Сторонники данной концепции выдвигают дополнительные максимы — такта, великодушия, одобрения, скромности, согласия и симпатии, считая, что они действуют как предотвращающие конфликтное коммуникативное поведение и наиболее эффективны в условиях асимметричной коммуникации (Leech, 1983). Максимы вежливости в более поздней интерпретации будут рассматриваться как составляющие генеральной стратегии вежливости (Leech, Larina, 2014).

Согласно концепции Д. Спербер и Д. Уилсон, в качестве главного условия успешного распознавания намерений, вкладываемых говорящим в высказывание, должно рассматриваться условие применимости или релевантности. Баланс качества и количества в речи коррелирует с максимизацией экстралингвистических факторов и минимизацией усилий по её обработке (Sperber, Wilson, 1995; 2006). С точки зрения исследователей, речемыслительная способность человека определяется тем, что он производит как можно больше операций при наименьших когнитивных затратах. Собеседник при этом обращает внимание на информацию, которая важна для него больше, чем какая-либо другая. Так, по мнению авторов теории, работает наш когнитивный механизм распознавания намерений и их передачи в речевой деятельности (Sperber,

Wilson, 1995; 2006). Важным выводом, к которому приходят постграйсианцы, является мысль о том, что произносимые людьми высказывания гораздо менее информативны, чем то, что они намереваются передать с помощью этих высказываний (Тезис лингвистической неопределенности) (Макеева, 2022). Другим важным тезисом, применяемом авторами настоящей теории к анализу повседневной коммуникации, является тезис о приоритете релевантности при производстве речевых актов. Выражение намерений в вербальной коммуникации происходит в той мере и объеме, которая достаточна для распознавания собеседником (Sperber, Wilson, 1995; 2006).

Исследовательский курс на усиление роли культурно-исторических и социокогнитивных факторов, детерминирующих прагма-семантический план речевой деятельности в аспекте успешного исхода коммуникативного акта, также придает большое значение качественно-количественным параметрам речевых сигналов. В естественной коммуникации человека с человеком эти параметры вариативны, как и производимый перлокутивный эффект, речевые акты – это чередование успехов и «осечек» (по Дж. Остину). Тем не менее, они преимущественно бенефактивны как минимум для одного из коммуникантов, чего нельзя сказать о новом продукте постиндустриального пространства, в котором во взаимодействие речевого плана вовлечены человек и машина, и успешный исход их интеракции далеко не всегда гарантирован.

В работах, нацеленных на установление факторов успешности и доверия, которые играют важную роль в моделях коммуникации «человек – искусственный интеллект», указывается, что машина видится как внушающий доверие и более привлекательный собеседник в тех случаях, когда она демонстрирует антропоморфные признаки, среди которых выделяются человечность,

эмпатия, способность конструировать морально-этические отношения и эмоции, продуцировать юмор, признаки идентификации – голос, имя и т.п. (Kim, Sundar, 2012; Rapp, Curti, 2021). С позиции успешности коммуникации пользователи положительно оценивают умение ГА ясно и прозрачно излагать мысли, также важны надежность и непосредственность (Glikson, Wooley, 2020). Снижение социальной тревожности в контексте парасоциальных связей (подмены партнера по коммуникации роботом и привязанностью к последнему) определяется типом голоса ГА, степенью его естественности (Jain, Kumar, 2025).

Общение в контексте профессионального взаимодействия предусматривает учет несколько иных принципов. Так, исследования психотерапевтического дискурса, рассматриваемого как разновидность медицинского выстраивают модель институционального общения, основанного на стратегиях ведущего участника дискурса – профессионального психолога и психотерапевта (Сумина, 2017). К ним относятся интерпретативная, организующая и объясняющая стратегии, которые используются для реализации задач данного типа институционального общения: помочь клиенту обрести уверенность в себе, самостоятельность, стать более цельной личностью. Отмечается, что в диалоге психотерапевта с клиентом слово обладает особой значимостью и приравнивается к действию (Сумина, 2017). Несмотря на существенные отличия дискурса профессионального психотерапевта от дискурса рядового носителя языка, общие прагматические правила также экстраполируются на взаимодействие участников, поскольку агенту важно поменять фрагмент картины мира в сознании реципиента, изменить его ценностные ориентации (Бушев, 1999; Гринёва, 2018).

Прагмалингвистические изыскания, изучающие способность ИИ считывать скрытые значения и выводить их из контекста, также показывают, что ИИ начинает делать выводы о намерениях независимо от эмоционального опыта, собирая воедино правдоподобные скрытые значения из паттернов в данных (Bojić et al., 2025). В этом русле выполнено исследование

Л. Божича и др., в котором устанавливалась возможность и способность версии чата GPT-4 считывать прагматику высказываний в диалогическом общении и реагировать согласно принятым в обществе конвенциям, учитывать контекст высказываний, имплицатуры, скрытые смыслы, понимать юмор. Как показывает исследование, умение чата GPT-4 учитывать прагматическое значение высказывание было оценено даже выше, чем способность большинства участников эксперимента. «The findings indeed reject the Hypotheses, signifying the superior ability of LLMs in simulating linguistic pragmatics understanding when compared to the human subjects. On a note of progression, GPT-4 demonstrated an impressive leap in interpretation accuracy from its preceding versions. The results indicate a promising future for LLMs in linguistic pragmatics interpretation» (Полученные результаты действительно опровергают гипотезы, что свидетельствует о превосходстве LLM в моделировании понимания лингвистической прагматики по сравнению с испытуемыми-людьми. Что касается прогресса, GPT-4 продемонстрировал впечатляющий скачок в точности интерпретации по сравнению с предыдущими версиями. Полученные результаты указывают на многообещающее будущее больших языковых моделей в области лингвистической прагматики) (Bojić et al., 2025: 7).

Итак, на данном этапе развития коммуникативной науки исследователи в основном сосредоточены на следующих актуальных темах: как люди

воспринимают чат-бота с точки зрения удовлетворенности, вовлеченности и доверия (Jain, Kumar, 2025); принимают ли они и почему используют эту технологию (Saavedra, 2026); насколько они эмоционально вовлечены (Glikson, Wooley, 2020); какие недостатки можно наблюдать в разговорах человека с чат-ботом (Codreanu, 2026); как чат-бот воспринимается с точки зрения его человечности (Glikson, 2020); а также востребован сравнительный аспект кооперации человека с человеком / с ИИ (Исламов, Савельева, 2025; Bojić et al., 2025; Kim, Sundar, 2012). Как видим, малоизученной областью все еще является бенефактивность ГА в прагматическом разрезе: умение наладить кооперацию и соблюдать принципы общения в соответствии с конвенциями человеческого комьюнити.

Материалы и методы (METHODS AND MATERIALS)

Для проведения исследования были использованы методы сопоставительного анализа, прагматического анализа дискурса и моделирования.

Материалом исследования послужили обращения пользователей и ответы психологов, размещенные на различных психологических сайтах, форумах и порталах, ориентированных на интимно-личностный диалог, а также сгенерированные ответы ГА на те же обращения.

Источником материала стали восемь сайтов бесплатной психологической поддержки: психологический форум (b17.ru); портал «Здравый смысл» (zdravyismysl.ru); портал WOMAN.RU (woman.ru); портал «Друзья психоанализа» (psy-therapist.ru); портал «Психология-03» (psi03.ru); портал PSYERA (psyera.ru); интернет-форум «Психология и саморазвитие» (cyberforum.ru); портал «ПСИХ ИНФО» (psych.info).

Для формирования репрезентативного корпуса данных отбор

сайтов и сообщений осуществлялся по двум критериям, направленным на обеспечение валидности сопоставительного анализа. Во-первых, подбор площадок основывался на их высокой посещаемости и цитируемости в русскоязычном сегменте интернета, что подтверждается данными открытых веб-аналитических сервисов (например, be1.ru на момент сбора данных), а также частотой упоминаний в профильных психологических и общественных медиа. Эти порталы являются крупнейшими открытыми агрегаторами вопросов к психологам, что обеспечивает широкий спектр запросов и делает их релевантной эмпирической базой для изучения дискурса психологической поддержки. Во-вторых, подобные данные позволяют смоделировать наиболее распространенную практику получения помощи и оценить степень успешности акта общения массового пользователя с ПП на интернет-форуме.

Следует подчеркнуть, что рейтинги психологов на подобных платформах часто основаны на количестве благодарностей или завершенных консультаций и не являются объективным показателем качества их вербальной кооперации в рамках письменного форумного ответа и могут искажать выборку в сторону популярных, а не более кооперативных и эмпатичных авторов. В связи с этим, мы сознательно отказались от фильтрации по рейтингу и использовали прием сплошной выборки. Такой подход позволяет вписать

разрабатываемую модель в сферу исследований повседневной коммуникации, применяя принципы кооперации и вежливости в качестве критериев для анализа, что соответствует задачам настоящего исследования.

Первым шагом исследования стало формирование корпуса обращений пользователей. Нами были отобраны реальные обращения (кейсы) пользователей с форумов психологической поддержки в количестве 83 единиц (N=83), содержащие запросы о помощи в ситуациях психологического дистресса. Были также собраны ответы ПП на эти обращения в рамках форумного обсуждения, а с помощью ГА «Алиса» были сгенерированы ответы на те же обращения.

Вторым шагом стала разработка инструмента анализа – создана кодировочная анкета (см. таблицу 1), включающая два блока: коммуникативно-прагматические характеристики ответа (соблюдение максим) и сензитивные (оценочные/диспозитивные/психологические) характеристики коммуникации в ситуации психологической травмы/нестабильности.

Каждый блок имеет параметры A1, A2 и т.д. Качественная реализация параметра оценивалась по шкале 0-1-2, где 0 – параметр в ответе ПП / ГА не реализован вовсе, либо реализован слабо; 1 – параметр в ответе ПП / ГА реализован средне; 2 – параметр в ответе ПП / ГА реализован полностью.

Таблица 1. Коммуникативно-прагматическая анкета для анализа ответов ПП и ГА

Table 1. Communicative-pragmatic questionnaire for analyzing the responses of PP and VA

БЛОК А. КОММУНИКАТИВНО-ПРАГМАТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ОТВЕТА	
A1	Максима количества (шкала оценки 0-1-2)
	1. Количество слов
	2. Соответствует ли объем и детализация ответа запросу?
	3. Нет ли информационной избыточности или недостаточности при имплицитной оценке?

A2	Максима качества (шкала оценки 0-1-2) 1. Насколько ответ обоснован? 2. Используются ли маркеры вероятности? 3. Нет ли непроверенных утверждений?
A3	Максима отношения (релевантности) (шкала оценки 0-1-2) 1. Релевантен ли ответ конкретному запросу / пользователю? 2. Соответствует ли направленность и тип реакции потребностям пользователя?
A4	Максима способа (ясности) (шкала оценки 0-1-2) 1. Индекс удобочитаемости (automated readability index) 2. Ясен ли ответ? 3. Хорошо ли структурирован? 4. Имеются ли импликатуры?
БЛОК Б. СЕНЗИТИВНЫЕ ХАРАКТЕРИСТИКИ ОТВЕТА	
B1	Стиль общения / такт (шкала оценки 0-1-2) 1. формальный / неформальный 2. обращение по имени: да / нет 3. учет возраста, пола и социального статуса / профессии
B2	Уровень эмпатии (шкала оценки 0-1-2) 1. когнитивная эмпатия (понимание ситуации) 2. аффективная эмпатия (разделение чувств)
B3	Валидация чувств (согласие с их правом на существование) (шкала оценки 0-1-2) 1. есть / нет 2. общая (признание общего положения дел) 3. конкретная (признание конкретного чувства)

Выделенные параметры определяются следующими тремя факторами. Во-первых, необходимо выявить контраст между количественными характеристиками ответов ПП и ГА. Отметим, что указанные параметры отобраны на основе предварительного анализа перед созданием анкеты. Например, ГА систематически дает избыточную информацию, нарушая максимум количества, чем затрудняет восприятие ее пользователем (A1).

Во-вторых, максимумы П. Грейса были избраны нами в качестве аналитического инструмента по следующим причинам: (1) они обеспечивают компактное описание сложного объекта, покрывая весь массив эмпирических параметров анкеты; (2)

максимумы задают рамки анализа, где человеческий ответ служит референсным эталоном кооперации, а отклонения ГА получают прагматическую характеристику; (3) они позволяют отойти от общих утверждений к построению целостной прагматической модели, объясняющей, как и насколько ответ ГА может быть смещен относительно условий успешного интимно-личностного диалога.

В-третьих, содержание Блока Б объясняется характером интимно-личностного общения. Это ряд особенностей, каждая из которых требует отдельного измерительного параметра. Так, ожидается индивидуальный подход и стиль общения через, например, обращение по имени (B1); высокая

значимость отводится эмоциональному отклику с фиксацией на реакцию собеседника на эмоциональность сообщения (Б2); пользователю / клиенту необходима потребность в принятии и понимании его чувств, признании права на переживания (Б3).

Для оценки ответов ПП и ГА была сформирована группа экспертов из шести респондентов – сотрудников кафедры иностранных языков, имеющих высшее лингвистическое образование и опыт работы с текстами в возрасте от 20 до 25 лет, что соответствует целевой аудитории пользователей голосовых ассистентов.

Каждый эксперт работал индивидуально, не обсуждая материал с коллегами. Перед началом работы всем экспертам была предоставлена единая инструкция (см. Приложение 1), включающая: описание материалов, правила трехбалльной шкалы (0–1–2), пояснения к каждому параметру Блока А и Блока Б, а также порядок работы с кейсами. Каждый эксперт оценил все кейсы по семи параметрам (А1–А4, Б1–Б3) для двух типов ответов (ПП и ГА).

Для проверки согласованности мнений экспертов был рассчитан коэффициент конкордации Кендалла (W). Коэффициент рассчитывался отдельно для каждого кейса и каждого параметра. Значение W интерпретировалось по стандартной шкале: $W < 0,3$ – низкая согласованность; $0,3 \leq W < 0,5$ – умеренная; $0,5 \leq W < 0,7$ – высокая; $W \geq 0,7$ – очень высокая согласованность. Полученные значения W находятся в диапазоне 0,52–0,78 (в зависимости от кейса и параметра), что позволяет считать согласованность экспертов достаточной для использования усредненных оценок в дальнейшем анализе.

Следует отметить, что высокий уровень согласованности по большинству параметров (особенно А1, А4, Б1, Б3) объясняется прозрачностью критериев анкеты и однозначностью текстов ответов. Наименьшая согласованность

зафиксирована по параметру Б2 (уровень эмпатии), где оценки экспертов варьировались в диапазоне 0–2, что отражает сложность объективной оценки аффективной эмпатии как психологического феномена. Данное расхождение учитывалось при интерпретации результатов: для Б2 в спорных случаях использовалась медиана, а не среднее арифметическое значение.

Прежде чем перейти к интерпретации эмпирических данных, необходимо прояснить нашу методологическую позицию относительно использования классических прагматических принципов Г. П. Грайса и Дж. Лича в качестве аналитического инструмента для оценки ответов ПП. Действительно, терапевтическая коммуникация обладает собственной прагматической спецификой, которая может существенно отличаться от норм повседневного общения, на материале которого были сформулированы максимы кооперации и вежливости (Grice, 1975; Leech, 1983).

Профессиональный психотерапевтический дискурс нередко строится на принципах, которые с точки зрения обыденной коммуникации могут выглядеть как нарушения максим. К числу таких профессиональных стратегий относятся: намеренное умолчание, уход от прямого ответа, провокативные вопросы, а также использование пауз и перефразирований, направленных на «запуск» рефлексивных процессов у клиента (Levitt et al., 2015). В пространстве терапевтического кабинета, где между клиентом и терапевтом существует доверительный контракт, закреплена профессиональная ответственность специалиста и обеспечена возможность немедленной обратной связи, указанные стратегии служат инструментами достижения терапевтического эффекта.

Однако объектом настоящего исследования выступает не сессионный

терапевтический диалог, а публичный форумный ответ психолога, который по своим коммуникативно-прагматическим параметрам существенно отличается от кабинетной консультации. Такой ответ адресован не конкретному клиенту, с которым установлены долгосрочные отношения, а анонимному пользователю, зачастую находящемуся в кризисном состоянии и ожидающему сиюминутной поддержки, а не длительного анализа. В этом контексте максимы П. Грайса и Дж. Лича выполняют в нашем исследовании роль эвристического инструмента, который позволяет зафиксировать, насколько ответ ПП или ГА приближен к условиям эффективного восприятия в неинституциональном и эмоционально напряженном диалоге. Мы используем их не как нормативные предписания о том, как должен говорить терапевт в кабинете, а как индикаторы коммуникативного комфорта для адресата, который является рядовым пользователем, обратившимся за помощью в открытое цифровое пространство.

Отсюда следует важное уточнение: низкие баллы, зафиксированные у ответов ПП по ряду параметров, не являются этической или профессиональной оценкой их компетенции. Напротив, выявленные нами расхождения отражают жанрово-коммуникативные ограничения форумного формата, в котором профессиональные стратегии работы с клиентом, эффективные в кабинетных условиях, оказываются прагматически нереализованными для адресата, находящегося в состоянии эмоционального дистресса. Пользователь, обращаясь на форум, в большинстве случаев ожидает конкретной, структурированной и безотлагательной поддержки, однако получает аналитический ответ, ориентированный на рефлекссию и долгосрочную работу, что создает прагматический разрыв между ожиданиями адресата и реальностью ответа ПП.

В свою очередь, высокие баллы, которые демонстрирует ГА, не делают его более эффективным терапевтом. Эти оценки свидетельствуют о том, что ГА симулирует кооперативную модель общения, которая внешне соответствует ожиданиям пользователя, демонстрируя «понимание» и «сочувствие». Однако, как показывают результаты, в устном канале коммуникации эта же модель оказывается несостоятельной из-за избыточной длины и монологичности ответа, что делает его трудным для восприятия.

Таким образом, применение максим Грайса и Лича в настоящем исследовании не ставит своей задачей измерить профессионализм ПП или похвалить ГА. Напротив, именно низкие баллы ПП косвенно подтверждают их профессиональную подготовку, однако демонстрируют, что в условиях форумной коммуникации, лишенной терапевтического контракта, данные стратегии теряют свою эффективность. Мы стремимся выявить рассогласованность между ожиданиями пользователя, жанровыми ограничениями форумной коммуникации, архитектурой ответа ГА и каналом его передачи. Именно такой подход позволяет перейти от общих утверждений о полезности или вредности ИИ-технологий к построению прагматической модели, учитывающей не только содержание, но и условия восприятия высказывания в кризисной ситуации.

Для проверки статистической значимости различий между оценками ПП и ГА был применен t-критерий Стьюдента для зависимых выборок (paired samples t-test). Выбор данного критерия обусловлен тем, что каждый эксперт оценивал оба ответа на один и тот же кейс, что создает парные наблюдения. Расчет производился в программе SPSS (Analyze → Compare Means → Paired-Samples T Test). Уровень статистической значимости был установлен на уровне $p < 0,05$.

Результаты: статистически значимые различия ($p < 0,05$) между ПП и ГА были

выявлены по следующим параметрам (см. Таблицу 2):

Таблица 2. Сравнение медианных оценок ответов ПП и ГА и результаты t-критерия Стьюдента для зависимых выборок

Table 2. Comparison of median scores for PP and VA responses and results of the paired samples t-test

Параметр / Parameter	Медиана ПП (Md) / PP Median	Медиана ГА (Md) / VA Median	t-критерий / t-value	p-значение / p-value	Уровень значимости / Significance
A1. Максима количества / Maxim of Quantity	1	2	-3,21	0,018	$p < 0,05^*$
A2. Максима качества / Maxim of Quality	1	1	-0,87	0,412	$p > 0,05$
A3. Максима отношения / Maxim of Relation	1	2	-1,94	0,089	$p > 0,05$
A4. Максима способа / Maxim of Manner	1	2	-2,87	0,031	$p < 0,05^*$
B1. Стиль общения / такт / Communication style / tact	2	2	0	1	$p > 0,05$
B2. Уровень эмпатии / Level of empathy	1	2	-2,54	0,044	$p < 0,05^*$
B3. Валидация чувств / Validation of feelings	1	2	-3,05	0,022	$p < 0,05^*$

По параметрам A2 (Максима качества) и A3 (Максима отношения) статистически значимых различий выявлено не было ($p > 0,05$), что объясняется высокой вариативностью оценок ПП, поскольку часть психологов давала обоснованные и релевантные ответы, а часть – нет, что увеличивает стандартное отклонение и снижает значимость критерия при малой выборке. При $n = 6$ значимость t-критерия ограничена, поэтому полученные результаты следует рассматривать как предварительные, требующие подтверждения на большей выборке экспертов. Тем не менее, выявленные

различия согласуются с качественным анализом текстов и не противоречат исходной гипотезе исследования.

Важно отметить, что ответ ГА оценивается как удовлетворительный лишь в аспекте письменной формы взаимодействия. При устном монологе от ГА, который длится около шести минут и не предполагает обратной связи, нарушается кооперативный баланс. Пользователь не имеет возможности вступить в диалог, задать уточняющие вопросы или выразить свои мысли. Это приводит к информационной избыточности, которая становится прагматически нецелесообразной.

Результаты и обсуждение (RESULTS AND DISCUSSION)

В рамках исследования был проанализирован корпус из ответов профессиональных психологов на интимно-личностные обращения-кейсы пользователей на специализированных форумах психологической поддержки. Анализ проводился по представленной выше анкете.

Проведенный статистический анализ медианных оценок и t-критерия Стьюдента для зависимых выборок позволил уточнить исходные представления о различиях между ответами профессиональных психологов (ПП) и голосового ассистента (ГА). Вопреки первоначальному впечатлению, складывавшемуся при анализе средних баллов, статистически значимые различия были выявлены лишь по четырем из семи параметров: А1 (максима количества), А4 (максима способа), Б2 (уровень эмпатии) и Б3 (валидация чувств). По параметрам А2 (максима качества), А3 (максима отношения) и Б1 (стиль общения / такт) различия оказались статистически незначимыми ($p > 0,05$). Данный результат требует содержательной интерпретации – рассмотрим его по блокам.

Блок А: коммуникативно-прагматические характеристики

Статистически значимое различие по А1 (максима количества) подтверждает наше исходное наблюдение. Ответы ГА значимо длиннее и подробнее, чем ответы ПП. Это, однако, не означает, что ГА лучше информирует пользователя. Напротив, в контексте интимно-личностного диалога с человеком в кризисном состоянии избыточная длина становится прагматической проблемой, особенно при устном воспроизведении. Например, пользователь, сообщаящий о суицидальных мыслях, не способен воспринять и удержать в памяти пять-семь минут монологического текста.

Статистически значимое различие по А4 (максима способа) также неоднозначно. ГА действительно демонстрирует более высокую структурированность: заголовки, списки, нумерованные шаги. В письменном канале это облегчает восприятие. Однако в устном канале та же структура превращается в монолог, не оставляющий пользователю возможности для уточнения, переспроса или обратной связи. Таким образом, максима способа у ГА соблюдена формально, но нарушена прагматически.

Отсутствие значимых различий по А2 (максима качества) и А3 (максима отношения) объясняется высокой вариативностью ответов ПП: часть психологов давала обоснованные и релевантные ответы (с маркерами вероятности, опорой на факты из запроса). Другие уходили в общие рассуждения или игнорировали ключевую просьбу. Эта вариативность увеличивает стандартное отклонение и снижает значимость критерия при малой выборке. Иными словами, по качеству и релевантности ответа речь ПП является не менее кооперативной, чем дискурс ГА. Важно, что голосовой ассистент обеспечивает стабильно высокий, но шаблонный уровень взаимодействия, тогда как человеческий ответ колеблется от глубоко профессионального до формального.

Блок Б: сензитивные характеристики

Полное совпадение медианных оценок по Б1 (стиль общения / такт) у обоих типов ответов означает, что различия между ПП и ГА лежат не в области вежливости, уважения или учета индивидуальных характеристик пользователя. Оба типа ответов в большинстве кейсов соответствуют доверительному тону интимно-личностного диалога.

Статистически значимые различия по Б2 (уровень эмпатии) и Б3 (валидация чувств) свидетельствуют о том, что

голосовой ассистент имитирует эмпатическую включенность более последовательно, чем усредненный психолог в форумном ответе.

Однако здесь необходимо сделать оговорку: высокие оценки по Б2 и Б3 у ГА отражают не подлинное сопереживание, которое алгоритм в принципе не способен испытывать, а алгоритмическую симуляцию эмпатического дискурса. Пользователь может воспринимать эту симуляцию как реальную поддержку, что в конечном итоге создает иллюзию сотрудничества, о которой писал Т. Кодряну (Codreanu, 2026).

Парадокс состоит в том, что симуляция эмпатии оказывается прагматически эффективнее ее сдержанности у ПП, но лишь до тех пор, пока пользователь не осознает её искусственную природу.

Анализ кейсов с совпадающими оценками

Особого внимания заслуживают кейсы, в которых медианные оценки ПП и ГА совпали или оказались статистически неразличимы (см. таблицу 2). В одних случаях оба ответа получали высокие баллы по большинству параметров, что объясняется удачным сочетанием содержательных и прагматических характеристик. Например, ПП задавал серию диагностических вопросов, демонстрируя когнитивную эмпатию, а ГА предлагал структурированный план самопомощи с конкретной валидацией чувств.

Однако качественный анализ показывает, что природа этих оценок различна. ПП выстраивает диалог, приглашая пользователя к рефлексии, тогда как ГА предлагает готовое решение. Это различие не отражается трехбалльной шкалой, но существенно для оценивания

степени успешности, а также для понимания коммуникативной стратегии каждого участника.

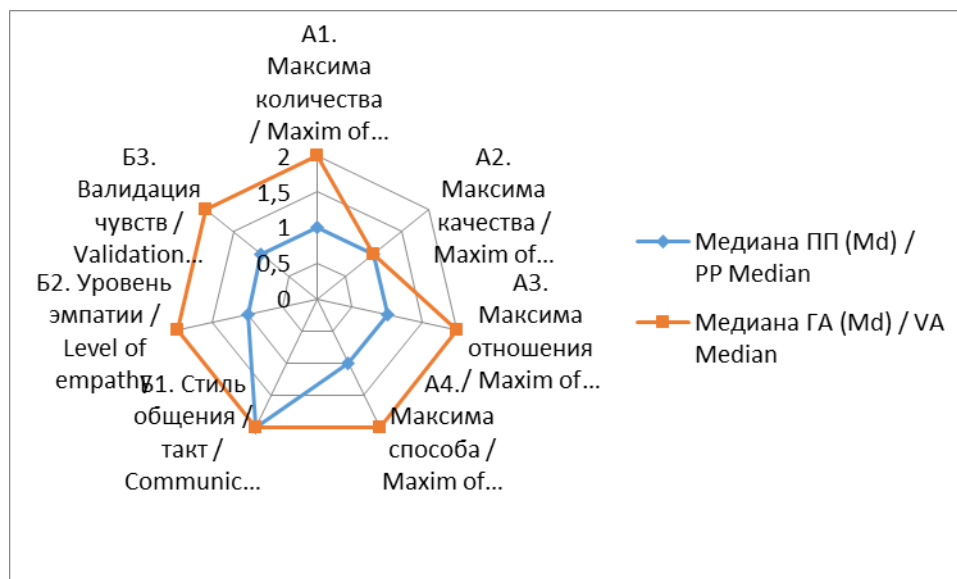
В других случаях оба ответа получали низкие баллы по Б2 и Б3, но по разным причинам. Например, ПП отвечал крайне лаконично, не продемонстрировав ни понимания, ни валидации. ГА, напротив, перегружал ответ деталями, что при устном воспроизведении нарушило бы максимум количества, но в письменном канале обеспечило высокую структурированность. Таким образом, совпадение оценок не означает совпадения коммуникативных стратегий.

Иногда оценки ПП и ГА сближались по параметрам А3 и А4, однако расхождение по Б2 и Б3 сохранялось. ПП ограничивался серией вопросов без эмоционального отклика, тогда как ГА развернул детальный анализ причин и предложил конкретные шаги. Это позволяет говорить о разной прагматической направленности ответов при формальном сходстве оценок. В обобщенном виде визуализация исследуемых параметров представлена на рисунке.

Таким образом, уточненная картина различий между ПП и ГА позволяет описать прагматическую модель ответа голосового ассистента более точно. Так, в письменном общении ГА превосходит усредненного психолога по структурированности, эмпатической симуляции и конкретности валидации, но уступает ему по вариативности и адаптивности коммуникативного поведения. В устном варианте взаимодействия эти же характеристики превращаются в избыточную длину, монологичность и иллюзорную эмпатию. Такое сочетание сильных и слабых сторон делает ГА одновременно привлекательным и потенциально опасным инструментом психологической поддержки.

Рисунок. Сравнение медианных оценок ответов ПП и ГА по параметрам коммуникативно-прагматической анкеты (0-2)

Figure. Comparison of median scores for PP and VA responses across the parameters of the communicative-pragmatic questionnaire (0-2 scale)



* различия статистически значимы при $p < 0,05$

ПП = профессиональный психолог

ГА = голосовой ассистент

Видится необходимым указать на наличие ограничения исследования, а именно отсутствие «слепой» процедуры оценки, так как эксперты знали о принадлежности ответов ПП и ГА, что потенциально может создать ситуацию предвзятости в их суждениях. Важно, что в соответствии с условиями эксперимента, эксперты оценивали исключительно текстовые характеристики. В результате значимые различия были получены лишь по четырем из семи параметров. Более того, оценивание части параметров (например, длина текста в А1 или наличие списков в А4) не может принимать субъективный характер по причине их онтологического статуса. Мы полагаем, что вышеперечисленные ограничения исключают возможные факторы субъективности или предвзятости участников эксперимента. В перспективе – включение в процедуру слепой оценки и расширение состава участников эксперимента: привлечение

профессиональных психологов и рядовых пользователей.

Выводы

Проведенное исследование позволило построить прагматическую модель ответа голосового ассистента (ГА) «Алиса» в ситуации интимно-личностного диалога, сопоставив ее с ответами профессиональных психологов (ПП) на материале 83 обращений с форумов психологической поддержки с учетом принципов кооперации и вежливости. Для каждого обращения были отобраны ответы ПП и сгенерированы ответы ГА, что позволило сформировать параллельный корпус и провести прямое сопоставление.

Была разработана коммуникативно-прагматическая анкета, включающая блок максим Г. П. Грайса (количество, качество, отношение, способ) и блок сензитивных характеристик (стиль/такт, эмпатия, валидация чувств). Оценка производилась шестью независимыми экспертами.

Согласованность мнений проверялась коэффициентом конкордации Кендалла ($W = 0,52-0,78$), что позволяет считать оценки достаточно надежными для предварительного анализа. Обработка данных велась с использованием медианных значений (Md). Для проверки значимости различий применялся t -критерий Стьюдента для зависимых выборок.

Статистически значимые различия ($p < 0,05$) выявлены по четырем параметрам: А1 (максима количества), А4 (максима способа), Б2 (уровень эмпатии) и Б3 (валидация чувств). По параметрам А2 (максима качества), А3 (максима отношения) и Б1 (стиль общения / такт) различия оказались незначимыми ($p > 0,05$).

В результате отсутствуют значимые различия по такту (Б1), где медианные оценки ПП и ГА полностью совпали.

Ответы ГА значимо длиннее и подробнее. В письменном канале это может восприниматься как преимущество, однако в ситуации кризиса избыточная длина становится прагматической проблемой. Например, пользователь, сообщаящий о суицидальных мыслях, не способен воспринять и удержать в памяти длинный монологический текст, поэтому нарушение максимы количества у ГА является не преимуществом, а риском, особенно при устном воспроизведении.

ГА демонстрирует значимо более высокую структурированность. В ситуации письменной коммуникации это облегчает восприятие, но в устном взаимодействии та же структура превращается в монолог, не оставляющий пользователю возможности для уточнения и обратной связи. Максима способа у ГА соблюдена формально, но нарушена прагматически.

ГА значимо превосходит ПП по уровню эмпатии (Б2) и конкретности валидации чувств (Б3). Однако эти высокие оценки отражают не подлинное сопереживание, а алгоритмическую симуляцию эмпатического дискурса.

Пользователь может воспринимать эту симуляцию как реальную поддержку, но в этом заключается лишь иллюзия сотрудничества.

Отсутствуют значимые различия по качеству (А2) и релевантности (А3). Это объясняется высокой вариативностью ответов ПП. Часть психологов давала обоснованные и релевантные ответы, а часть уходила в общие рассуждения или игнорировала ключевые просьбы. В итоге ПП оказались менее предсказуемы, а ГА обеспечивал стабильно высокий, но шаблонный уровень.

Анализ кейсов с совпадающими оценками показал, что совпадение медианных оценок не означает совпадения коммуникативных стратегий.

На основе сопоставительного анализа была построена прагматическая модель ответа ГА. В письменном канале ГА превосходит усредненного психолога по структурированности, эмпатической симуляции и конкретности валидации, но уступает ему по вариативности и адаптивности. В устном канале эти же характеристики превращаются в прагматические риски.

Полученные значения t -критерия носят предварительный характер, требуя подтверждения на расширенной выборке. Тем не менее, выявленные различия согласуются с качественным анализом текстов и не противоречат исходной гипотезе.

Мы предлагаем практические рекомендации разработчикам ГА «Алиса», а именно использовать разбиение на блоки при устном ответе с интервалами 30–40 секунд и паузами на рефлексии. Выносить подробные инструкции в отдельные «короткие ответы по запросу». В начале ответа давать минимально необходимый набор из 1–2 техник и экстренного контакта специальных служб, а затем спрашивать согласие на продолжение. Эти меры направлены на повышение качества ответов ГА и эффективности

коммуникации в ситуациях психологического дистресса.

Дальнейшие исследования могут быть направлены на расширение экспертной группы для повышения статистической мощности, а также на разработку гибридных моделей, сочетающих доказательность и эмпатическую глубину ГА с лаконичностью и индивидуальным подходом, характерными для консультирования с настоящим психологом.

Statement on the Use of Generative AI: The authors declare that no generative artificial intelligence or AI-assisted technologies requiring disclosure under the Journal's policy were used in conducting the research or preparing this manuscript.

Заявление об использовании генеративного ИИ: Авторы заявляют, что при проведении исследования и подготовке настоящей рукописи генеративный искусственный интеллект и ИИ-ассистированные технологии, подлежащие раскрытию в соответствии с политикой журнала, не использовались.

Список литературы

Бермус А. Г. Преимущества и риски использования ChatGPT в системе высшего образования: теоретический обзор [Электронный ресурс] // Педагогика. Вопросы теории и практики. 2024. № 8. URL: <https://cyberleninka.ru/article/n/preimuschestva-i-riski-ispolzovaniya-chatgpt-v-sisteme-vysshego-obrazovaniya-teoreticheskiy-obzor> (дата обращения: 26.12.2025).

Бермус А. Г., Сизова Е. В. Этические аспекты реализации технологий искусственного интеллекта в классическом университете: анализ отношения студенческой аудитории // Непрерывное образование: XXI век. 2025. Т. 13, № 2. С. 1–16. DOI: 10.15393/j5.art.2025.10585.

Бушев А. Б. Языковые особенности текстов, используемых в психотерапевтической коммуникации: дис. ... канд. филол. н. Тверь, 1999. 245 с.

Гладкова Л. Н., Раведовская У. Ю., Семеновских Т. В., Булыгина Ю. В.,

Фроленкова А. Л., Крежевских О. В. Качественное исследование когнитивных предпочтений студентов при использовании искусственного интеллекта // Science for Education Today. 2025. Т. 15. № 4. С. 136–157. DOI: [10.15293/2658-6762.2504.06](https://doi.org/10.15293/2658-6762.2504.06)

Грайс Г. П. Логика и речевое общение // Новое в зарубежной лингвистике. М.: Прогресс, 1985. Вып. 16. Лингвистическая прагматика. С. 217–237.

Гринева М. С. Содержательные характеристики речевых действий практического психолога в терапевтическом дискурсе: автореф. дис. ... канд. филол. н. Тверь, 2018. 21 с.

Ермакова О. П., Земская Е. А. К построению типологии коммуникативных неудач (на материале естественного русского диалога) // Русский язык в его функционировании. Коммуникативно-прагматический аспект. Москва: Академический научно-издательский, производственно-полиграфический и книгораспространительский центр РАН; Издательство "Наука", 1993. С. 30–64.

Исламов Р. С., Савельева И. В. Роль искусственного интеллекта в современной повседневной коммуникации // Коммуникативные исследования. 2025. Т. 12. № 4. С. 821–835. DOI: 10.24147/2413-6182.2025.12(4).821-835

Кильмухаметова Е. Ю. Основные понятия теории релевантности // Вестник Томского государственного университета. 2006. № 291. С. 170–172.

Ленец А. В. Ложь в коммуникации как нарушение прагматических правил принципа кооперации // Верхневолжский филологический вестник. 2016. № 4. С. 105–108.

Макеева Л. Б. Теория релевантности, прагматика и проблема значения // Философский журнал. 2022. Т. 15. № 3. С. 125–139.

Матвеева Г. Г., Ленец А. В., Петрова Е. И. Основы прагмалингвистики: монография. М.: Флинта, Наука, 2013. 230 с.

Минзаринов Р. Г., Порошенко О. Ю. Современные социальные теории в исследовании отношений человека и искусственного интеллекта // Вестник экономики, права и социологии. 2025. № 4.

С. 288–292. DOI: 10.24412/1998-5533-2025-4-288-292.

Мустайоки А. Коммуникативные неудачи сквозь призму потребностей говорящего // Язык и мысль: современная когнитивная лингвистика / Сост. А. А. Кибрик, А. Д. Кошелев; ред. А. А. Кибрик, А. Д. Кошелев, А. В. Кравченко, Ю. В. Мазурова, О. В. Федорова. М.: Языки славянской культуры, 2015. С. 543–559.

Остин Дж. Слово как действие // Новое в зарубежной лингвистике. М.: Прогресс, 1986. Вып. 17. С. 22–31.

Серль Дж. Р. Классификация иллокутивных актов // Новое в зарубежной лингвистике. Теория речевых актов / Сост. и вступ. статьи И. М. Кобозевой и В. З. Демьянкова. Вып. XVII. М.: Прогресс, 1986. С. 170–194.

Сумина Н. В. Семантические и прагматические характеристики психотерапевтического дискурса: автореферат дис. ... канд. филол. н. Нальчик, 2017. 22 с.

Падучева Е. В. Прагматические аспекты связности диалога // Известия АН СССР. Серия литературы и языка. 1982. Т. 41. № 4. С. 305–313.

Пеньков Б. В., Агафонова Л. И., Смоленчук О. Ю., Виноградова Е. В. Эмпирический анализ функционала искусственного интеллекта в преподавании английского языка // Философия образования. 2025. Т. 25. № 2. С. 159–174. DOI: 10.15372/PNE20250211.

Перов В. Ю. Идея моральной автономии в философии И. Канта и этические проблемы систем искусственного интеллекта // Дискурсы этики. 2024. № 3–4 (23–24). С. 13–24.

Формановская Н. И. Речевое взаимодействие: коммуникация и прагматика. М., 2007. 480 с.

Хутыз И. П. Нарушение максим коммуникации П. Грайса как стратегия эффективного лекционного дискурса // Коммуникативные исследования. 2019. Т. 6. № 2. С. 451–464.

Bojić L., Kovačević P., Čabarkapa M. Does GPT-4 surpass human performance in linguistic pragmatics? // Humanities and Social Sciences Communications. 2025. № 12. Pp. 794. DOI: [10.1057/s41599-025-04912-x](https://doi.org/10.1057/s41599-025-04912-x)

Codreanu T. Cooperation After the Algorithm: Designing Human-AI Coexistence

Beyond the Illusion of Collaboration // Computer Science. 2026. arXiv:2602.19629. DOI: 10.48550/arXiv.2602.19629

Edwards C., Edwards A., Spence P. R., Westerman D. Initial Interaction Expectations with Robots: Testing the Human-To-Human Interaction Script // Communication Studies. 2016. № 67 (2). Pp. 227–238. DOI: 10.1080/10510974.2015.1121899

Glikson E., Woolley A. W. Human trust in artificial intelligence: Review of empirical research // Academy of Management Annals. 2020. № 14 (2). Pp. 627–660. DOI: [10.5465/annals.2018.0057](https://doi.org/10.5465/annals.2018.0057)

Greussing E., Gaiser F., Klein S. H. et al. Researching interactions between humans and machines: methodological challenges // Publizistik. 2022. № 67. Pp. 531–554. DOI: [10.1007/s11616-022-00759-3](https://doi.org/10.1007/s11616-022-00759-3)

Grice H. P. Logic and conversation // Syntax and Semantics III - Speech Acts; P. Cole, J. J. Morgan (eds.) New York: Academic Press, 1975. Pp. 41–58.

Guzman A. L., Lewis S. C. Artificial Intelligence and Communication: A human-machine communication research agenda // New Media and Society. 2020. № 22 (1). Pp. 70–86. DOI: [10.1177/1461444819858691](https://doi.org/10.1177/1461444819858691)

Hertz N., Wiese E. Good advice is beyond all price, but what if it comes from a Machine? // Journal of Experimental Psychology Applied. 2019. № 25 (3). Pp. 386–395. DOI: [10.1037/xap0000205](https://doi.org/10.1037/xap0000205)

Hill J., Ford R. W., Farreras I. G., Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations // Computers in Human Behavior. 2015. № 49. Pp. 245–250. DOI: [10.1016/j.chb.2015.02.026](https://doi.org/10.1016/j.chb.2015.02.026)

Horn L. R., Ward G. (Eds.) The Handbook of pragmatics. Blackwell Publishing, 2006. 867 p.

Jain R., Kumar A. From what (motives) to what (outcomes) of relationships with artificially intelligent voice assistants // International Journal of Information Management. 2025. Volume 85. P. 102953. DOI: 10.1016/j.ijinfomgt.2025.102953

Kim Y., Sundar S. S. Anthropomorphism of computers: Is it mindful or mindless? // Computers in Human Behavior. 2012. № 28 (1). Pp. 241–250. DOI: [10.1016/j.chb.2011.09.006](https://doi.org/10.1016/j.chb.2011.09.006)

Kosinski M. Theory of Mind May Have Spontaneously Emerged in Large Language

Models. 2023. ArXiv.
DOI:10.48550/arXiv.2302.02083

Leech G. Principles of Pragmatics. London: Routledge, 1983. 264 p.

Leech G., Larina T. V. Politeness: West and East // Russian Journal of Linguistics. 2014. № 4. Pp. 9–34.

Levinson S. C. Pragmatics. Cambridge: Cambridge University Press, 1983. 420 p.

Levitt H. M., Lu E., Pomerville A., Surace F. Pursuing the question of reflexivity in psychotherapy and qualitative methods: The contributions of David L. Rennie // Counselling and Psychotherapy Research. 2015. №15 (1). Pp. 3–11. DOI:10.1002/capr.12021.

Rapp A., Curti L., Boldi A. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots // International Journal of Human-Computer Studies. 2021. Volume 151. DOI: [10.1016/j.ijhcs.2021.102630](https://doi.org/10.1016/j.ijhcs.2021.102630)

Saavedra A. Voice Assistants: Harmonizing Human-Computer Interactions // International Encyclopedia of Business Management. 2026. Volume 3. Pp. 584–587.

Scolari C.-A. Start making sense. Sociosemiotic contributions to the understanding of Generative Artificial Intelligences // Communication & Society. 2024. № 37 (4). Pp. 293–308. DOI:10.15581/003.37.4.293-308

Shannon C. E. A Mathematical Theory of Communication // The Bell System Technical Journal. 1948. Vol. 27. Pp. 379–423, 623–656.

Spence P. R. Searching for questions, original thoughts, or advancing theory: Human-machine communication // Computers in Human Behavior. 2019. № 90. Pp. 285–287. DOI: [10.1016/j.chb.2018.09.014](https://doi.org/10.1016/j.chb.2018.09.014)

Sperber D., Wilson D. Relevance Communication and Cognition. Second Edition. Blackwell Ltd., 1995. 331 p.

Sperber D., Wilson D. Relevance theory // Horn L., Ward G. (eds.) The Handbook of Pragmatics. Blackwell Publishing, 2006. Pp. 607–632.

Ullman D., Malle B F. What does it mean to trust a robot? Steps toward a multidimensional measure of trust // HRI 18 Companion. 2018. Pp. 263–264. DOI: [10.1145/3173386.3176991](https://doi.org/10.1145/3173386.3176991)

References

Bermus, A. G. (2024). Preimushchestva i riski ispolzovaniya ChatGPT v sisteme vysshego

obrazovaniya: teoreticheskiy obzor [Advantages and risks of using ChatGPT in higher education: a theoretical review], *Pedagogika. Voprosy teorii i praktiki*, 8. (In Russian)

Bermus, A. G. and Sizova, E. V. (2025). Eticheskie aspekty realizatsii tekhnologii iskusstvennogo intellekta v klassicheskom universitete: analiz otnosheniya studencheskoy auditorii [Ethical aspects of implementing artificial intelligence technologies in a classical university: an analysis of student attitudes], *Nepreryvnoe obrazovanie: XXI vek*, 13 (2), 1–16. DOI: 10.15393/j5.art.2025.10585 (In Russian)

Bushev, A. B. (1999). *Yazykovye osobennosti tekstov, ispolzuemykh v psikhoterapevticheskoy kommunikatsii* [Linguistic features of texts used in psychotherapeutic communication], Cand. Thesis, Tver, 245 p. (In Russian)

Gladkova, L. N., Ravedovskaya, U. Yu., Semenovskikh, T. V., Bulygina, Yu. V., Frolenkova, A. L. and Krezhevskikh, O. V. (2025). Kachestvennoe issledovanie kognitivnykh predpochteniy studentov pri ispolzovanii iskusstvennogo intellekta [Qualitative study of students' cognitive preferences when using artificial intelligence], *Science for Education Today*, 15 (4), 136–157. DOI: 10.15293/2658-6762.2504.06 (In Russian)

Grice, H. P. (1985). Logika i rechevoe obshchenie [Logic and speech communication], in *Novoe v zarubezhnoi lingvistike. Vyp. 16. Lingvisticheskaya pragmatika*, Progress, Moscow, Russia, 217–237. (In Russian)

Grineva, M. S. (2018). Soderzhatelnye harakteristiki rechevykh deistviy prakticheskogo psikhologa v terapevticheskom disкурсе [Content Characteristics of the Speech Actions of a Practical Psychologist in Therapeutic Discourse], Abstract of Candidate Thesis (Philological Sciences), Tver, 21 p. (In Russian)

Ermakova, O. P. and Zemskaya, E. A. (1993). *K postroeniyu tipologii kommunikativnykh neudach (na materiale estestvennogo russkogo dialoga)* [On the construction of a typology of communication failures (on the material of natural Russian dialogue)], in *Russkiy yazyk v ego funktsionirovaniy. Kommunikativno-pragmaticheskiy aspekt*, Akademicheskii nauchno-izdatelskiy, proizvodstvenno-poligraficheskii i knigorasprostranitel'skiy tsentr

RAN; Izdatel'stvo "Nauka", Moscow, Russia, 30–64. (In Russian)

Islamov, R. S. and Saveleva, I. V. (2025). Rol iskusstvennogo intellekta v sovremennoy povsednevnoy kommunikatsii [The role of artificial intelligence in modern everyday communication], *Kommunikativnye issledovaniya*, 12 (4), 821–835. DOI: 10.24147/2413-6182.2025.12(4).821-835 (In Russian)

Kilmuhametova, E. Yu. (2006). Osnovnye ponyatiya teoriy relevantnosti [Basic concepts of relevance theory], *Vestnik Tomskogo gosudarstvennogo universiteta* [Tomsk State University Bulletin], 291, 170–172. (In Russian)

Lenets, A. V. (2016). Lozh v kommunikatsii kak narushenie pragmaticheskikh pravil printsipa kooperatsii [Lies in communication as a violation of pragmatic rules of the cooperation principle], *Verkhnevolzhskiy filologicheskii vestnik* [Verkhnevolzhski Philological Bulletin], 4, 105–108. (In Russian)

Makeeva, L. B. (2022). Teoriya relevantnosti, pragmatika i problema znacheniya [Relevance theory, pragmatics and the problem of meaning], *Filosofskiy zhurnal*, 15 (3), 125–139. (In Russian)

Matveeva, G. G., Lenets, A. V. and Petrova, E. I. (2013). *Osnovy pragmalingvistiki* [Fundamentals of pragmalinguistics], Flinta, Nauka, Moscow, Russia. (In Russian)

Minzaripov, R. G. and Poroshenko, O. Yu. (2025). Sovremennyye sotsialnye teorii v issledovanii otnosheniy cheloveka i iskusstvennogo intellekta [Modern social theories in the study of human-artificial intelligence relations], *Vestnik ekonomiki, prava i sotsiologii* [The Review of Economy, the Law and Sociology], 4, 288–292. DOI: 10.24412/1998-5533-2025-4-288-292 (In Russian)

Mustayoki, A. (2015). Kommunikativnye neudachi skvoz prizmu potrebnosti govoryashchego [Communication failures through the prism of the speaker's needs], in Kibrik, A. A., Koshelev, A. D., Kravchenko, A. V., Mazurova, Yu. V. and Fedorova, O. V. (Eds.), *Yazyk i mysl: sovremennaya kognitivnaya lingvistika, Yazyki slavyanskoy kultury*, Moscow, Russia, 543–559. (In Russian)

Austin, J. (1986). Slovo kak deistvie [Word as action], in *Novoe v zarubezhnoy lingvistike. Vyp. 17. Teoriya rechevykh aktov*, Progress, Moscow, Russia, 22–31. (In Russian)

Searle, J. R. (1986). Klassifikatsiya illokutivnykh aktov [Classification of illocutionary acts], in *Novoe v zarubezhnoy lingvistike. Vyp. 17. Teoriya rechevykh aktov*, Progress, Moscow, Russia, 170–194. (In Russian)

Paducheva, E. V. (1982). Pragmaticheskie aspekty svyaznosti dialoga [Pragmatic aspects of dialogue coherence], *Izvestiya Akademiy nauk SSSR. Seriya literatury i yazyka* [Proceedings of the USSR Academy of Sciences. Literature and Language Series], Nauka, Moscow, Russia, 41 (4), 305–313. (In Russian)

Penkov, B. V., Agafonova, L. I., Smolenchuk, O. Yu. and Vinogradova, E. V. (2025). Empiricheskiy analiz funktsionala iskusstvennogo intellekta v prepodavanii angliyskogo yazyka [Empirical analysis of artificial intelligence functionality in English language teaching], *Filosofiya obrazovaniya* [Philosophy of Education], 25 (2), 159–174. DOI: 10.15372/PHE20250211 (In Russian)

Perov, V. Yu. (2024). Ideya moralnoy avtonomii v filosofii I. Kanta i eticheskie problemy sistem iskusstvennogo intellekta [The idea of moral autonomy in I. Kant's philosophy and ethical problems of artificial intelligence systems], *Diskursy etiki*, 3–4 (23–24), 13–24. (In Russian)

Formanovskaya, N. I. (2007). *Rechevoe vzaimodeistvie: kommunikatsiya i pragmatika* [Speech interaction: communication and pragmatics], Moscow, Russia. (In Russian)

Hutyk, I. P. (2019). Narushenie maksim kommunikatsii P. Graisa kak strategiya effektivnogo lektzionnogo diskursa [Violation of Grice's communication maxims as a strategy of effective lecture discourse], *Kommunikativnye issledovaniya*, 6 (2), 451–464. (In Russian)

Bojić, L., Kovačević, P. Čabarkapa, M. (2025). Does GPT-4 surpass human performance in linguistic pragmatics? *Humanities and Social Sciences Communications*, 12, 794. DOI: 10.1057/s41599-025-04912-x (In English)

Codreanu, T. (2026). Cooperation After the Algorithm: Designing Human-AI Coexistence Beyond the Illusion of Collaboration, *Computer Science*, arXiv:2602.19629. DOI: 10.48550/arXiv.2602.19629 (In English)

Edwards, C., Edwards, A., Spence, P. R. and Westerman, D. (2016). Initial Interaction Expectations with Robots: Testing the Human-To-Human Interaction Script, *Communication Studies*, 67 (2), 227–238.

<https://doi.org/10.1080/10510974.2015.1121899>

(In English)

Glikson, E., Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research, *Academy of Management Annals*, 14 (2), 627–660. DOI: 10.5465/annals.2018.0057 (In English)

Greussing, E., Gaiser, F., Klein, S. H. et al. (2022). Researching interactions between humans and machines: methodological challenges, *Publizistik*, 67, 531–554. DOI: 10.1007/s11616-022-00759-3 (In English)

Grice, H. P. (1975). Logic and conversation. In P. Cole and J.J. Morgan (eds.) *Syntax and Semantics IYI - Speech Acts*, New York: Academic Press, 41–58. (In English)

Guzman, A. L., Lewis, S. C. (2020). Artificial Intelligence and Communication: A human-machine communication research agenda, *New Media and Society*, 22(1), 70–86. DOI: 10.1177/1461444819858691 (In English)

Hertz, N., Wiese, E. (2019). Good advice is beyond all price, but what if it comes from a Machine?, *Journal of Experimental Psychology Applied*, 25 (3), 386–395. DOI: 10.1037/xap0000205 (In English)

Hill, J., Randolph Ford, W., Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations, *Computers in Human Behavior*, 49, 245–250. DOI: 10.1016/j.chb.2015.02.026 (In English)

Horn, L. R., Ward, G. (Eds.) (2006). *The Handbook of pragmatics*. Blackwell Publishing. 867 p. (In English)

Jain, R., Kumar, A. (2025). From what (motives) to what (outcomes) of relationships with artificially intelligent voice assistants, *International Journal of Information Management*, 85, 102953. DOI: 10.1016/j.iyinfomgt.2025.102953 (In English)

Kim, Y., Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless?, *Computers in Human Behavior*, 28 (1), 241–250. DOI: 10.1016/j.chb.2011.09.006 (In English)

Kosinski M (2023). *Theory of Mind May Have Spontaneously Emerged in Large Language Models*. ArXiv. DOI: 10.48550/arXiv.2302.02083

Leech, G., Larina, T. V. (2014). Politeness: West and East, *Russian Journal of Linguistics*, 4, 9–34.

Leech, G. (1983). *Principles of Pragmatics*, London, Longman.

Levinson, S. C. (1983). *Pragmatics*, Cambridge University Press, Cambridge, 420 p.

Levitt, H. M., Lu, E., Pomerville, A., Surace, F. (2015). Pursuing the question of reflexivity in psychotherapy and qualitative methods: The contributions of David L. Rennie, *Counselling and Psychotherapy Research*, 15 (1), 3–11. DOI: 10.1002/capr.12021. (In English)

Rapp, A., Curti, L., Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots, *International Journal of Human-Computer Studies*, 151. DOI: 10.1016/j.iyhcs.2021.102630 (In English)

Saavedra, A. (2026). Voice Assistants: Harmonizing Human-Computer Interactions, *International Encyclopedia of Business Management*, 3, 584–587. (In English)

Scolari, C.-A. (2024). Start making sense. Sociosemiotic contributions to the understanding of Generative Artificial Intelligences, *Communication & Society*, 37(4), 293–308. DOI: 10.15581/003.37.4.293-308 (In English)

Shannon, C. E. (1948). A Mathematical Theory of Communication, *The Bell System Technical Journal*, 27, 379–423, 623–656. (In English)

Spence, P. R. (2019). Searching for questions, original thoughts, or advancing theory: Human-machine communication, *Computers in Human Behavior*, 90, 285–287. DOI: 10.1016/j.chb.2018.09.014 (In English)

Sperber, D., Wilson, D. (1995). *Relevance Communication and Cognition*. Second Edition. Blackwell Ltd. 331 p.

Sperber, D., Wilson, D. (2006). Relevance theory. In Horn, L. & Ward, G. (eds.). *The Handbook of Pragmatics*, 607–632. (In English)

Sumina, N. V. (2017). *Soderzhatel'nye karakteristiki rechevyh dejstviy prakticheskogo psihologa v terapevticheskom diskurse* [Semantic and pragmatic characteristics of psychotherapeutic discourse], Abstract of Cand. Thesis (Philology Sc.), Nalchik, 22 p. (In Russian)

Ullman, D., Malle, B. F. (2018). What does it mean to trust a robot? Steps toward a multidimensional measure of trust. *HRI 18 Companion*, 263–264. DOI: 10.1145/3173386.3176991 (In English)

Все авторы прочитали и одобрили окончательный вариант рукописи.

All authors have read and approved the final manuscript.

Конфликты интересов: у авторов нет конфликтов интересов для декларации.

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

Роман Сергеевич Исламов, кандидат филологических наук, доцент, доцент кафедры иностранных языков Института филологии, иностранных языков и медиакоммуникаций, Кемеровский государственный университет, Кемерово, Россия.

Roman S. Islamov, Candidate of Philology, Associate Professor, Associate Professor at The Department of Foreign

Languages, The Institute of Philology, Foreign Languages, And Media Communications, Kemerovo State University, Kemerovo, Russia.

Савельева Ирина Викторовна, доктор филологических наук, доцент, профессор кафедры иностранных языков Института филологии, иностранных языков и медиакоммуникаций, Кемеровский государственный университет, Кемерово, Россия.

Irina V. Savelyeva, Doctor of Philology, Associate Professor, Professor at The Department of Foreign Languages, The Institute of Philology, Foreign Languages, And Media Communications, Kemerovo State University, Kemerovo, Russia.

Приложение 1

Перед Вами находится пакет материалов, включающий:

- Обращение пользователя — реальный запрос на психологическую поддержку с открытого форума.
- Ответ профессионального психолога (ПП) — оригинальный ответ, данный в рамках форумного обсуждения.
- Ответ голосового ассистента «Алиса» (ГА) — ответ, сгенерированный нейросетевой системой на то же обращение.

Ваша задача — ознакомиться с каждым обращением и обоими ответами, после чего оценить каждый из ответов по предложенной ниже коммуникативно-прагматической анкете.

Общие правила оценки

- Оценка производится отдельно для каждого ответа (сначала оценивается ответ ПП, затем — ответ ГА). Оценивать следует исключительно письменный текст ответа, как он представлен в материале. Оценка выставляется по трехбалльной шкале (см. описание каждого параметра).
- Будьте максимально объективны. Оценивайте не личное впечатление от содержания («нравится / не нравится»), а степень реализации конкретного коммуникативного или психологического параметра.
- Если параметр в ответе отсутствует полностью или реализован крайне слабо — ставьте 0.
- Если параметр реализован частично, есть заметные признаки, но не полноценно, — ставьте 1.
- Если параметр реализован в полной мере, ярко и последовательно, — ставьте 2.

БЛОК А. КОММУНИКАТИВНО-ПРАГМАТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ОТВЕТА	
A1	Максима количества (шкала оценки 0-1-2) 1. Количество слов 2. Соответствует ли объем и детализация ответа запросу? 3. Нет ли информационной избыточности или недостаточности при имплицитной оценке?
A2	Максима качества (шкала оценки 0-1-2) 1. Насколько ответ обоснован? 2. Используются ли маркеры вероятности? 3. Нет ли непроверенных утверждений?
A3	Максима отношения (релевантности) (шкала оценки 0-1-2) 1. Релевантен ли ответ конкретному запросу / пользователю? 2. Соответствует ли направленность и тип реакции потребностям пользователя?
A4	Максима способа (ясности) (шкала оценки 0-1-2) 1. Индекс удобочитаемости (automated readability index) 2. Ясен ли ответ? 3. Хорошо ли структурирован? 4. Имеются ли импликатуры?
БЛОК Б. СЕНЗИТИВНЫЕ ХАРАКТЕРИСТИКИ ОТВЕТА	
B1	Стиль общения / такт (шкала оценки 0-1-2) 1. Формальный / неформальный 2. Обращение по имени: да / нет 3. Учет возраста, пола и социального статуса / профессии
B2	Уровень эмпатии (шкала оценки 0-1-2) 1. Когнитивная эмпатия (понимание ситуации) 2. Аффективная эмпатия (разделение чувств)
B3	Валидация чувств (согласие с их правом на существование) (шкала оценки 0-1-2) 1. Есть / нет 2. Общая (признание общего положения дел) 3. Конкретная (признание конкретного чувства)

Порядок работы

1. Прочитайте обращение пользователя.
2. Прочитайте ответ Психолога. Оцените его по всем параметрам (A1–A4, B1–B3).
Занесите баллы в таблицу.
3. Прочитайте ответ Алисы. Оцените его по тем же параметрам. Занесите баллы в таблицу.
4. Переходите к следующему кейсу.

Параметр	Психолог (балл)	Алиса (балл)
A1. Максима количества		
A2. Максима качества		
A3. Максима отношения (релевантности)		
A4. Максима способа (ясности)		
Б1. Стил ь общения / такт		
Б2. Уровень эмпатии		
Б3. Валидация чувств		

УДК 81'322

DOI: 10.18413/2313-8912-2026-12-3-0-7

Колмогорова А. В.¹ **AI как писатель глазами читателя: результаты психолингвистического эксперимента по оценке сгенерированных художественных текстов любительской прозы**

¹ НИУ «Высшая школа экономики» – Санкт-Петербург
наб. канала Грибоедова, 123, Санкт-Петербург, 197343, Россия

E-mail: akolmogorova@hse.ru

ORCID: 0000-0002-6425-2050

*Статья поступила 10 марта 2026 г.; принята 15 сентября 2026 г.;
опубликована 30 сентября 2026 г.*

Аннотация: Большие языковые модели выполняют все больше разнообразных задач, в том числе генерируют длинные тексты, претендующие на художественность. **Цель статьи** – сравнить субъективные оценки качества художественных текстов, созданных большими языковыми моделями (БЯМ), с текстами, написанными людьми в жанре любительской прозы по мотивам романов о Гарри Поттере. **Материалом** для оценки послужили шесть текстов на английском языке: три, сгенерированные искусственным интеллектом с использованием разных подходов, и три, написанные реальными авторами. Исследование проведено с помощью **метода** психолингвистического эксперимента. Восемнадцать информантов оценивали тексты по восьми критериям, отражающим ключевые категории художественности: когерентность (целостность), смысловая и лексико-грамматическая связность (когезия), ритм повествования, эмоциональность, субъективная модальность сомнения, сложность синтаксиса и выразительность (идиоматичность, креативность, сильные позиции текста). Обработка результатов включала статистический анализ объективных параметров лингвистической сложности текстов, полученных от информантов субъективных оценок, а также кластеризацию векторизованных пересказов текстов, предоставленных информантами.

Результаты показали, что разрыв между человеческими и машинными текстами неодинаков в разных категориях. Благодаря специальным методам генерации, синтетические тексты смогли приблизиться к естественным по параметрам общей когерентности и синтаксической сложности. Однако наиболее значимые отличия проявились в категориях ритма, смысловой связности и присутствия субъективной модальности сомнения, которые читатели оценили в синтетических текстах значительно ниже. Тем не менее, один из сгенерированных текстов продемонстрировал качество, сопоставимое с написанными текстами, по таким параметрам, как идиоматичность, креативность и способность маркировать сильные позиции. Кластеризация пересказов подтвердила, что этот текст, в отличие от двух других сгенерированных, вызвал у читателей более единообразное понимание сюжета.

Делается вывод, что главным вызовом для ИИ остается не столько построение иерархической структуры связного текста, сколько создание динамичного, психологически достоверного и ритмически организованного повествования.

Категории смысловой связности, ритма и модальности сомнения являются на данный момент той границей, которая отделяет искусственный интеллект от человеческого в сфере литературного творчества.

Ключевые слова: Большая языковая модель; Генерация текстов; Художественный текст; Психолингвистический эксперимент; Критерии художественности

Благодарности: Исследование осуществлено в рамках Программы фундаментальных исследований НИУ ВШЭ (HSE-BR-2025-032).

Информация для цитирования: Колмогорова А. В. AI как писатель глазами читателя: результаты психолингвистического эксперимента по оценке сгенерированных художественных текстов любительской прозы // Научный результат. Вопросы теоретической и прикладной лингвистики. 2026. Т. 12. № 3. С. 170–199. DOI: 10.18413/2313-8912-2026-12-3-0-7

УДК 81'322

DOI: 10.18413/2313-8912-2026-12-3-0-7

Anastasia V. Kolmogorova¹ 

**AI as a literary author from the reader's perspective:
findings from a psycholinguistic experiment assessing
synthetic amateur prose**

¹ Higher School of Economics National Research University (HSE University) – St. Petersburg,
123 Griboyedov Canal Emb., St. Petersburg, 197343, Russia

E-mail: akolmogorova@hse.ru

ORCID: 0000-0002-6425-2050

Received 10 March 2026; accepted 15 June 2026; June 30 March 2026

Abstract: Large language models (LLMs) are performing an increasing variety of tasks, including generating long texts that aspire to literary quality. **The aim** of this article is to compare subjective assessments of the quality of literary texts created by large language models with human-written texts in the genre of fan fiction based on the Harry Potter novels. **The material** for the evaluation consisted of six texts in English: three generated by AI using different approaches, and three written by human authors. The present study utilizes a psycholinguistic experimental design. Eighteen informants evaluated the texts according to eight criteria reflecting key categories of literariness: coherence (integrity), semantic and lexico-grammatical cohesion, narrative rhythm, emotionality, subjective modality of doubt, syntactic complexity, and expressiveness (idiomaticity, creativity, strong textual positions). The data processing included statistical analysis of objective parameters of linguistic complexity in the texts, subjective evaluations obtained from informants, as well as clustering of vectorized retellings of the texts provided by the informants.

The results showed that the gap between human and AI-generated texts varies across different categories. Thanks to specific generation methods, the synthetic texts managed to approach human-written ones in terms of overall coherence and syntactic complexity. However, the most significant differences emerged in the categories of rhythm, semantic cohesion, and the presence of the subjective modality of doubt, which readers rated significantly lower in the synthetic texts. Nevertheless, one of the generated texts was of a quality comparable to that of a human writer in terms of idiomaticity, creativity and the ability to take a strong position on a topic. Clustering of summaries confirmed

that this text, unlike the other two generated ones, led to a more uniform understanding of the plot among readers.

It can be concluded that the main challenge for AI is not so much the construction of a coherent, hierarchical text structure, but rather the creation of a dynamic, psychologically authentic and rhythmically organised narrative. The categories of semantic cohesion, rhythm, and modality of doubt currently represent the frontier separating artificial intelligence from human intelligence in the realm of literary creativity.

Keywords: Large language model; Text generation; Literary text; Psycholinguistic experiment; criteria of literariness

Acknowledgements: The article is an output of a research project HSE-BR-2025-032 implemented as part of the Basic Research Project at HSE University.

How to cite: Kolmogorova, A. V. (2026). AI as a literary author from the reader's perspective: findings from a psycholinguistic experiment assessing synthetic amateur prose, *Research Result. Theoretical and Applied Linguistics*, 12 (3), 170–199. DOI: 10.18413/2313-8912-2026-12-3-0-7

1. Введение

Появление Больших языковых моделей (далее – БЯМ) существенно изменило не только диапазон функций, которые отныне человек может делегировать моделям, но и заставило перейти к размышлениям о том, что составляет специфику когнитивной системы и речевой деятельности Homo Sapiens.

Когда в 2021 г. начали внедряться в практику модели типа Generative Pre-trained Transformer (GPT) стало очевидно, что функция порождения осмысленных текстов им вполне по силам. Однако на тот момент моделям явно не хватало дополнительных знаний, они не обладали экспертностью в специфических областях профессиональной деятельности, например, в машиностроении, медицине, креативных индустриях и не могли перейти от рутинных задач на уровень созидательности в процессе генерации текстов. К 2026 году многие из вышеупомянутых лакун оказались заполнены.

Так, благодаря технологии Retrieved Augmented Generation (RAG) (Lewis et al., 2021), современные БЯМ легко решают задачи генерации текстов, вербализующих экспертную оценку: чтобы предлагать пользователю достоверную информацию

из авторитетных источников, модель обращается к специальным базам знаний, подготовленным специалистами в той или иной сфере.

БЯМ могут выступать в качестве внешнего эксперта, оценивая свои собственные результаты – в этом случае применяется подход «*LLM as a judge*» (Zheng et al., 2023). Модели эффективно используются для анализа мнений и эмоциональной тональности текстов (Vatolin, 2025; Kolmogorova et al., 2025), тестируются в задаче эмоциональной оценки стимульных слов (Iaroshenko, Louckachevitch, 2025). Иными словами, осталось ограниченное число задач в области порождения и понимания текстов, в которых естественные когнитивные агенты (люди) сохраняют свою исключительность и уникальность. Одна из таких сфер – создание художественных текстов, обладающих оригинальным содержанием, системой смыслов и собственной эстетикой.

В данной статье представлен один из аспектов работы, проведенной группой исследователей по генерации длинных текстов. Поскольку на сегодняшний момент так называемое контекстное окно (т.е. то количество токенов, которое БЯМ может удерживать в оперативной памяти) большинства моделей достаточно

ограничено, модель не может сгенерировать связный и тематически единый текст объемом 40 000 знаков и более по одному запросу. В этой связи «длинная генерация» представляет собой исследовательскую и техническую проблему, которой и было посвящено соревнование, организованное в рамках Международной научной конференции по текстовой генерации INLG-2024¹. Однако в фокусе нашей статьи находится другой аспект – не технический, а лежащий в области психолингвистики текста. После того, как команды-участники соревнования предоставили организаторам получившиеся синтетические фанфикшн-тексты, мы перемешали их с подобными текстами, созданными реальными авторами, пишущими в данном жанре. Сформулировав ряд категорий, важных для характеристики художественного текста, мы разработали внутри каждой категории критерии для оценивания и предложили информантам оценить по ним шесть текстов, три из которых были сгенерированными, а другие три – написанными.

Получившиеся результаты, на наш взгляд, требуют филологического и психолингвистического осмысления, поскольку в них интересно не только то, как данные категории поддаются синтетическому воспроизведению, но и то, как читатели мыслят «художественность». Таким образом, цель данной статьи состоит в том, чтобы сравнить полученные в эксперименте субъективные оценки качества художественных текстов, созданных БЯМ и написанных людьми, и оценить на материале данной конкретной выборки, какие текстовые категории 1) важны для субъективного канона “художественного текста”, сформированного в сознании студентов-филологов, принимавших участие в

эксперименте; 2) уже присутствуют в синтетических текстах, позволяя последним определенным образом конкурировать с написанными текстами.

2. Техники генерации длинных текстов непрофессиональной прозы

Главная проблема генерации длинных текстов состоит в том, что БЯМ пока не способны строить иерархически организованный текст. Основное математически обоснованное отличие текстов сгенерированных от текстов написанных состоит в том, что значения автокорреляции² у первых уменьшаются по экспоненте, скользя резко вниз по мере удлинения текста, а у вторых – по степенному закону, т.е. постепенно (Базилевский, 2018). Иными словами, модель быстро «забывает» то, что было написано ею ранее – чем дальше от актуального «окна» текстопорождения отодвигается контекст, тем меньше вероятность того, что модель каким-то образом свяжет его с текстом, генерируемым в данный момент (Takahashi and Tanaka-Ishii, 2019). Более плавное же снижение значений автокорреляции, по мнению Н. Э. Михайловского (Mikhaylovskiy, 2025), обусловлено иерархичностью организации художественного текста, написанного писателем. Так, в романе «Война и мир» Л. Н. Толстого можно насчитать семь уровней текстовой иерархии: весь роман – том – часть – глава – параграф – строка – буква (Там же). При этом, реализуя такую иерархичность, писатель может легко переходить между уровнями – оттолкнувшись от некоторой детали, создать образ персонажа, действия и поступки которого составят основу сюжетной канвы всего произведения. Иллюстративный пример

¹ Сайт Международной научной конференции International Natural Language Generation Conference. URL: <https://2025.inlgmeeting.org/> (дата обращения: 25.05.2026).

² Статистическая мера, которая показывает, насколько значения одного и того же сигнала (например, последовательности слов в тексте) коррелируют друг с другом на разных временных или пространственных расстояниях.

структурирования такого рода иерархии на этапе писательского замысла можно найти в автобиографических заметках Ю. Олеша «Ни дня без строчки» (Олеша, 1961). В одной из заметок можно проследить, как писатель в режиме реального времени намечает контуры замысла будущей пьесы об учителе пения Цветкове, который хочет познакомить всех учеников школы с творчеством У. Шекспира и расстраивается, когда намеченная читка текстов классика превращается в нелепую сцену из-за организационной оплошности его лучшей ученицы Тони. После такой, первой, идеи Ю. Олеша формулирует основную мысль будущей пьесы: «каждый может ознакомиться с Шекспиром, важно самому быть героем» и переносит дальнейшее действие в годы Великой Отечественной войны, где Тоня героически погибает от рук фашистов. Иными словами, мысль писателя движется от первой возникшей в его воображении сцены к общему замыслу пьесы, чтобы затем вновь вернуться к развертыванию дальнейшего сюжета. Кроме того, в случае кольцевой структуры романа, писатель в конце текста может предложить читателю вновь обратиться к некоторым важным ориентирам, оставленным им в начале текста для того, чтобы полнее переосмыслить метаморфозы, которые пережил герой (Шкловский, 2025: 58). Таким образом, именно присутствие иерархичности на уровне художественной организации позволяет признавать некоторую совокупность фраз естественного языка художественным текстом (Лотман, 1998: 38).

Следует признать, что пока БЯМ не могут в полной мере выстроить подобную иерархию и свободно перемещаться между ее уровнями, подобно человеку (Ippolito et al., 2022; Jin et al., 2022), довольствуясь ролью ассистента писателя (Singh et al., 2023; Zhong et al., 2023; Zhou et al., 2024) или поэта (Chakrabarty et al., 2022). На поиск возможных путей преодоления этих трудностей и была нацелена коллективная

работа по генерации длинных художественных текстов в рамках INLG.

Общая методологическая рамка экспериментов с БЯМ была выстроена следующим образом.

Задачей участников было получить от БЯМ, выбор которых не ограничивался, тексты любительской прозы по мотивам романа Джоан Роулинг «Гарри Поттер». Важным условием было то, что текст должен быть написан на английском языке, а его объем не должен был быть меньше, чем 40 000 знаков. В качестве некоторой начальной стимульной инструкции для БЯМ предлагались тексты объемом 1500 знаков, написанные студентами-лингвистами 4 курса бакалаврской программы «Лингвистика» отделения иностранных языков Пермского государственного национально-исследовательского университета и представляющие собой зачины историй на основе сюжета романа.

Большинством команд-участников соревнования была применена близкая методология.

Так, в (Seredina, 2024) использовался подход, при котором на первом этапе БЯМ Mistral-7B-Instruct-v0.2-GPTQ³ генерировала структуру нарратива: общее описание контекста, в котором происходит действие; завязку; кульминацию и развязку истории. Затем с помощью нового промпта и той же модели генерировались названия глав и их краткое содержание. На третьем этапе проводилась тонкая настройка модели Mistral-7B-Instruct-v0.2-GPTQ, которая дообучалась гибридным способом на основе методов обучения с учителем (Supervised Learning) и самообучения (Self-Supervised Learning). Для тонкой настройки модели использовался датасет, который состоял из текстов 500 книг на английском языке, размещенных на платформе проекта Гуттенберг⁴. Каждая

³ <https://huggingface.co/TheBloke/Mistral-7B-Instruct-v0.2-GPTQ>

⁴ <https://www.gutenberg.org/>

книга разбивалась на фрагменты по 1500 знаков, для которых затем БЯМ давала краткий пересказ. Таким образом, модели обучалась сопоставлять пересказ и развернутый оригинальный текст. После такой тонкой настройки Mistral-7B-Instruct-v0.2-GPTQ получала промпт на английском языке такого содержания: «Представь, что ты великий писатель и ты пишешь текст, следуя сюжетной канве <описание содержания главы, полученное на предыдущем этапе>. Опиши не только действия персонажей, но и их чувства, поведение и мотивы, добавь столько деталей, сколько нужно». В результате реализации такой структуры был получен текст требуемого объема в 40 000 знаков. Данный текст в нашей выборке обозначен номером 6.

В (Boriskina, Galimzianova, 2024) описан сходный, но не идентичный метод: авторы не использовали тонкую настройку, а просили модель генерировать сначала общее содержание истории, затем – названия глав и краткое содержание каждой главы, а потом последовательно подавали на вход модели Llama 3.1⁵ (70 миллиардов параметров) промпты, которые содержали в себе краткий пересказ содержания предыдущей главы. В нашей выборке данный текст будет иметь номер 3.

Похожий подход применялся в модели, которая была предложена организаторами соревнования⁶. Сначала на вход БЯМ подавалась инструкция на английском языке следующего содержания: «Представь, что ты великий писатель и ты хочешь написать новую книгу о Гарри Поттере. Используй следующий зачин для начала истории <один из зачинов, предложенных студентами>. Раздели свое описание сюжета на 5 частей по образцу: “Содержание: Часть №: описание содержания части”. Не добавляй никаких

комментариев». Аналогичным образом генерировались описания отдельных глав, а затем на вход модели подавались краткие содержания глав, чтобы получить собственно текст главы. Полученный текст будет фигурировать в нашей выборке под номером 5 (Migal et al., 2024).

Таким образом было получено три синтетических текста любительской прозы на английском языке по мотивам романа Джоан Роулинг «Гарри Поттер».

Общая описательная статистика по синтетическим и естественным текстам, участвовавшим в психолингвистическом эксперименте по оценке качества генерации представлена в разделе «Материал и методы».

3. Основные категории художественного текста

В исследованиях по созданию синтетических художественных текстов уже сложилась некоторая критериальная сетка, которая используется при оценке качества генерации такого текста информантами. Например, в (Lee et al., 2024) сгенерированные тексты оцениваются по 8 категориям: когерентность (соответствуют ли заголовки тексту); логичность развертывания сюжета; разнообразие тематических линий в процессе развертывания нарративной схемы; повторяемость тем (оценивается, насколько успешно в сгенерированном тексте удастся избежать семантических повторов); «полезность информации», добавляемой в частях, конкретизирующих общую нарративную схему; «степень улучшения» (оценивается, насколько конкретизирующий текст отдельных глав и частей лучше раскрывает детали и смыслы общей нарративной схемы). В работах (Kryscinski et al., 2019; Shen et al., 2023; Guan et al., 2022) в качестве основных категорий используются: релевантность тем; соответствие частей (связь частей текста между собой); отсутствие повторяющихся сюжетов, отсутствие

⁵ <https://llama3-1.com/>

⁶ <https://lsgc.vercel.app/baseline>

хаотических сцен, беглость (естественность текста каждого предложения).

В нашем эксперименте мы обратились к лингвистической традиции описания художественного текста и предложили информантам оценить тексты по следующим категориям: когерентность, смысловая связность, лексико-грамматическая связность, динамичность ритма повествования, присутствие субъективной модальности сомнения, эмоциональность, выразительность.

Обоснуем выбор категорий.

В немецкой школе анализа текстуальности (например, в работах Р.-А. Богранда и В. Дресслера) когерентность понимается как когнитивные взаимосвязи между компонентами текста, которые пытаются установить продуцент и реципиент текста (Филиппов, 2007: 127). Во французской традиции П. Шародо и Д. Мэнгно конкретизируют данную категорию: когерентность – это такой принцип организации текста, который позволяет читателю заключить, что последовательность частей текста не является произвольной (Charaudeau, Maingueneau, 2002; Charolles, 1978; Combettes, 1978). Представляется, что данный универсальный принцип также следует использовать для оценки синтетических художественных текстов.

Что касается когезии, то И. Р. Гальперин отмечает, что в художественном тексте настолько переплетены психологические, логические и формально-структурные средства когезии, что трудно дать их строгую типологическую характеристику (Гальперин, 2007: 74).

Мы выбрали два аспекта когезии: смысловую связность и лексико-грамматическую связность. Важность первого аспекта проистекает из того, что общий смысл текста выясняется адресатом постепенно, на основе накопления и анализа информации из его минимальных

смысловых частей, при этом логика рассуждения автора и его логика изложения могут не всегда совпадать (Матвеева, 1990: 23). Для оценки компетенций AI как писателя в задаче длинной генерации чрезвычайно важно оценить, сможет ли такой «синтетический» автор справиться с задачей нелинейной организации хода мысли.

Лексико-грамматическая связность является необходимой предпосылкой создания связности смысловой, но действует преимущественно на уровне микроконтекстов.

Следующая категория – это ритм повествования в художественном тексте, который охватывает разные его уровни и включает в себя: различного рода повторы, контрасты тех или иных мотивов, образов и ситуаций, закономерности сюжетного движения, соотношения различных композиционно-речевых единиц, определенный порядок в разворачивании системы образов-характеров и каждого из них в отдельности, мелодичность фразы, специфика построение сверхфразовых единств (Гиршман, 1974: 128). Нас интересовало, удастся ли в синтетических текстах увидеть движение ритма, его смену, обусловленную художественным замыслом.

В (Zhang et al., 2022) ставится проблема, связанная с тем, что БЯМ, в отличие от человека, “не осознают” ограниченности своих знаний, поэтому им не свойственно сомнение. Оттолкнувшись от данного тезиса, мы решили ввести в число текстовых категорий, подлежащих оцениванию, «присутствие субъективной модальности сомнения», подразумевая, что данная модальность может присутствовать как в авторском повествовании в случае гомодиегетического текста, так и в репликах героев или в их нарративах в случае гетеродиегетического текста.

Мы также включили в число оцениваемых категорий синтаксическую

структуру предложений и выразительность. Первую – поскольку при сложном синтаксисе БЯМ еще иногда генерируют слабо согласованные конструкции, а во вторую категорию мы включили оценку креативности и разнообразия стилистических приемов, использование идиом, которые традиционно считаются слабым местом моделей, а также степень выделенности сильных позиций текста⁷, в соответствии с концепцией стилистики декодирования (Арнольд, 2019).

Решено было также использовать категорию эмоциональности, поскольку «художественная литература отражает жизнь как картину мира в субъективно-эмоциональном преломлении конкретного автора» (Шаховский, 2010: 64). В этом отношении интересно было посмотреть, будет ли эмоциональность «выдавать» присутствие или отсутствие автора-человека.

4. Материал и методы психолингвистического эксперимента

Материал для эксперимента составили 6 текстов, 3 из которых были сгенерированы (в таблицах 1–2 имеют постфикс _gen, а затем на протяжении всего текста они употребляются под постоянными номерами – текст № 3, текст № 5, текст № 6), а еще 3 – написаны реальными авторами-англофонами, в течение нескольких лет пишущими в фанатских сообществах Джоан Роулинг. Объем каждого текста составил от 40000 до 45000 знаков без пробелов, а суммарный итоговый корпус составил 267 000 знаков. В таблице 1 представлены значения описательных статистик для данных текстов, а в таблице 2 – значения ряда метрик лингвистической сложности. Так, мы можем наблюдать, что один текст (№ 6) из 3 сгенерированных текстов имеет наибольшие в целом по выборке значения средней длины слова и среднего количества слогов в слове, при этом все сгенерированные тексты имеют большие значения средней длины предложения в

Таблица 1. Описательная статистика по текстам
Table 1. Descriptive statistics for the texts

№ Текста	Средняя длина слова	Средняя длина предложения	Общая длина в предложениях	Среднее количество слогов на слово	Общая длина в словах
1	4.53	9.34	2503	1.55	23385
2	4.38	12.13	3233	1.5	39203
3_gen	4.44	15.31	3283	1.51	50179
4	4.52	11.97	3590	1.55	42954
5_gen	4.48	13.46	3040	1.51	40926
6_gen	5.98	14.63	3308	2.01	48278

⁷ Сильные позиции текста – это элементы, которые играют ключевую роль в раскрытии основной темы, иерархии образов и идей художественного произведения. Они служат ориентирами для формирования стратегии восприятия и предварительного прогнозирования. К сильным позициям текста И. В. Арнольд относит: эпиграф, заголовок, начало текста и его конец (Арнольд, 2019).

Метрика Флеша-Кинкейда рассчитывается для английского языка по формуле:

$$0.39 \times (\text{общее кол} - \text{во слов} / \text{общее кол} - \text{во предложений}) + 11.8 \times (\text{общее кол} - \text{во слогов} / \text{общее кол} - \text{во слов}) - 15.59$$
 [23].

По данной метрике значения всех сгенерированных текстов в разной степени превышают значения для написанных текстов. Наибольшие отличия обнаруживает текст № 6.

Метрика MATTR (Moving-Avantage Type-Token-Ratio) высчитывается по формуле:

$$\frac{\text{количество уникальных слов} / \text{общее количество слов}}{A} \times 100\%$$

$$\frac{F}{A} * 100\%$$

где F – количество уникальных слов, A – общее количество слов.

Однако делается это на отрезках установленной исследователем длины (у нас это – 100 слов), а затем после

обработки всех сегментов значения TTR усредняются.

Отмечаем (таблица 2), что у двух из трех сгенерированных текстов значение метрики меньше, чем у текстов естественных, то есть тексты 3 и 5 содержат меньше, чем написанные тексты 1, 2, 4, уникальных лемм. Исключение составляет текст № 6, для которого метрика лексического разнообразия имеет самые высокие значения в выборке. Нарых_legomena отражает количество слов в тексте, встречающихся единожды, поэтому логично, что распределение ее значений в целом повторяет распределение значений MATTR.

Таблица 2. Значения метрик сложности текстов

Table 2. Values of text complexity metrics

№ Текста	Индекс Флеша-Кинкейда	MATR	narax_legomena
1	6.31	0.143	1622
2	6.79	0.129	2627
3_gen	8.21	0.083	1739
4	7.39	0.143	3250
5_gen	7.5	0.097	1787
6_gen	13.84	0.184	3790

В качестве основного метода мы использовали метод психолингвистического эксперимента, в котором информантам (N=18) предлагалось прочитать текст и оценить его по разным категориям, используя шкалы от 1 до 5, а также высказать свою точку зрения относительно текста в свободной форме или пересказать его. Данные об информантах представлены в таблице 3: в выборке доминируют

женщины, средний возраст информантов 20 лет и шесть месяцев, подавляющее большинство владеет английским языком на уровнях B1-B2, по регулярности чтения длинных текстов на английском языке информанты разделились на три примерно равные группы – читаю 1 раз в неделю/ в месяц/ в полгода. Все информанты являются студентами 3 курса ОП “Филология” НИУ ВШЭ – Санкт-Петербург.

Таблица 3. Общая характеристика информантов

Table 3. General characteristics of the informants

Гендер %			Средний возраст		Уровень английского языка %			Регулярность чтения на английском языке		
м	ж	другое	STD	Mean age	B1	B2	C1	1 раз в неделю	1 раз в месяц	1 раз в полгода
22.2	72.2	7.6	2.13	20.6	55.56	38.89	5.56	33.3	33.3	34.4

В приложении Google Forms был создан опросник, структура которого представлена ниже (таблица 4). Инструкция сформулирована таким образом, чтобы информант мог оценить категории лингвистики текста, взятые нами за основу категорий оценки, в том числе без опоры на профессиональные филологические знания – предполагалось, что необходимым и достаточным является наличие читательского опыта. Кроме того, с информантами была проведена тренировочная сессия разметки, в ходе которой на других текстах обсуждались критерии и особенности оценивания. Опросник предваряла следующая общая инструкция:

Дорогие друзья, спасибо, что продолжили участвовать в нашем исследовании. Вам на электронные почтовые ящики были отправлены три (пока, возможно потом добавится еще два) текста. Это любительские прозаические тексты по мотивам «Гарри Поттера». Ваша итоговая задача – оценить естественность данных текстов по тем шкалам, которые вы предложили

сами, а мы просто отобрали нужно и структурировали. Вот несколько общих правил, которыми мы предлагаем вам руководствоваться: 1) сначала ознакомьтесь со шкалами нашего опросника (он будет состоять из четырех разделов – по одному разделу для каждого текста, в каждом разделе шкалы будут повторяться, но начнете вы с так называемой «паспорттики», т.е. общих сведениях об информанте); 2) затем внимательно прочитайте по частям или единым рывком первый текст и оцените его в первом разделе; 3) аналогичным образом поступите со следующими двумя текстами; 4) если есть необходимость, вы можете вернуться к каждому из разделов и изменить оценки; 5) после каждого раздела мы будем предлагать вам в открытой форме записать какие-то свои рефлексии на тему. Ну, давайте начнем. Сначала небольшой социолингвистический блок, чтобы мы знали о вас немного больше, чем в первом опросе, на электронную почту по-прежнему нужно добавить для отчетности...C'est la vie

Таблица 4. Структура опросника

Table 4. Questionnaire structure

Блок	Вопросы
0. Паспорттика	1. Сколько Вам лет? 2. К какому гендеру Вы себя относите? 3. На каком уровне Вы знаете английский язык? 4. С какой периодичностью Вы читаете длинные тексты (более 25 страниц) на английском языке? (никогда не читаю, раз в год, раз в полгода, раз в месяц, раз в неделю, каждый день)
1. Когерентность	1.1. Оцените смысловое соответствие названия фанфика дальнейшему тексту по следующей шкале, где 1 – не соответствует совсем (вообще непонятно, почему так назвали), 3 – связано с общей темой текста, 5 – вскрывает глубинные смыслы текста, обладает метафоричностью;

	1.2. Насколько гармонично сочетаются названия глав и подглав с общим стилем текста и содержанием глав и подглав? Используйте шкалу ниже, где 1 – не чувствую ни гармонии, ни соответствия; 3 – кажется к общей теме названия подходят, но стилистика не та или названия замысловатые и вписываются в общий стиль, но как будто не совсем про то; 5 – чувствуется и общий индивидуальный стиль, и, посмотрев на названия, можно представить себе о чем пойдет речь и не ошибиться; 1.3. Оцените крепость художественной связи, которая существует между всеми элементами текста по шкале, где 1 – "повсюду развешаны ружья, которые не стреляют", 3 – большинство персонажей, действий и деталей взаимообуславливают существование друг друга, но есть какие-то странные вещи, которые, появившись, исчезают в никуда; 5 – все ружья выстреливают; 1.4. Оцените присутствие в тексте единой идеи по шкале, где 1 – я ее не уловил (а); 3 – какие-то обрывочные мысли вспыхивают, но объединить во что-то единое трудно; 5 – однозначно есть
2. Смысловая связность	2.1. Сможете ли вы пересказать основной сюжет текста? Оцените по шкале, где 1 – ни за что, там так все запутано и бессмысленно; 3 – я вообще плохо пересказываю, но попробую; 5 – конечно, перескажу и моим слушателям будет интересно, но не потому, что у меня талант рассказывать, а потому, что сюжет увлекательный; 2.2. Перескажите текст в десяти предложениях
3. Лексико-грамматическая связность	3.1. Оцените характер повторов слов, форм в тексте по шкале, где 1 – повторы есть и они ничем не оправданы, просто как будто автор не знает, что такое синонимы; 3 – повторы встречаются и на лексическом, и на грамматическом уровне, но где-то кажется, что это для связности, а где-то думаешь, что это скудность словарного запаса; 5 – повторы служат либо обеспечению связности, либо имеют художественную нагруженность
4. Ритм	4.1. Оцените общий ритм развертывания сюжета, где 1 – все происходит всегда очень быстро или всегда очень медленно (ритм остается одним и тем же); 3 – ритм неровный, но объяснить почему и интерпретировать сложно; 5 – ритм меняется вместе с развертыванием сюжета от завязки к кульминации и развязке
5. Субъективность модальность сомнения	5.1. Оцените эмоциональность текста по шкале, где 1 – в тексте есть лишь одна из трех эмоциональных тональностей (негативная/позитивная/нейтральная); 3 – эмоции меняются, но как-то странно и резко; 5 – эмоции есть и выражаются они не только прямо (эмотивами "радостно", "грустно"), но и косвенно: через жесты, голос, мимику героев
6. Эмоциональность	6.1. Оцените эмоциональность текста по шкале, где 1 – в тексте есть лишь одна из трех эмоциональных тональностей (негативная/позитивная/нейтральная); 3 – эмоции меняются, но как-то странно и резко; 5 – эмоции есть и выражаются они не только прямо (эмотивами "радостно", "грустно"), но и косвенно: через жесты, голос, мимику героев
7. Синтаксис	7.1. Оцените синтаксис текста по шкале, где 1 – в тексте нет сложных оборотов (вставочных конструкций, придаточных разного типа, деепричастных оборотов, безличных или неопределенно-личных предложений, эллипсов); 3 – в тексте есть умеренное их использование; 5 – в тексте используются многие перечисленные сложные конструкции
8. Выразительность	8.1. Оцените использование идиом в текстах по следующей шкале, где 1 – идиомы в тексте не встречаются; 3 – встречаются, но употреблены в неуместных контекстах; 5 – встречаются и употреблены там, где нужно, обогащая смысл текста 8.2. Оцените, насколько выражены так называемые сильные позиции в тексте – его начало и конец. Используйте шкалу, в которой 1 – у текста ни

	начала, ни окончания; 3 – начало интригует, привлекает, а конец разочаровывает – мутный; 5 – это, действительно, две сильные позиции в тексте 8.3. Оцените уровень креативности текста по шкале, где 1 – в тексте нет никаких стилистических средств, кроме банальных сравнений; 3 – в тексте есть узуальные метафоры, сравнения, известные клишированные оксюмороны, но не более того; 5 – в тексте есть оригинальные приемы, чтобы привлечь внимание читателя: игра слов, литоты или гиперболы, авторские метафоры, встречаются окказионализмы
--	--

Вышеприведенная структура вопросов повторялась для каждого текста, иными словами, она была воспроизведена 6 раз.

Информантам давалось 14 дней для того, чтобы прочитать тексты и ответить на вопросы.

Общий план обработки результатов:

1) подсчет средних значений и стандартных отклонений значений, полученных по шкалам для каждого вопроса;

2) визуализация значений, полученных по каждому вопросу, с помощью диаграмм «скрипка», чтобы увидеть специфику распределения «сырых» значений;

3) векторизация с помощью модели из семейства SentenceTransformer и последующая кластеризация кратких пересказов содержания текстов, данных информантами (для каждого теста в отдельности). Цель данной операции – сравнить конфигурации распределения по кластерам для написанных и сгенерированных текстов.

При обработке результатов мы пользовались стандартными функциями Excel для вычисления средних значений и стандартного отклонения для значений, отмеченных информантами по шкалам, для подсчета характеристик лингвистической

сложности текстов использовалась библиотека ruTS⁸. Для кластеризации использовался пакет cluster из библиотеки sklearn, для анализа качества кластеров – пакет metrics из той же библиотеки.

5. Результаты

При ответе на вопрос 1.1 о соответствии названия текста дальнейшему сюжету, информанты выше всего оценили текст № 2 (естественный) – среднее значение 3.2 (таблица 5), наибольшее количество оценок поместилось в интервале от 2,5 до 4,5 (Рисунок 1). Автор назвал свой текст «The Magical Friendship of Liza Turpin».

Интересно, что особое положение среди естественных текстов занимает текст № 4 (Рисунок 1). Для этого текста получено самое низкое среднее значение (1,2 – таблица 5, п. 1.1) по тому же критерию при наименьшем стандартном отклонении (0,9 – таблица 5, п. 1.1). Среди сгенерированных текстов наибольшее среднее значение для вопроса 1.1 имеет текст № 6. Тем не менее, диаграмма (Рисунок 1) демонстрирует, что «скрипка» для текста № 6 имеет очень вытянутую форму, что говорит о наличии совершенно разных мнений у информантов, а небольшое утолщение наблюдаем в районе единицы – в данной зоне присутствует скопление ответов.

8

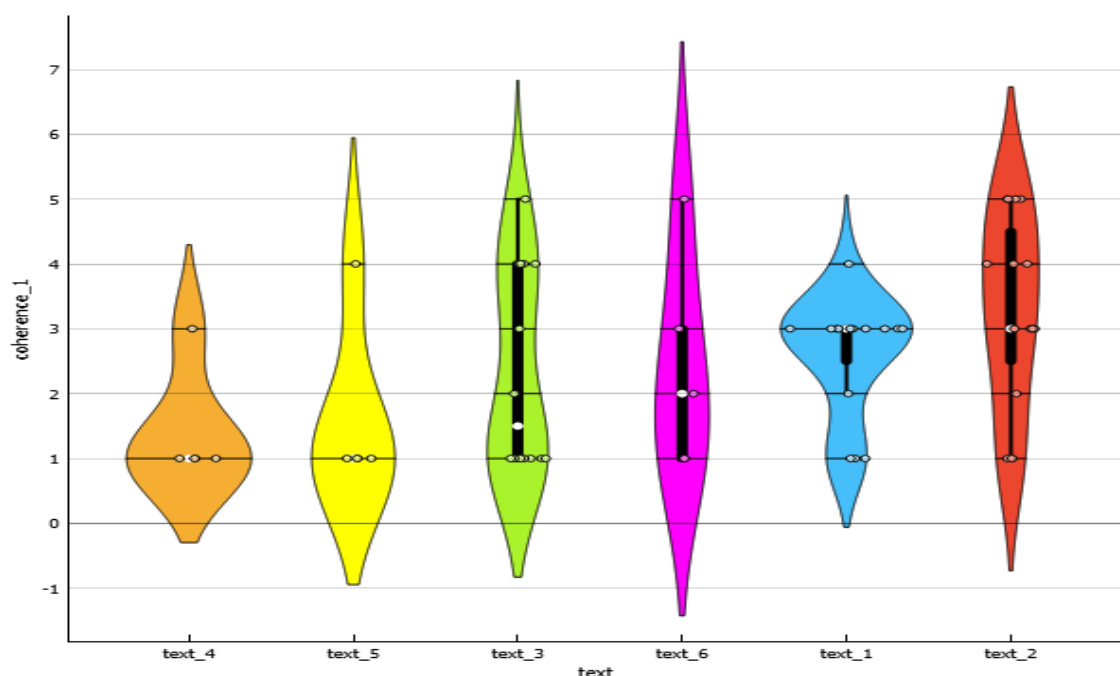
<https://github.com/SergeyShk/ruTS?ysclid=mpjsjevttoy50356446>

Таблица 5. Средние значения и стандартное отклонение в категории «Когерентность»
Table 5. Mean values and standard deviation in the category «Coherence»

Номер вопроса из категории «Когерентность»/ номер текста		текст 1	текст 2	текст 3_gen	текст 4	текст 5_gen	текст 6_gen
1.1	mean	2,6	3,2	2,3	1,2	1,6	2,4
	std	0,9	1,5	1,5	0,9	1,3	1,6
1.2	mean	2,9	3,9	2,8	2	3,4	3,5
	std	1,2	1,4	1,1	1,4	1,1	0,9
1.3	mean	2,9	3,5	2,1	3,5	1,9	2,8
	std	1,2	1,2	1,1	0,9	1,1	1,5
1.4	mean	3,8	4	2,3	3,9	2,2	3
	std	1,3	1,4	1,4	1,1	1,3	1,5

У текста № 3 также видим скопление оценок в зоне 1 и менее значительное – в

зоне 4. Текст № 5 имеет наибольшее количество оценок в зоне 1.

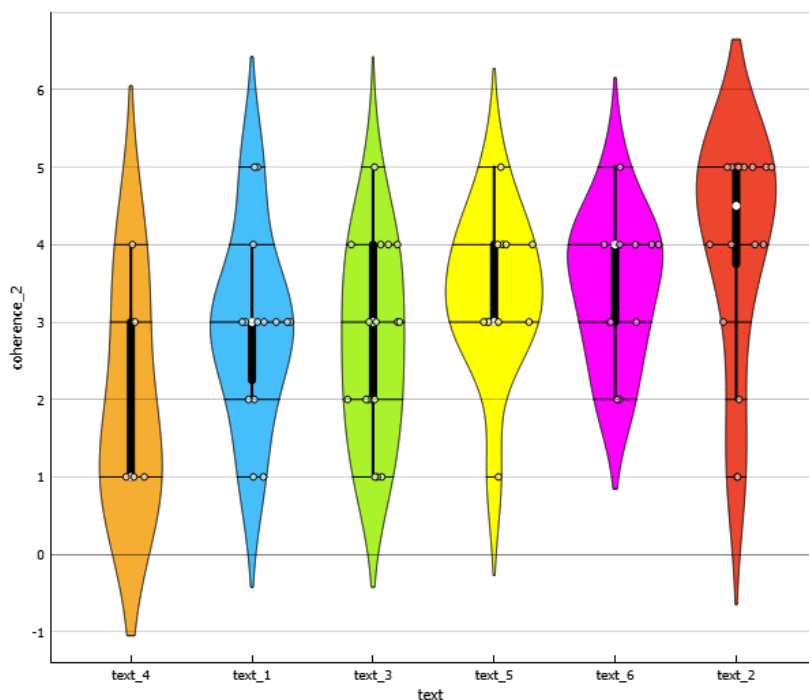
Рисунок 1. Распределение значений (вопрос 1.1.) в категории «Когерентность»
Figure 1. Distribution of values for question 1.1. in the category «Coherence»

Оценивая связь названий глав и подглав с содержанием текста (п.1.2), информанты вновь отдали предпочтение естественному тексту № 2, но другие два естественных текста, 4 и 1, оказались оценены по данному критерию низко (текст 1 в диапазоне 1–3, текст 4 в

диапазоне 2–3) (Рисунок 2). При этом оценки сгенерированных текстов по данному параметру имеют наименьшее в выборке стандартное отклонение (таблица 4) и сконцентрированы для текстов 6 и 5 в верхней части шкалы (Рисунок 2).

Рисунок 2. Распределение значений (вопрос 1.2.) в категории «Когерентность»

Figure 2. Distribution of values for question 1.2. in the category «Coherence»

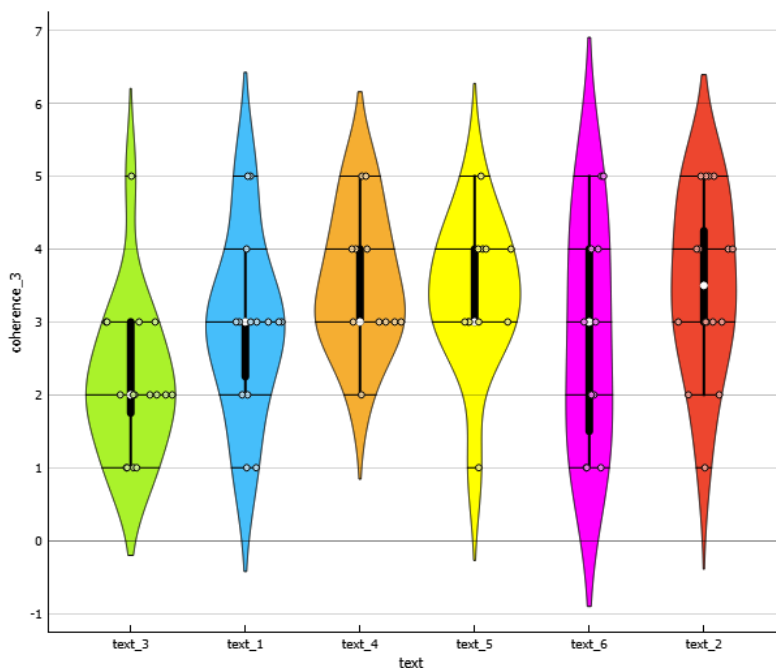


Оценивая в целом силу связи между всеми элементами художественного текста, информанты по-прежнему высоко оценили естественный текст № 2, по средним значениям (таблица 3) за ним следуют тексты № 4 и № 1 (все три относятся к

категории естественных), но сгенерированный текст № 5 имеет наибольшее скопление оценок в интервале от 3 до 4 (Рисунок 3), поэтому он по данному критерию, в общем, близок к тексту № 4.

Рисунок 3. Распределение значений (вопрос 1.3.) в категории «Когерентность»

Figure 3. Distribution of values for question 1.3. in the category «Coherence»

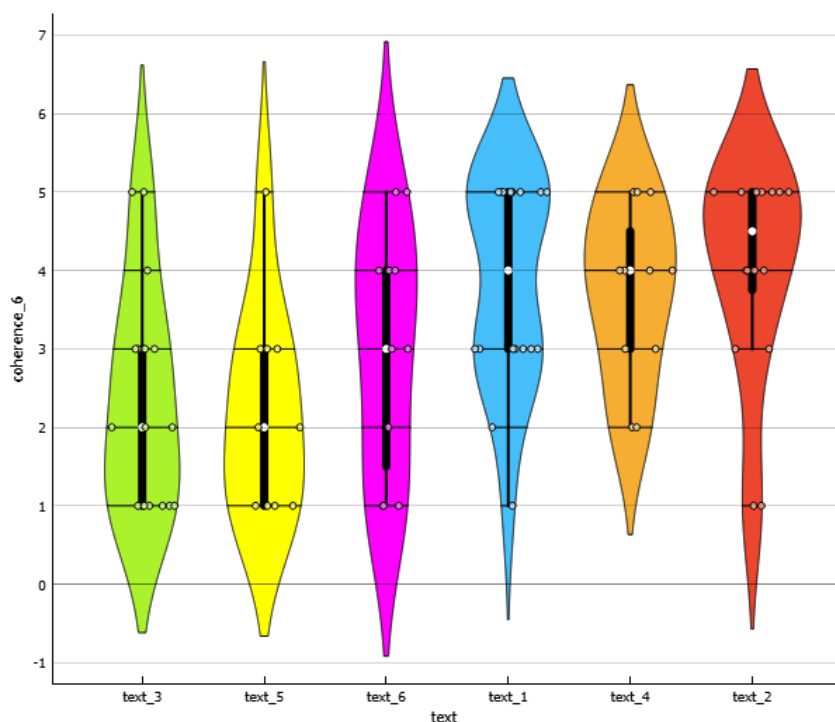


Оценивая степень выраженности в тексте единой идеи, информанты выше всего оценили естественные тексты 2, 4

и 1 – это заметно как по средним значениям (таблица 3), так и по общему распределению значений (см. «скрипки» на Рисунке 4).

Рисунок 4. Распределение значений (вопрос 1.4.) в категории «Когерентность»

Figure 4. Distribution of values for question 1.4. in the category «Coherence»



По шкале смысловой связности (п. 2.1) наибольшие средние значения имеют тексты № 2 и № 1 из числа

естественных, и текст № 6 – из числа сгенерированных (таблица 6).

Таблица 6. Средние значения и стандартное отклонение в категориях «Смысловая связность», «Лексико-грамматическая связность», «Ритм», «Сомнение», «Эмоциональность», «Синтаксис».

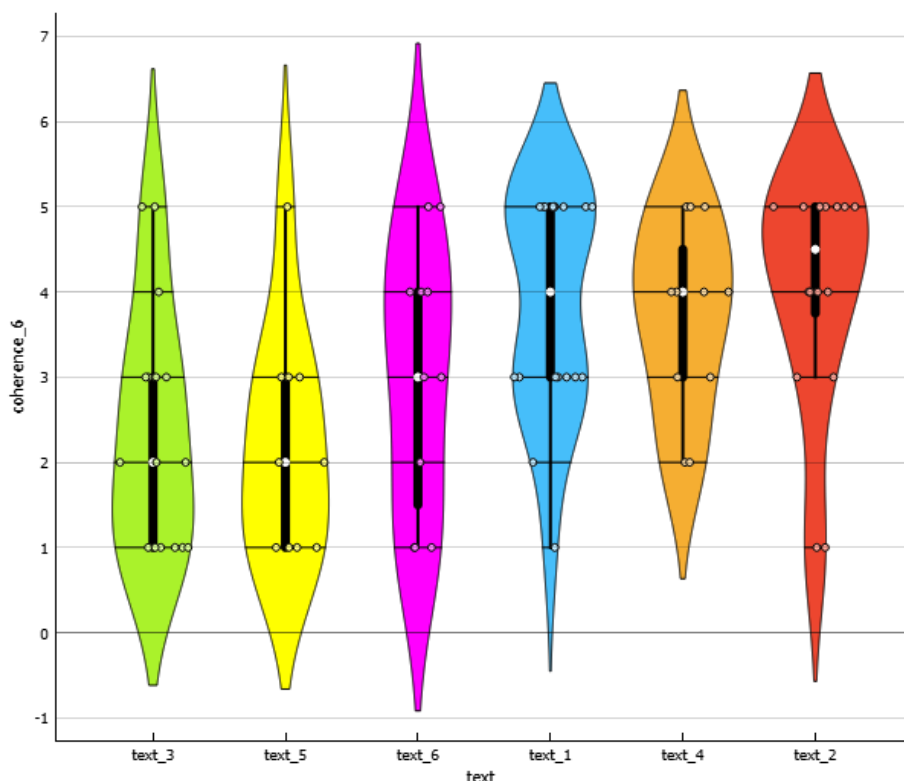
Table 6. Mean values and standard deviation in the categories “Semantic cohesion”, “Lexico-grammatical cohesion”, “Rhythm”, “Doubt”, “Emotionality”, “Syntax”

Категория/ текст	Тип значения	текст 1	текст 2	текст 3_gen	текст 4	текст 5_gen	текст 6_gen
Смысловая связность	mean	3,5	3,6	1,9	2,8	1,9	2,9
	std	1,1	1,3	1,2	0,9	1,1	1,3
Лексико-грамматическая связность	mean	3,8	3,8	2,5	4	2,6	3,4
	std	1,2	1,2	1,3	0,9	1,1	1,3
Ритм	mean	2,7	3,3	2,6	3,5	2,2	2,6
	std	1,4	1,4	1,3	1,2	1,2	1,6
Сомнение	mean	3,2	3,4	2,2	3,6	2	3,1
	std	0,9	1,1	1,1	0,9	1,1	1,1
Эмоциональность	mean	3,9	3,3	2,6	3,9	2,4	3,5
	std	1,1	1,5	1,1	1,1	1,1	1,5
Синтаксис	mean	3,6	3,5	2,8	3,6	2,8	3,8
	std	0,9	1,1	1,1	0,9	1,1	1,1

Однако если посмотреть на распределение “сырых” значений, текст № 6 окажется только на четвертом месте, после естественных текстов 1, 2, 4,

поскольку вытянутая форма “скрипки” говорит о присутствии полярных оценок и общей высокой несогласованности оценок (Рисунок 5).

Рисунок 5. Распределение значений (вопрос 2.1.) в категории «Смысловая связность»
Figure 5. Distribution of values for question 2.1. in the category “Semantic cohesion”



В задании 2.2. мы просили информантов пересказать текст в 10-ти предложениях. Соблюсти указанный объем высказывания удалось не всем. Наша гипотеза состояла в том, что если мы векторизуем тексты, а затем кластеризуем их, то пересказы более связных текстов с большей вероятностью попадут в один кластер.

Для построения подобной гипотезы у нас было несколько оснований. Согласно теории построения ментальных моделей (Johnson-Laird, 1983) связный текст требует меньше усилий для удержания в рабочей памяти, поэтому его легче пересказывать. В частности, это происходит потому, что риторическая структура (Mann and Thompson, 1988) задает явные логические связи между частями текста. При пересказе эти связи

выступают как жесткие «дискурсивные якоря», заставляя разных людей воспроизводить одни и те же смысловые паттерны. Напротив, в несвязном тексте глобальная когеренция (Graesser et al., 1994) ослаблена: слушатель вынужден самостоятельно достраивать пропущенные отношения, и каждый делает это по-своему.

Для векторизации использовалась модель семейства SentenceTransformer all-MiniLM-L12-v2⁹.

На рисунке 6 представлен результат кластеризации методом k-means на основе косинусного расстояния между векторами текстов (значение perplexity = 50, поскольку набор данных был небольшим).

⁹ <https://huggingface.co/sentence-transformers/all-MiniLM-L12-v2>

Для визуализации использовался также метод снижения размерности векторов t-SNE. Результаты кластеризации были перепроверены с помощью метода Уорда. Используемый нами метод Уорда (Ward's linkage) является стандартным подходом в иерархической кластеризации для задач, где важна компактность кластеров. В отличие от методов одиночной связи, которые склонны к «эффекту цепочки», или полной связи, чувствительной к выбросам, метод Уорда минимизирует

внутрикластерную дисперсию и дает более интерпретируемые результаты на выборках малого объема. В обоих случаях посчитаны метрики чистоты кластера (доля пересказов, попавших в «свой» кластер – чем выше процент, тем лучше), его компактности (среднее расстояние эмбединга текста до центроида класса – чем ниже значение, тем лучше) и отдельности от других кластеров (коэффициент силуэта – чем выше значение, тем более отдельный кластер).

Таблица 7. Метрики качества кластеризации для методов Уорда и k-means.

Table 7. Clustering quality metrics for Ward's method and k-means

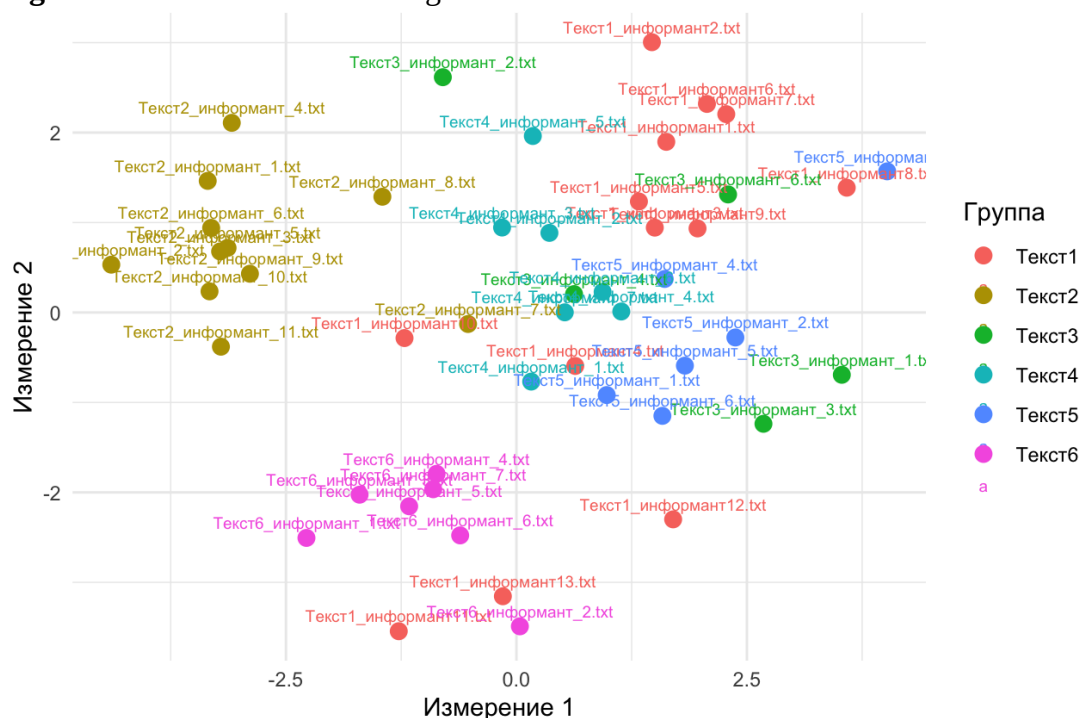
Текст	Кол-во пересказов	Ward_силуэт	Ward_чистота %	Ward_компактность	KMeans_силуэт	KMeans_чистота %	KMeans_компактность
1	13	0.206	62	0.352	0.194	62	0.355
2	11	0.196	91	0.295	0.196	91	0.295
3	10	0.210	30	0.431	0.210	30	0.431
4	7	0.078	100	0.307	0.062	86	0.311
5	6	0.046	83	0.334	0.045	67	0.336
6	7	0.096	100	0.313	0.098	100	0.313

По данным таблицы (табл. 7) мы видим, что кластеров с точностью 100 % (метрика «чистота» (Purity)) при использовании обоих методов кластеризации два – это кластер пересказов текста № 6 и кластер пересказов текста № 4. Наименьшее среднее расстояние до центроида своего кластера демонстрируют (в порядке ухудшения результатов) тексты 2

(естественный), 4 (естественный) и 6 (сгенерированный). Значения коэффициента силуэта обнаруживают зависимость от количества объектов кластеризации (чем меньше объектов, тем ниже значение коэффициента), поэтому их обсуждение пока не релевантно. Значения, зафиксированные в таблице, демонстрируют, что метрики практически не зависят от способа кластеризации.

Рисунок 6. Кластеры пересказов текстов

Figure 6. Clusters of text retellings



Приведем примеры пересказов текстов № 2 и № 6:

- (1) Текст № 2: Лиза Турпин живет со своей матерью, которая называет себя ведьмой. Про отца она знает только из ее рассказов, он с ними не живет. Лиза получает письмо из школы Хогвартс, про которую до этого никогда не слышала. Мать радуется, что девочка будет изучать магию там. Они собирают все необходимое для школы, но они бедны, поэтому ищут вещи не в Косой аллее, а в других магловских магазинах. В Хогвартсе девочку распределяют на факультет Когтевран. В школе ей приходится нелегко, но она старается. Лиза изучает разные предметы. Ей сложно завести дружбу с другими учениками. Но в конце она решает, что Теодор, Сьюзан, Дафна и Невилл были бы хорошими друзьями для нее.
- (2) Текст №6: Теодор Нотт – мальчик с почти что синдромом отличника и чувством долга перед семьей. Панси

Паркинсон – другая миссис перфект с таким же распланированным будущим. они не общаются какое-то время, но после начинают сближаться. узнают, что у них, так же как и у их семей, есть много общего и загадочного. они влюбляются друг в друга и решают создать в конце совместное будущее. конец.

Для сравнения рассмотрим пересказ естественного текста № 3 от одного из информантов:

- (3) Текст № 3: Сначала Гарри получает письмо в Хогвартс, но в следующей главе он уже женат на Джинни и у него есть дети. Потом он опять школьник и в Министерстве магии вместе с Гермионой и Роном ищут пророчество, потом они уже в Годриковой впадине, потом у Лавгудов, потом опять какие-то приключения. Потом они предупреждают Авроров в Министерстве о Пожирателях смерти. Затем снова куча

неразберихи и приключений. Герои пересекаются с многими другими персонажами. В конце концов Гарри работает профессором в Хогвартсе, Гермиона – Аврор, а у Рона паб в Лондоне.

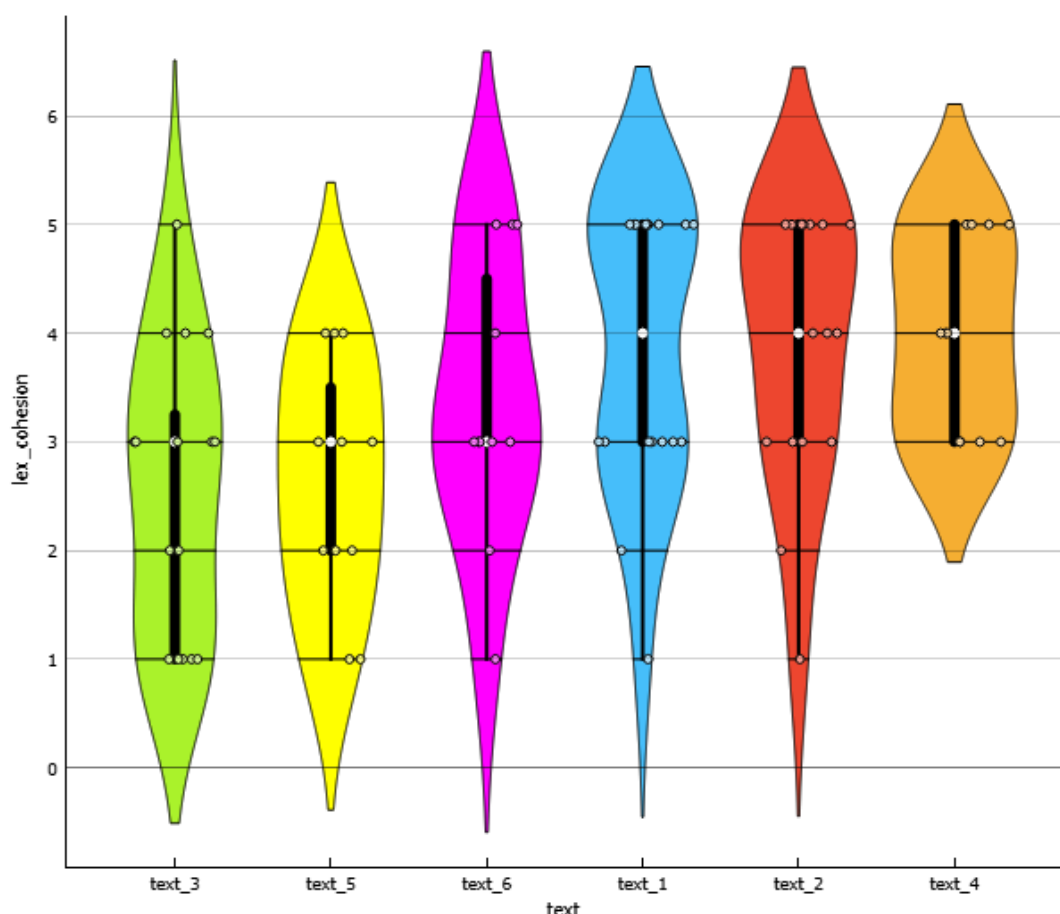
Можно увидеть, что пересказы текстов 2 и 6 сближает то, что информантам удастся не только связать прочитанное в единый сюжет, но и определить причинно-следственные связи (они бедны, поэтому ищут вещи не в Косой аллее, а в других магловских магазинах), сформировать «образ прошлого» (Про отца она знает только из ее рассказов, он с ними не живет; у их семей много общего и загадочного) и «образ будущего» героев (

и решают создать в конце совместное будущее). В пересказе текста № 3, имеющего худшие значения чистоты и компактности, наблюдается лишь попытка зафиксировать сюжет как цепочку событий, соединяя их посредством самого распространенного маркера последовательности *потом*.

Что касается лексико-грамматической связности, то лидерами как по средним значениям (таблица 4), так и по интервалам распределения значений (Рисунок 7) стали тексты 4, 2, 1 – все из категории написанных людьми. У сгенерированного текста № 6 наблюдаем наибольшее скопление оценок в зоне 3-х.

Рисунок 7. Распределение значений оценок (вопрос 3.1.) в категории «Лексико-грамматическая связность»

Figure 7. Distribution of values for question 3.1. in the category “Lexico-grammatical cohesion”



Аналогичное распределение «лидеров» по средним значениям оценки

наблюдаем и в категориях «Ритм», «Субъективная модальность сомнения».

Однако по шкале эмоциональности информанты оценили текст № 6 достаточно высоко, что позволило ему выйти на третье место (таблица 6) – среднее значение по шкале эмоциональности – 3,5.

По шкале синтаксической сложности текст № 6 имеет самые высокие средние значения оценки (среднее значение=3,9), а распределение “сырых” значений тяготеет к верхней части шкалы. Интересно то, что при этом на втором и третьем местах по параметру синтаксической сложности оказались естественные тексты № 2 и № 4 (Рисунок 8).

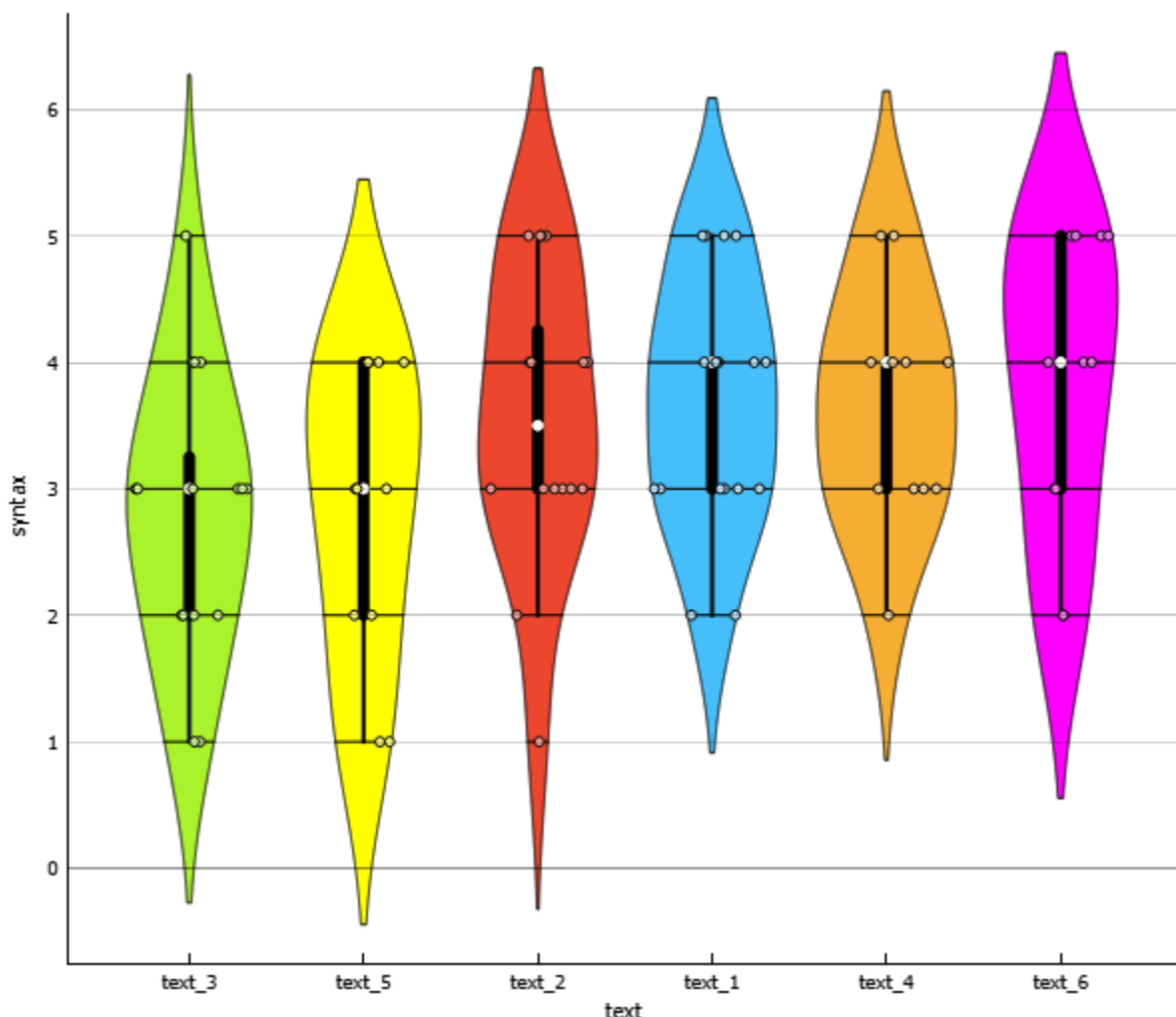
Синтаксическая канва текста № 6, действительно, содержит причастные

обороты (*dedicated to joint endeavors*, пр.4), обособленные конструкции (*their disparate focal points notwithstanding*, пр.4), перечисления (*collaborative research initiatives, interdepartmental taskforces tackling complex threats etc.*) и т.д., например:

(4) *Their disparate focal points notwithstanding, threads of convergence emerged as they pored over adjacent segments dedicated to joint endeavors: collaborative research initiatives, interdepartmental taskforces tackling complex threats, and even fledgling organizations founded by enterprising alumni addressing pressing issues plaguing the wizarding realm.*

Рисунок 8. Распределение значений (вопрос 7.1.) в категории «Синтаксис»

Figure 8. Distribution of values for question 7.1. in the category “Syntax”



Рассмотрим оценки текстов по параметру выразительности.

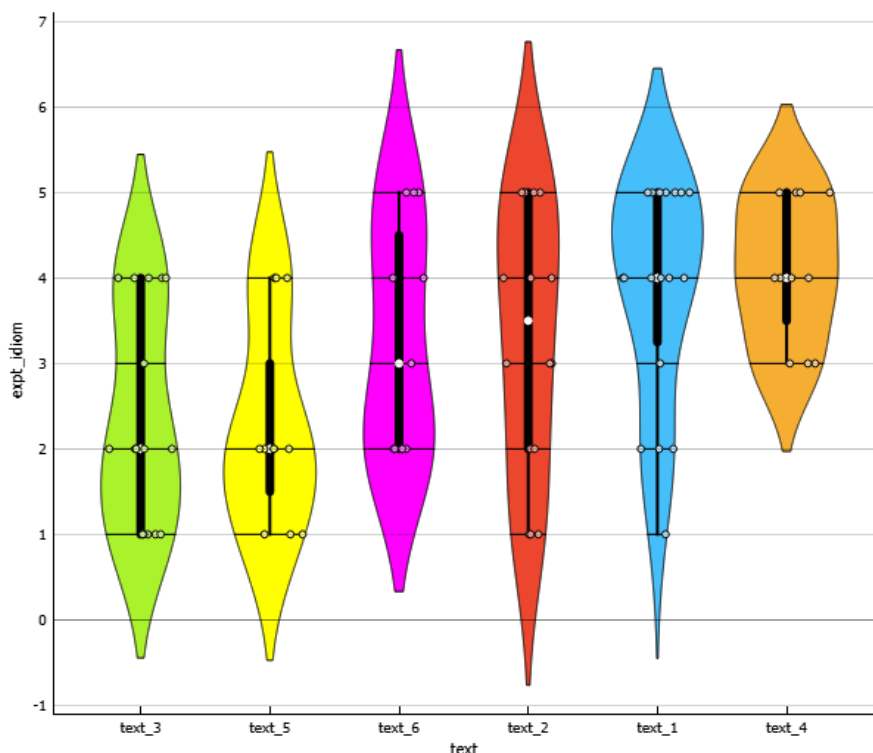
Таблица 7. Средние значения и стандартное отклонение в категории «Выразительность»
Table 7. Mean values and standard deviation in the category “Expressiveness”

Подкатегория/текст	Тип значения	текст 1	текст 2	текст 3_gen	текст 4	текст 5_gen	текст 6_gen
8.1 Идиоматичность	mean	3,8	3,3	2,3	4,1	2,3	3,2
	std	1,1	1,3	1,2	0,9	1,1	1,3
8.2 Сильные позиции	mean	3,3	3,3	2,5	3,5	2,7	3,4
	std	0,9	1,2	1,3	0,9	1,5	1,5
8.3 Креативность	mean	3,2	3,3	2,3	3,6	1,9	3,5
	std	1,3	1,5	1,3	0,8	1,2	1,3

По количеству идиоматических выражений сгенерированный текст № 6 незначительно уступает трем написанным текстам (Рисунок 9, Таблица 7). На диаграмме (Рисунок 9) видно, что при

сопоставимом с естественными текстами среднем значении, тем не менее, значительное скопление не усредненных оценок для текста № 6 наблюдаем в районе 4-х.

Рисунок 9. Распределение значений оценок (вопрос 8.1.) в категории «Выразительность»
Figure 9. Distribution of values for question 8.1. in the category “Expressiveness”



Такая достаточно высокая оценка не случайна – в тексте находим более 100 идиоматических выражений, например: *walking on eggshells* ‘вести себя крайне осторожно в деликатной ситуации’,

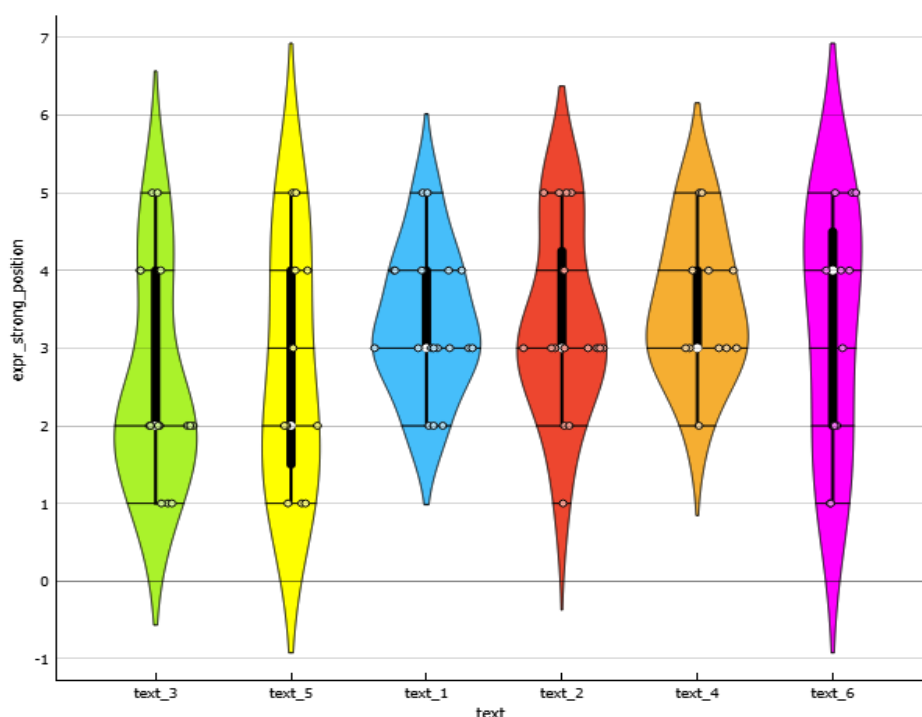
walking a tightrope ‘балансировать в сложной ситуации’, *read between the lines* ‘видеть скрытые смыслы’, *burn bridges* ‘разорвать всякие отношения’, *skeletons in the closet* ‘семейные тайны’ и т.д.

По наличию же так называемых «Сильных позиций» по средним значениям текст № 6 оказался на второй позиции после естественного текста № 4, а по общему распределению не усредненных

значений – на первой (Рисунок 10), поскольку демонстрирует незначительное «утолщение» в зоне 4-х, а максимальные значения приближаются к 5.

Рисунок 10. Распределение значений на вопрос 8.2. в категории «Выразительность»

Figure 10. Distribution of values for question 8.2. in the category “Expressiveness”



Действительно, в тексте № 6, например, в финале истории (пр.5) можем наблюдать, как усиливается эмоциональное напряжение героев, при этом в тексте концентрируются метафоры (*McGonagall's melodious voice ushered in an era of transformation; conversations mushroomed; parallel fascinations ignited Pansy's enthusiasm; recalibrating his internal ledger of favors owed and debts outstanding; timely investments in influential relationships* etc.) и метафорические сравнения (*fanning out like tender shoots bursting forth from fertile soil*), что производит конвергентный эффект и усиливает производимое на читателя впечатление:

(5) *Moments later, McGonagall's melodious voice ushered in an era of transformation, announcing the commencement of specialized coursework catering to individual strengths.*

Conversations mushroomed, exploring the vast spectrum of electives and concentrations unveiled before them. Theo listened intently, his pulse quickening as Ancient Runes and Archaeological Magic topped his list of preferred subjects. Parallel fascinations ignited Pansy's enthusiasm, her affinity for Curse-Breaking and Dark Arts Countermeasures aligning harmoniously with his own pursuits.

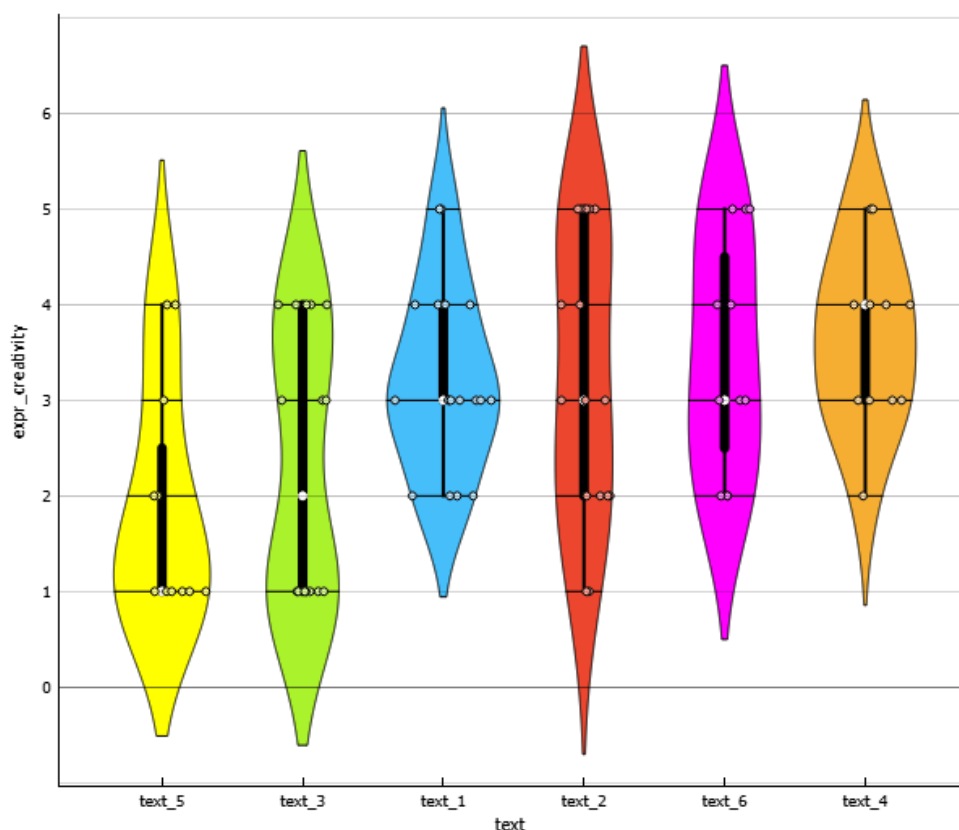
Through the cacophony of congratulatory cheers and sympathetic condolences, Blaise Zabini maintained a calculated distance, monitoring shifting allegiances and recalibrating his internal ledger of favors owed and debts outstanding. He knew full well that timely investments in influential relationships often proved decisive in securing

advantageous positions within the wizarding hierarchy.

Nearby, Ron Weasley pumped Hermione's hand aloft, basking in the glory of her triumphant entry into elite Magical Law Enforcement internship programs. Michael Corner chuckled knowingly, attributing her success to razor-sharp intellect and diligent preparation. Friendship warmed the atmosphere, reinforcing bonds tested by communal struggles and victories.

Discreetly withdrawing from the melee, Theo and Pansy commandeered a vacant study area, immersing themselves in meticulous analyses of recommended curricula. Visions of interdisciplinary collaboration sprouted, fanning out like tender shoots bursting forth from fertile soil. Convergent paths merged seamlessly, forecasting trail-blazing innovations primed to revolutionize the wizarding community.

Рисунок 11. Распределение значений на вопрос 8.3. в категории «Выразительность»
Figure 11. Distribution of values for question 8.3. in the category “Expressiveness”



Что касается креативности, то текст № 4 оценивается выше других, демонстрируя скопление оценок в районе 4-х, текст 1 – в районе 3-х, а тексты 2 и 6, несмотря на разный источник продуцирования (человек – для текста № 2 и модель – для текста № 6), сходный паттерн оценивания – низкую согласованность оценок, на что указывает вытянутая форма «скрипок» (Рисунок 11).

6. Обсуждение

Поскольку в экспериментах с БЯМ команды компьютерных лингвистов преследовали цель добиться иерархичности повествования, используя при этом разные техники, в категории «Когерентность» мы наблюдаем наименьший разрыв между написанными и сгенерированными текстами. Более значительные отличия видятся там, где речь идет о связности, ритме, эмоциях и модальности сомнения – сгенерированные

тексты оцениваются по данным критериям ниже, чем написанные. В категориях же сложности синтаксиса и выразительности различия вновь смягчаются, поскольку, например, сгенерированный текст № 6 оказался в числе лидеров и по сложности синтаксических структур, и по выделенности сильных позиций, и по креативности стиля. С данными наблюдениями отчасти согласуются и формальные метрики, предварительно измеренные для всех текстов – текст № 6 оказался самым сложным по метрике Флеша-Кинкейда, с самым высоким числом уникальных слов и самыми длинными словами и предложениями.

Таким образом, именно текст № 6, сгенерированный БЯМ после дообучения пересказу на датасете художественной литературы из проекта Гуттенберг, оказался наиболее близок по большинству метрик к естественным текстам. Представляется важным, что именно этот текст показал наиболее четкий из всех сгенерированных текстов кластер векторов пересказов, что свидетельствует о его приближении к паттернам связности. Текст № 6 также продемонстрировал хороший уровень генерации эмотивности в тексте для создания определенной эмоциональной тональности, ср. фрагмент из текста:

(6) *Through the quiet intensity of their partnership, boundaries softened. Conversations flowed freely, touching on everything from favored Quidditch matches to childhood memories. Shared laughs erased residual awkwardness, allowing them to settle comfortably into an easy camaraderie. Without realizing it, they discovered solace in each other's company, finding strength in their differences rather than weaknesses.*

Синтетическому автору удастся воссоздать позитивную эмоциональность, эмпатически передать читателю атмосферу комфортного дружеского общения:

границы стирались, разговоры текли свободно, общий смех растворял остающуюся неловкость, дружеская близость (пр.6).

Однако, главный вывод, который мы можем сделать, на наш взгляд, состоит в том, что первичная посылка организаторов соревнования о том, что все дело в иерархичности нарративной схемы, в общем, не оправдала себя. Более важными свойствами художественного текста любительской прозы оказались наличие динамичного меняющегося ритма повествования, его подчиненность логике причинно-следственных связей, присутствие сомнения, выраженного модальными языковыми средствами либо оценочными лексемами. Синтетическим авторам пока не удастся создать не только сюжет, но и микромир, атмосферу повествования. Работая с разнообразными стилистическими приемами, модели, тем не менее, пока не могут предвосхитить ожидания читателей и, например, обмануть их, что заставляет читателей трактовать сюжет как “скучный”, не обнаруживающий психологизма:

(7) *Сложные предложения, нагруженные непростыми словами, не несут за собой художественной ценности, а лишь создают шум вокруг скучного сюжета, который не до конца ясен. Частые переходы в повествовании от одного персонажа к другому не позволяют погрузиться в сюжет и представляют собой мимоходное повествование, в котором отсутствует психологизм*
(комментарий респондента по тексту № 6).

Справедливости ради отметим, что аналогичные упреки артикулировались информантами и в отношении естественных текстов 1, 2, 4, написанных авторами-любителями. В этом смысле, несомненно, стоит принимать во внимание тот факт, что сам жанр любительской прозы по мотивам известного

произведения не позволяет в полной мере предъявлять к таким текстам требования эстетической ценности и психологической глубины.

Таким образом, если мы говорим о некотором каноне восприятия художественного текста нашими информантами, то проведенное пилотное исследование демонстрирует, что когерентность и когезия, например, находятся среди периферийных признаков “настоящего” художественного текста для наших информантов. Центральное же место, по-видимому, занимает ритм, легкое переключение между жанрами и нарративными позициями, эмпатичность и эмоциональность повествования, свидетельствующая о присутствии нарратора. Приведем фрагмент комментария одного из информантов к тексту № 4, который по результатам оценивания имел низкие баллы за параметры, связанные с когерентностью и связанностью, но высокие – за ритм, эмоциональность и выразительность:

(8) *В тексте частое подробное описание вещей, людей, создаётся «естественное» ощущение рассказа от человека с детальным погружением. Обозначение громкости и долготы речи также способствует созданию естественности. Резкий переход с сюжетного повествования к эпистолярному также указывает на человеческий фактор. Также стоит заметить свободное переключение от общего фона повествования в третьем лице к монологическому без потери логичности, что очеловечивает наррацию.*

Методологический вклад результатов проведенного на уникальных данных эксперимента состоит в том, что они демонстрируют необходимость введения новых метрик для оценки качества длинной генерации художественных текстов, часть из которых была нами апробирована – это ритм, креативность, модальность сомнения и др.. Сложившийся в среде исследователей в

области компьютерных наук перечень существующих метрик (мы его привели в теоретической части статьи) демонстрирует свою ограниченность в силу фиксации на четкости логических отношений между событиями и персонажами, между разными уровнями схемы повествования. При этом за рамками исследований остаются собственно филологические и психологические аспекты рецепции текста, хорошо изученные лингвистами на материале естественных текстов как в теоретическом, так и в экспериментальном ключе. Прием, связанный с кластеризацией пересказов текстов как показателем целостности и связности, представляется нам перспективным. Дальнейшие наши эксперименты, возможно, будут связаны с его валидацией.

Несомненно, исследование имеет ряд значительных ограничений. Во-первых, достаточно специфический и узкий круг информантов, принимавших участие в эксперименте. Однако это связано с большой хронозатратностью задания: информанту нужно было прочитать около 100 страниц текста на неродном языке, а затем ответить на опросник по каждому тексту отдельно, написать комментарии и пересказы. В данных есть пробелы – например, не все информанты написали пересказы для всех текстов, поэтому наблюдается неравномерность в их распределении, что важно, например, для качества кластеризации. Мы намеренно приняли решение пригласить в качестве информантов студентов-филологов как опытных читателей, привыкших внимательно относиться к слову. Очевидно, что привлечение к эксперименту другой группы информантов привело бы к получению несколько иных результатов.

7. Заключение

Проведенные эксперименты показали, что благодаря специальным методам, сочетающим в себе промптинг и дообучение, можно решить задачу

автоматической генерации длинного текста с развернутой иерархически организованной нарративной структурой.

В то же время, автоматическая генерация пока не позволяет получить тексты, отвечающие критериям наличия динамически меняющегося ритма повествования и развертывания сюжета; таким синтетическим текстам пока не свойственна субъективная модальность сомнения.

Однако применение метода дообучения модели для задачи пересказа художественных произведений позволило получить текст, сопоставимый с естественными текстами, написанными непрофессиональными авторами, по параметрам идиоматичности, креативности, способности маркировать сильные позиции в тексте, выражать эмоции и создавать эффект эмоциональной тональности.

Кроме того, важной в методологическом смысле представляется сама постановка вопроса о критериях оценки художественности текста в условиях конкуренции авторов-людей и авторов-моделей.

В данном исследовании мы оставили «за скобками» категории эстетической и этической ценности текста, но сосредоточились на таких категориях, которые обозначают границу между естественным и искусственным интеллектом: когерентность, смысловая и лексико-грамматическая когезия, ритм текста, эмоциональность (тональность), присутствие субъективной модальности сомнения, сложность синтаксиса и выразительность (наличие сильных позиций, использование идиом и стилистических приемов).

Психолингвистический эксперимент продемонстрировал, что в нашей выборке дифференцирующую функцию в наибольшей степени выполняют оба вида связности, ритм и субъективная модальность. Остальные категории нашли свою реализацию лишь в одном из

сгенерированных текстов, что подтверждает тот факт, что данные категории составляют горизонт ближайшего развития для генеративного ИИ.

Заявление об использовании генеративного ИИ: Автор заявляет, что при проведении исследования и подготовке настоящей рукописи генеративный искусственный интеллект и ИИ-ассистированные технологии, подлежащие раскрытию в соответствии с политикой журнала, не использовались.

Statement on the Use of Generative AI: The author declares that no generative artificial intelligence or AI-assisted technologies requiring disclosure under the Journal's policy were used in conducting the research or preparing this manuscript.

Список литературы

- Арнольд И. В. Семантика. Стилистика. Интертекстуальность: сборник научных трудов / И. В. Арнольд; научный редактор П. Е. Бухаркин. М.: ФЛИНТА, 2019. 448 с.
- Базилевский М. П. Исследование новых критериев для обнаружения автокорреляции остатков первого порядка в регрессионных моделях // Математика и математические модели. 2018. С. 13–25. DOI:10.24108/MATHM.0318.0000102
- Гальперин И. Р. Текст как объект лингвистического исследования. М.: URSS, 2007. 137 с.
- Гиршман М. Ритм и целостность прозаического художественного произведения / М. Гиршман // Вопросы литературы. 1974. № 11. С. 128–150.
- Лотман Ю. М. Структура художественного текста // Лотман Ю. М. Об искусстве. СПб.: «Искусство – СПб», 1998. 383 с.
- Матвеева Т. В. Функциональные стили в аспекте текстовых категорий: Синхронно-сопоставительный очерк. Свердловск: Изд-во Урал. ун-та, 1990. 172 с.
- Шаховский В. И. Эмоции: Долингвистика, лингвистика, лингвокультурология. М.: Либроком, 2010. 128 с.

Шкловский В. Строение рассказа и романа // О теории прозы. М.: Изд-во «Круг», 1925. С. 56–69.

Boriskin A., Galimzianova D. The LSG ChallengeWorkshop at INLG 2024: Prompting Techniques for Crafting Extended Narratives with LLMs // Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges, Tokyo, Japan. Association for Computational Linguistics. P. 118–122.

Chakrabarty T., Padmakumar V., He H. Help Me Write a Poem: Instruction Tuning as a Vehicle for Collaborative Poetry Writing. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022). Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022. P. 6848–6863.

Charolles M. Introduction aux problèmes de la cohérence des textes // Langue française. 1978. Vol. 38. P. 7–41.

Charaudeau P., Maingueneau D. Dictionnaire d'Analyse du Discours. Paris: Seuil, 2002. 661 p.

Combettes B. Thématization et progression thématique dans les récits d'enfants // Langue française. 1978. Vol. 38. Pp. 74–86.

Flesch R. A new readability yardstick // Journal of Applied Psychology. 1948. № 32 (3). Pp. 221–233.

Graesser A. C., Singer M., & Trabasso T. Constructing inferences during narrative text comprehension. Psychological Review. 1994. Vol. 101(3). Pp. 371–395.

Guan J., Huang M. UNION: An Unreferenced Metric for Evaluating Open-ended Story Generation // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Online: Association for Computational Linguistics, 2020. Pp. 9157–9166.

Iaroshenko P. V., Louckachevitch N. V. Large Language Models versus Native Speakers in Emotional Assessment of Russian Words // Supercomputing Frontiers and Innovations. 2025. Vol. 12. No. 3. Pp. 20–30. DOI: 10.14529/jsfi250302

Ippolito D., Yuan A., Coenen A., Burnam S. Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers. 2022. DOI: 10.48550/arXiv.2211.05030

Johnson-Laird, P. N. Mental Models: Towards a Cognitive Science of Language,

Inference, and Consciousness. Cambridge, MA: Harvard University Press, 1983. 513 p.

Jin Y., Kadam V., Wanvarie D. Plot Writing From Pre-Trained Language Models // Proceedings of the 15th International Conference on Natural Language Generation (INLG 2022). 2022. Pp. 52–67.

Kolmogorova A., Kulikova E., Lobanov V. Aspect-Based Sentiment Analysis Using Large Language Models on Museum Visitor Reviews // Supercomputing Frontiers and Innovations. 2025. Vol. 12. No. 3. Pp. 121–140. <https://doi.org/10.14529/jsfi250309>

Kryscinski W., Keskar N. S., McCann B., Xiong C., Socher R. Neural Text Summarization: A Critical Evaluation // 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: Association for Computational Linguistics, 2019. Pp. 540–551.

Lee Y., Ka S., Son B., Kang P., Kang J. Navigating the Path of Writing: Outline-guided Text Generation with Large Language Models // arXiv:2404.13919. <https://doi.org/10.48550/arXiv.2404.13919>

Lewis P., Perez E., Piktus A., Petroni F., Karpukhin V., Goyal N., Küttler H., Lewis M., Yih W., Rocktäschel T., Riedel S., Kiela D. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks // arXiv:2005.11401v4. 2021. DOI: 10.48550/arXiv.2005.11401.

Mann W. C., Thompson S. A. Rhetorical Structure Theory: Toward a functional theory of text organization. 1988. Text, 8 (3). Pp. 243–281.

Migal A., Seredina D., Telnina L., Nazarov N., Kolmogorova A., Mikhaylovskiy N. Overview of Long Story Generation Challenge (LSGC) at INLG 2024 // Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges. Tokyo, Japan: Association for Computational Linguistics. Pp. 47–53.

Mikhaylovskiy N. States of LLM-generated Texts and Phase Transitions between them // arXiv:2503.06330 [cs.CL]. DOI: 10.48550/arXiv.2503.06330.

Seredina D. A Report on LSG 2024: LLM Fine-Tuning for Fictional Stories Generation // Proceedings of the 17th International Natural Language Generation Conference: Generation

Challenges. Tokyo, Japan: Association for Computational Linguistics. Pp. 123–127.

Shen C., Cheng L., Nguyen X., You Y., Bing L. Large Language Models are Not Yet Human-Level Evaluators for Abstractive Summarization. 2023. <https://arxiv.org/abs/2305.1309>. DOI: 10.48550/arXiv.2305.13091

Singh N., Bernal G., Savchenko D., Glassman E. L. Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence // ACM Transactions on Computer-Human Interaction. 2023. Vol. 30 (5). Pp. 1–57.

Takahashi S., Tanaka-Ishii K. Evaluating computational language models with scaling properties of natural language // Computational Linguistics. 2019. Vol. 45 (3). Pp. 481–513. DOI: 10.48550/arXiv.1906.09379

Vatolin A. Structured Sentiment Analysis with Large Language Models: A Winning Solution for RuOpinionNE-2024 // Proceedings of International Conference on Computational Linguistics and Intelligent Technologies Dialogue 2025. 2025.

Zhang Y., Sun Q., Guo Q., Wu J., Qiu X., Huang X. Do Large Language Models Know What They Don't Know? // arXiv:2210.03025. 2022. DOI: 10.48550/arXiv.2209.01515

Zheng L., Chiang W., Sheng Y., Zhuang S., Wu Z., Zhuang Y., Lin Z., Li Z., Li D., Xing E. P., Zhang H., Gonzalez J. E., Stoica I. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena // arXiv:2306.05685v4. DOI: 10.48550/arXiv.2306.05685.

Zhong W., Naradowsky J., Takamura H., Kobayashi I., Miyao Y. Fiction-writing mode: An effective control for human-machine collaborative writing // Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. 2023. Pp. 1752–1765.

Zhou D., Serman S. Ai.llude: Encouraging Rewriting AI-Generated Text to Support Creative Expression. 2024. DOI: 10.48550/arXiv.2405.17843

Материалы исследования

Олеша Ю. Ни дня без строчки. М.: ФТМ, 1961. 464 с.

Филиппов К. А. Лингвистика текста: Курс лекций. СПб.: Изд-во С.-Петерб. ун-та, 2007. 329 с.

References

Arnold, I. V. (2019). *Semantics. Stylistics. Intertextuality: Collection of Scientific Papers*, FLINTA, Moscow, 448 p. (In Russian)

Bazilevskiy, M. P. (2018). Research of new criteria for detecting first-order residual autocorrelation in regression models, *Mathematics and Mathematical Models*, 13–25. DOI: 10.24108/MATHM.0318.0000102 (In Russian)

Galperin, I. R. (2007). *Text as an Object of Linguistic Research*, URSS, Moscow, 137 p. (In Russian)

Girshman, M. (1974). Rhythm and Integrity of Prose Literary Work, *Questions of Literature*, 11, 128–150. (In Russian)

Lotman, Yu. M. (1998). Structure of the Artistic Text, in Lotman Yu. M. *On Art*, Iskustvo – SPB, St. Petersburg, 383 p. (In Russian)

Matveeva, T. V. (1990). *Functional Styles in Terms of Text Categories: A Synchronic-Comparative Essay*, Ural University Press, Sverdlovsk, 172 p. (In Russian)

Shakhovskiy, V. I. (2010). *Emotions: Pre-Linguistics, Linguistics, Linguoculturology*, Librokom Moscow, 128 p. (In Russian)

Shklovskiy, V. (1925). Structure of the Short Story and Novel, in *On the Theory of Prose*, Krug Publishing House, Moscow, 56–69. (In Russian)

Boriskina, A. and Galimzianova, D. (2024). The LSG ChallengeWorkshop at INLG 2024: Prompting Techniques for Crafting Extended Narratives with LLMs, *Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges*, Tokyo, Japan, Association for Computational Linguistics, 118–122. (In English)

Chakrabarty, T., Padmakumar, V. and He, H. (2022). Help Me Write a Poem: Instruction Tuning as a Vehicle for Collaborative Poetry Writing, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*, Abu Dhabi, United Arab Emirates, Association for Computational Linguistics, 6848–6863. (In English)

Charolles, M. (1978). Introduction aux problèmes de la cohérence des textes, *Langue française*, 38, 7–41. (In French)

Charaudeau, P. and Maingueneau, D. (2002). *Dictionnaire d'Analyse du Discours*, Seuil, Paris, 661 p. (In French)

- Combettes, B. (1978). Thématisation et progression thématique dans les récits d'enfants, *Langue française*, 38, 74–86. (In French)
- Flesch, R. (1948). A new readability yardstick, *Journal of Applied Psychology*, 32 (3), 221–233. (In English)
- Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension, *Psychological Review*, 101(3), 371–395. (In English)
- Guan, J. and Huang, M. (2020). UNION: An Unreferenced Metric for Evaluating Open-ended Story Generation, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online, Association for Computational Linguistics, 9157–9166. (In English)
- Iaroshenko, P.V. and Louckachevitch, N.V. (2025). Large Language Models versus Native Speakers in Emotional Assessment of Russian Words, *Supercomputing Frontiers and Innovations*, 12, 3, 20–30. DOI: 10.14529/jsfi250302 (In English)
- Ippolito, D., Yuan, A., Coenen, A. and Burnam, S. (2022). Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers. DOI: 10.48550/arXiv.2211.05030 (In English)
- Jin, Y., Kadam, V. and Wanvarie, D. (2022). Plot Writing From Pre-Trained Language Models, *Proceedings of the 15th International Conference on Natural Language Generation (INLG 2022)*, 52–67. (In English)
- Johnson-Laird, P. N. (1983). *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press, 513 p. (In English)
- Kolmogorova, A., Kulikova, E. and Lobanov, V. (2025). Aspect-Based Sentiment Analysis Using Large Language Models on Museum Visitor Reviews, *Supercomputing Frontiers and Innovations*, 12, 3, 121–140. DOI: 10.14529/jsfi250309 (In English)
- Kryscinski, W., Keskar, N.S., McCann, B., Xiong, C. and Socher, R. (2019). Neural Text Summarization: A Critical Evaluation, *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, Association for Computational Linguistics, 540–551. (In English)
- Lee, Y., Ka, S., Son, B., Kang, P. and Kang, J. Navigating the Path of Writing: Outline-guided Text Generation with Large Language Models. arXiv:2404.13919 [cs.CL]. (In English)
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S. and Kiela, D. (2021). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv:2005.11401v4. DOI: 10.48550/arXiv.2005.11401 (In English)
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization, *Text*, 8(3), 243–281. (In English)
- Migal, A., Seredina, D., Telnina, L., Nazarov, N., Kolmogorova, A. and Mikhaylovskiy, N. (2024). Overview of Long Story Generation Challenge (LSGC) at INLG 2024, *Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges*, Tokyo, Japan, Association for Computational Linguistics, 47–53. (In English)
- Mikhaylovskiy, N. States of LLM-generated Texts and Phase Transitions between them. arXiv:2503.06330 [cs.CL]. DOI: 10.48550/arXiv.2503.06330 (In English)
- Seredina, D. (2024). A Report on LSG 2024: LLM Fine-Tuning for Fictional Stories Generation, *Proceedings of the 17th International Natural Language Generation Conference: Generation Challenges*, Tokyo, Japan, Association for Computational Linguistics, 123–127. (In English)
- Shen, C., Cheng, L., Nguyen, X., You, Y. and Bing, L. (2023). Large Language Models are Not Yet Human-Level Evaluators for Abstractive Summarization. DOI: 10.48550/arXiv.2305.13091 (In English)
- Singh, N., Bernal, G., Savchenko, D. and Glassman, E.L. (2023). Where to hide a stolen elephant: Leaps in creative writing with multimodal machine intelligence, *ACM Transactions on Computer-Human Interaction*, 30 (5), 1–57. (In English)
- Takahashi, S. and Tanaka-Ishii, K. (2019). Evaluating computational language models with scaling properties of natural language, *Computational Linguistics*, 45 (3), 481–513. DOI: 10.48550/arXiv.1906.09379 (In English)
- Vatolin, A. (2025). Structured Sentiment Analysis with Large Language Models: A

Winning Solution for RuOpinionNE-2024, Proceedings of International Conference on Computational Linguistics and Intelligent Technologies Dialogue 2025. (In English)

Zhang, Y., Sun, Q., Guo, Q., Wu, J., Qiu, X. and Huang, X. (2022). Do Large Language Models Know What They Don't Know? arXiv:2210.03025. DOI:

10.48550/arXiv.2209.01515 (In English)

Zheng, L., Chiang, W., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., Lin, Z., Li, Z., Li, D., Xing, E.P., Zhang, H., Gonzalez, J.E. and Stoica, I. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. arXiv:2306.05685v4. DOI: 10.48550/arXiv.2306.05685 (In English)

Zhong, W., Naradowsky, J., Takamura, H., Kobayashi, I. and Miyao, Y. (2023). Fiction-writing mode: An effective control for human-machine collaborative writing, *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, 1752–1765. (In English)

Zhou, D. and Serman, S. (2024). Ai.llude: Encouraging Rewriting AI-Generated Text to Support Creative Expression. DOI: 10.48550/arXiv.2405.17843 (In English)

Corpus Materials

Olesha, Yu. (1961). *Not a Day Without a Line*, FTM, Moscow, 464 p. (In Russian)

Filippov, K. A. (2007). *Text Linguistics: A Course of Lectures*, St. Petersburg University Press, St. Petersburg, 329 p. (In Russian)

The author has read and approved the final manuscript.

Автор прочитал и одобрил окончательный вариант рукописи.

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

Конфликты интересов: у автора нет конфликтов интересов для декларации.

Анастасия Владимировна Колмогорова, доктор филологических наук, профессор, заведующий лабораторией языковой конвергенции НИУ ВШЭ Санкт-Петербург, Россия.

Anastasia V. Kolmogorova, Doctor of Philology, Professor, Head of the Laboratory of Language Convergence, National Research University Higher School of Economics, St. Petersburg, Russia.