

ISSN 2313-8912

НАУЧНЫЙ РЕЗУЛЬТАТ

ВОПРОСЫ ТЕОРЕТИЧЕСКОЙ И ПРИКЛАДНОЙ ЛИНГВИСТИКИ

RESEARCH RESULT. THEORETICAL AND APPLIED LINGUISTICS

10(3) 2024

16+

Сайт журнала:
rilinguistics.ru

сетевой научный рецензируемый журнал
online scholarly peer-reviewed journal



НАУЧНЫЙ РЕЗУЛЬТАТ. ВОПРОСЫ ТЕОРЕТИЧЕСКОЙ И ПРИКЛАДНОЙ ЛИНГВИСТИКИ

RESEARCH RESULT. THEORETICAL AND APPLIED LINGUISTICS

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



Том 10, № 3. 2024
СЕТЕВОЙ НАУЧНО-ПРАКТИЧЕСКИЙ
ЖУРНАЛ
Издается с 2014 г.
ISSN 2313-8912



Volume 10 (3). 2024
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

Воркачёв 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

Порхомовский В.Я., доктор филологических наук, профессор, Институт языкознания РАН, г. Москва, Россия

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. Ополье, Польша

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

EDITORIAL TEAM:

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

EXECUTIVE SECRETARY: **Elena V. Lukyanova**, Ph.D. in Philology, Federal State Autonomous Educational Institution of Higher Education "Belgorod State National Research University", Associate Professor, 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

Alexandr 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. Ogneva, Doctor of Philology, Federal State Autonomous Educational Institution of Higher Education "Belgorod State National Research University", Professor, Belgorod, Russia

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

Victor Ya. Porkhomovskiy, Doctor of Philology, professor, Linguistics Institute Moscow, Moscow, Russia

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

ЯЗЫКОЗНАНИЕ

LINGUISTICS

РАЗДЕЛ I. ТЕОРИЯ ЯЗЫКА

SECTION I. THEORY OF LANGUAGE

Аматов А. М., Кураш С. Б., Данилова Е. С. Институциональность дискурса как методологическая проблема	4	Amatov Alexander M., Kurash Sergey B., Danilova Elena S. Institutionality of discourse as a methodological problem	4
Дубинина М. Н. Цифровые и национальные домены: тенденции ономастики в интернет- коммуникации	26	Dubinina Maria N. Numeric Domain Names and National Domains: Trends in Internet Onomastics	26
Савельева И. В., Мельник Н. В. Рецептивно-продуктивная деятельность субъектов непрофессионального интернет- дискурса	42	Savelyeva Irina V., Mel'nik Natalya V. Receptive and productive activity of non-professional Internet discourse participants	42

РАЗДЕЛ II.

SECTION II.

ПРИКЛАДНАЯ ЛИНГВИСТИКА

APPLIED LINGUISTICS




Гатиятуллина Г. М., Гафиятова Э. В., Данилов А. В. Динамика уровня сложности как функция морфологических параметров текста на русском языке	70	Gatiyatullina Galiya M., Gafiyatova Elzara V., Danilov Andrei V. Text complexity increase in Russian texts as a function of morphological changes	70
Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма: библиометрический обзор с использованием инструментов R и VOSviewer	91	Dekhnich Olga V., Litvinova Tatyana A. A keystroke analysis to study writing: a bibliometric review using R and VOSviewer	91
Sanosi A. B., Mohammed A. S. E. ASAWEC: towards a corpus of Arab scholars' academic written English	116	Sanosi A. B., Mohammed A. S. E. ASAWEC: towards a corpus of Arab scholars' academic written English	116
Фирсанова В. И. Новый графовый подход к генерации текстов узкой предметной области на естественном языке	135	Firsanova Victoria I. A graph-based approach to closed-domain natural language generation	135

РАЗДЕЛ III. ЛИТЕРАТУРЫ НАРОДОВ МИРА	SECTION III. WORLD LITERATURES
Karen A., Christopher G. Anthropocentric Views on Women and Nature through the Concepts of Gyn/Ecology and Ecofeminism in Anuradha Roy's <i>The Folded Earth</i> 168	Karen A., Christopher G. Anthropocentric Views on Women and Nature through the Concepts of Gyn/Ecology and Ecofeminism in Anuradha Roy's <i>The Folded Earth</i> 168
Chakravarty K., Thenmozhi M. Interpretation of Human Emotion: An Implementation of Compassion through Karuna Rasa in Manjula Padmanabhan's <i>Lights Out</i> 180	Chakravarty K., Thenmozhi M. Interpretation of Human Emotion: An Implementation of Compassion through Karuna Rasa in Manjula Padmanabhan's <i>Lights Out</i> 180
Шулятьева Д. В. Опыт погони: нарративная скорость и ритм в современных разветвленных нарративах 194	Shulyatyeva Dina V. The chase experience: narrative speed and rhythm in contemporary forking-path narratives 194

ЯЗЫКОЗНАНИЕ
РАЗДЕЛ I. ТЕОРИЯ ЯЗЫКА
LINGUISTICS
SECTION I. THEORY OF LANGUAGE

УДК 81'42

DOI: 10.18413/2313-8912-2024-10-3-0-1

Аматов А. М.¹ 
Кураш С. Б.² 
Данилова Е. С.³ 

**Институциональность дискурса как
методологическая проблема**

¹ Белгородский государственный национальный исследовательский университет
ул. Победы, 85, Белгород, 308015, Россия
E-mail: 1
ORCID: 0000-0002-5530-3274

² Мозырский государственный педагогический университет имени И. П. Шамякина
ул. Студенческая, 28, Мозырь, 247760, Республика Беларусь
E-mail: text2005@mail.ru
ORCID: 0000-0003-2317-7872

³ Белгородский государственный национальный исследовательский университет
ул. Победы, 85, Белгород, 308015, Россия
E-mail: elena_danilova@bsu.edu.ru
ORCID: 0000-0003-1556-3116

*Статья поступила 30 августа 2024 г.; принята 15 сентября 2024 г.;
опубликована 30 сентября 2024 г.*

Аннотация: Институциональность дискурса, являясь одним из ключевых понятий современного дискурс-анализа и неотъемлемой характеристикой целого ряда видов дискурса, нередко представляет собой исследовательскую проблему не только практического, но и методологического свойства. Динамичность дискурса, его социальный характер часто не вмещаются в рамки институциональности, расширяя или даже нарушая их. Помимо общих методологических проблем, исследователь нередко сталкивается со случаями сознательного нарушения писанных и неписанных правил речевого поведения. Подобные примеры представляют собой определенную проблему классификации дискурса и методологии дискурс-анализа, но в то же время позволяют взглянуть на проблему институциональности под другим углом. В статье рассматриваются проблемы методологии институционального дискурса, связанные, с одной стороны, с неопределенностью (или недостаточно четкой определенностью) ряда понятий и видов дискурса, а с другой стороны – с нарушением статуса участников вплоть до смены их ролей. На первый взгляд, подобные нарушения идут вразрез с самим принципом институциональности. В работе будет показано, что с функциональной точки зрения институциональность не только сохраняется, но и служит орудием

манипуляции. Рассмотрены примеры таких коммуникативных нарушений, представляющих как методологические, так и практические затруднения дискурс-исследований. Показано, что формальная классификация видов дискурса через соответствующие им социальные институты при ее простоте неизбежно приводит к ряду проблем, таких как пересечение нескольких типов дискурса в одном дискурсивном акте или трудности отнесения к какому-либо типу по причине изменения речевых тактик участников дискурса.


Ключевые слова: Методология дискурс-анализа; Институциональность дискурса; Политический дискурс; Медиадискурс; Педагогический дискурс; Статус коммуникантов; Манипуляция

Информация для цитирования: Амагов А. М., Кураш С. Б., Данилова Е. С. Институциональность дискурса как методологическая проблема // Научный результат. Вопросы теоретической и прикладной лингвистики. 2024. Т. 10. № 3. С. 4-25. DOI: 10.18413/2313-8912-2024-10-3-0-1

UDC 81'42

DOI: 10.18413/2313-8912-2024-10-3-0-1

Alexander M. Amатов¹ 

Sergey B. Kurash² 

Elena S. Danilova³ 

Institutionality of discourse as a methodological problem

¹ Belgorod State National Research University,
85 Pobedy St., Belgorod, 308015, Russia
E-mail: amatov@bsu.edu.ru
ORCID: 0000-0002-5530-3274

² I. P. Shamyakin Mozyr State Pedagogical University,
28 Studencheskaya St., Mozyr, 247760, Belarus
E-mail: text2005@mail.ru
ORCID: 0000-0003-2317-7872

³ Belgorod State National Research University,
85 Pobedy St., Belgorod, 308015, Russia
E-mail: elena_danilova@bsu.edu.ru
ORCID: 0000-0003-1556-3116

Received 30 August 2024; accepted 15 September 2024; published 30 September 2024

Abstract: Institutionality of discourse, as one of the key concepts of modern discourse analysis and an integral property of a number of discourse types, often poses a research problem of practical and/or methodological character. The dynamics of discourse and its social nature do not always fit into the framework of institutionality, extending it or even transgressing it. Apart from general methodological problems, the researcher often encounters cases of deliberate violation of written and unwritten rules of communication. Such examples, however, not only represent an issue, but also grant the possibility of looking at the problem of institutionality from a different angle. The article considers the problems of the methodology of institutional discourse associated, on the one hand, with the uncertainty (or an insufficiently clear definition) of a number of concepts and types

of discourse, and on the other hand, with shifts of status and roles of communicants. It might seem that such shifts make the principle of institutionality rather volatile. Our study shows, however, that functionally all the institutional properties of discourse stand and even serve as a tool for manipulation. We consider examples of such violations that present a methodological issue and hamper practical discourse research. It is shown that a formal classification of discourse types via corresponding social institutions, despite its simplicity, tends to produce a number of issues including the overlapping of several types of discourse in one discursive act or difficulties in attributing it to any particular type due to changes in the communicative tactics of participants.

Keywords: Methodology of discourse analysis; Institutionalinity of discourse; Political discourse; Media discourse; Academic discourse; Communicators' status; Manipulation

How to cite: Amatov, A. M., Kurash, S. B., Danilova, E. S. (2024). Institutionalinity of discourse as a methodological problem, *Research Result. Theoretical and Applied Linguistics*, 10 (3), 4-25. DOI: 10.18413/2313-8912-2024-10-3-0-1

Введение

Большое количество исследований в области дискурса, с одной стороны, предоставляет в наше распоряжение богатый инструментарий. С другой стороны, всякое серьезное исследование в области дискурса начинается с определения исследуемого понятия, что порождает большое количество дефиниций. Это нормально, учитывая то обстоятельство, что каждый исследователь стремится выделить в исследуемом явлении в первую очередь те характеристики, которые отвечают целям и задачам исследовательской работы. Но, как и везде, в этом есть своя обратная сторона: в многочисленных определениях дискурса становится все сложнее ориентироваться. Мы не будем приводить множество этих определений, дабы не создавать еще одно нагромождение понятий и их трактовок. Вместо этого мы отметим наиболее существенные, на наш взгляд, характеристики дискурса, чтобы сразу прояснить, о чем именно пойдет речь в дальнейшем.

Итак, мы рассматриваем дискурс как коммуникативное действие, законченное или длящееся, происходящее в определенном контексте (историческом, культурном, политическом и т.д.), воспринимаемое и интерпретируемое

адресатом.

Данное определение вовсе не претендует на новизну и оригинальность: в частности, аналогичным образом определяет дискурс Т. ван Дейк (van Dijk, 1998: 192). Здесь мы лишь отметим те характеристики, отличающие дискурс от близких, но не тождественных ему понятий текста (речи) и жанра.

В отличие от текста как законченного речевого произведения, дискурс динамичен. Это не только и даже не столько сам продукт речевой деятельности, сколько взаимодействие, т.е. явление социальное. Проще говоря, дискурс – это не только сама речь (текст), взятая в определенном контексте, но и реакция на нее, ее интерпретация слушающими, читающими, смотрящими – аудиторией в целом и отдельным человеком в частности. Едва ли есть смысл отрицать, что один и тот же текст может по-разному интерпретироваться разными людьми, вызывать у них разные мысли, реакции, чувства. Это объясняется тем, что дискурсы в таких случаях разные, хотя лежащий в их основе текст один и тот же.

Иногда разные дискурсы, связанные тематически или по тем или иным формальным признакам, объединяют в общий вид, когда говорят и пишут о политическом дискурсе, педагогическом

дискурсе, юридическом дискурсе, спортивном дискурсе и т.д. Это сближает дискурс с понятием жанра, однако и между ними есть существенная разница. Жанр как совокупность структурных и содержательных особенностей текста неизменен, в то время как дискурс динамичен. И здесь, как и в случае с текстом, важную роль играет восприятие. К примеру, «Гамлет» является пьесой по форме и трагедией по содержанию независимо от зрителя или читателя, тогда как дискурс (театральный или художественный), как сказано в определении выше, обязательно включает восприятие и интерпретацию текста или театрального представления читателями и/или зрителями, их реакцию.

Восприятие и интерпретация не просто важная часть дискурса, но и в большинстве случаев – его цель. Писатель (если не брать в расчет случаи графомании, когда «пишут в стол») создает текст не просто так, а в расчете на определенную реакцию читателей. Еще ярче это прослеживается, например, в политическом дискурсе, где вообще практически ничего не говорится и не пишется «просто так». Таким образом, восприятие и интерпретация превращаются в инструмент обратной связи: говорящий или пишущий с самого начала принимает в расчет реакцию аудитории, и если он при этом правильно учитывает контекст и использует нужные коммуникативные средства, то добивается нужных ему результатов.

Наличие обратной связи необходимо учитывать при анализе любого дискурса, но прежде всего – институционального, поскольку именно благодаря ей социальные институты получают возможность влиять на общество. С одной стороны, дискурс институтов в значительной мере определяет поведение социальных групп и общества в целом, но с другой стороны, сам институциональный дискурс в известной мере определяется этим поведением, как мы постараемся

показать в дальнейшем.

С формальной стороны институциональный дискурс изучен в достаточной степени. В отличие от персонального дискурса, институциональность подразумевает наличие некоего социального института, который определяет его характер. При этом под определение социального дискурса подпадает как общение между внутри института, так и коммуникация между представителями института с одной стороны и общественности в самом широком смысле с другой стороны. Под общественностью здесь понимаются не только те, кто обращается к тому или иному институту (пациенты в медицинских учреждениях, студенты в образовательных учреждениях, прихожане церкви и т.д.), но и любые сторонние люди, так или иначе вовлеченные в коммуникативное взаимодействие с институтом – например, аудитория СМИ. Представителей института обычно называют агентами, а обращающихся к ним сторонних людей (например, за услугами) – клиентами. (Freed, 2015, Аматов, Свищев, 2023). В отличие от персонального дискурса, в диалоге «агент-клиент» последний воспринимает информацию как официальную, исходящую от компетентного органа. Соответственно, авторитет агента обычно (хотя и не всегда) выше, чем у клиента (van Dijk, 2008). Таким образом, институциональный дискурс «по умолчанию» подразумевает неравенство участников. Разумеется, речь идет о неравенстве позиций, хотя социальное неравенство может также присутствовать в институциональном дискурсе в том или ином виде, явно или неявно.

Социолингвистический подход подразумевает установление определенного соответствия между социальными институтами и преобладающими в их рамках видами дискурса, и на первый взгляд такая классификация вполне оправдана. В

частности, В.И. Карасик выделяет следующие типы современного институционального дискурса: политический, административный, юридический, военный, педагогический, религиозный, мистический, медицинский, деловой, рекламный, спортивный, научный, сценический и медийный (Карасик, 2000а: 24).

Впрочем, у данной классификации, как и у всех ей подобных, сразу обнаруживаются слабые стороны.

Во-первых, трудно назвать данный перечень полным и окончательным по причине изменчивости социальных институтов, возникновения новых и упразднения старых, что отмечает и автор классификации. Таким образом, перед нами скорее эскиз, нежели строгая структура, призванная охватить все существующие типы институционального дискурса.

Во-вторых, что более важно, за таким формализованным подходом ускользает из внимания важная существенная сторона институционального дискурса: характер отношений (а дискурс – это прежде всего отношение) между институтами и обществом. Институты конструируют дискурс и определяют нарративы не только в политике и СМИ, но и в самых различных областях повседневной жизни, влияя на самые разные процессы и отрасли общественной жизни: мобильность населения (Brömmelstroet et al., 2022), планирование будущего (Oomen et al., 2021), развитие городской инфраструктуры и транспорта (Egan, 2021), ресторанный бизнес (Kauhanen et al., 2024) и многое другое.

Сам В.И. Карасик, вводя данную классификацию, определяет основные признаки институционального дискурса следующим образом: «Институциональный дискурс выделяется на основании двух системообразующих признаков: цели и участники общения. Цель политического дискурса – завоевание и удержание власти, педагогического

дискурса – социализация нового члена общества, медицинского дискурса – оказание квалифицированной помощи больному и т.д. Основными участниками институционального дискурса являются представители института (агенты) и люди, обращающиеся к ним (клиенты). Например, это учитель и ученик, врач и пациент, политик и избиратель, священник и прихожанин» (Карасик, 2000б: 11).

Здесь также необходимо отметить и те ограничения, которые накладывает на нас характер исследуемого явления. Будучи социальным действием, т. е. динамичной, а не статичной сущностью, дискурс вообще с трудом поддается регламентации. Чтобы говорить о наличии, отсутствии или нарушении институциональности, требуется некий четкий критерий помимо принадлежности кого-то из участников к тем или иным социальным институтам, а такие критерии в большинстве случаев отсутствуют. Существующие инструкции, регламенты и уставы, как правило, затрагивают лишь формальную сторону речи, но не ее семантику. В частности, уставы вооруженных сил предписывают военнослужащим, как они должны обращаться друг к другу, но при этом никак не регламентируют то содержание информации, которое они передают. Более того, они и не могут регламентировать это содержание, поскольку оно всегда ситуативно и не может быть предсказано заранее. Аналогичная ситуация наблюдается и в других видах дискурса, например, в политическом: смена нарратива вовсе не обязательно подразумевает изменение каких-то формальных признаков общения. К тому же сами эти формальные признаки нередко представляют собой обычные правила речевого этикета, большей частью неписанные. В связи с этим проблема, как нам представляется, носит двоякий характер. С одной стороны, не каждый дискурсивный акт можно однозначно отнести к какому-то одному типу – здесь

дискурс-анализ сталкивается с проблемой, схожей с трудностями определения жанровой принадлежности текста. С другой стороны, дискурс – это взаимодействие людей, которые далеко не всегда, условно говоря, давали обязательство строго следовать протоколу, а потому в анализе даже самых «безобидных» форм дискурса невозможно гарантировать отсутствие разного рода неожиданностей, сбоев, нарушения коммуникации и прочих «подводных камней».

В дальнейшем мы рассмотрим как пример дискурса, плохо поддающегося систематизации в рамках дихотомии «институциональный-персональный», так и случаи намеренного или невольного нарушения принципа институциональности в дискурсе, когда ситуация вышла из-под контроля или же была сознательно направлена по неожиданному и не предусмотренному никакими протоколами руслу.

Цели и задачи

Цель работы – выработка критерия институциональности, позволяющего не только классифицировать и далее исследовать тот или иной вид дискурса, но и помогающего раскрыть тактики манипуляции – в первую очередь, но не только политической. На наш взгляд, данный критерий следует искать не в формальной отнесенности дискурса к тому или иному социальному институту, а в его функциональных особенностях, позволяющих участникам добиваться своих целей, распространять и отстаивать свои интересы. В свою очередь, определение сущностных, а не формальных характеристик институционального дискурса (в том числе путем их противопоставления) является основной задачей данной работы.

Методы и материалы исследования

Помимо стандартных методов классификации и стратификации, в работе применяется метод критического дискурс-анализа (КДА). Данный метод

применяется преимущественно к тем видам дискурса, где явной или неявной целью выступает утверждение, поддержка и оправдание неравенства любого рода (социального, национального, гендерного и т.д.), а также воспроизводство действующей власти (Fowler, 1991; Fairclough, 1995; van Dijk, 2008; Wodak & Meyer, 2009; Catalano & Waugh, 2020; Sherwani, 2021).

Доминирование и подчинение в общественном смысле всегда подразумевает организацию, а в современном обществе – определенный социальный институт. Таким образом, институциональность удобнее всего анализировать с позиций критического дискурс-анализа, выявляя те функции, которые как сама институциональная принадлежность, так и отклонения от нее, выполняют при осуществлении стратегий доминирования, отстаивания и/или навязывания своих интересов.

В качестве фактического материала рассматриваются примеры медийного и политического дискурса, в которых при помощи разного рода манипуляций осуществляется, на первый взгляд, нарушение принципов институциональности, а на деле же манипулятор просто оборачивает институциональность против своего оппонента.

Результаты и их обсуждение

Институциональный дискурс и проблема классификации

Первая проблема, с которой сталкивается исследователь дискурса, в том числе институционального – это систематизация объекта исследования. На первый взгляд, социолингвистический подход позволяет систематизировать виды дискурса без особых проблем: насколько структурировано общество, настолько же структурирован и его дискурс. В.И. Карасик, как уже упоминалось выше, выделяет персональный и институциональный дискурс. Первый, в свою очередь, разделяется на бытовой и

бытийный и в основном выпадает из сферы нашего анализа, тогда как второй выделяется по социальным институтам, агенты которых являются его

репрезентантами: все просто, понятно и укладывается в стройную схему, как на рисунке ниже:

Рисунок 1. Типы дискурса согласно системе В. И. Карасика
Figure 1. Discourse types according to V. I. Karasik



Схема основана на классическом бинарном членении и подразумевает открытость, особенно в правой, наиболее интересующей нас части. Сюда можно добавлять новые типы дискурса, соответствующие новым общественным институтам, или делить уже имеющиеся согласно исследовательским целям и задачам. В целом все выглядит достаточно просто и удобно, а главное – хорошо структурировано.

Однако все, что хорошо структурируется, так же хорошо привлекает внимание критиков. Не стала исключением и предложенная В. И. Карасиком классификация. Критики не обошли вниманием определенную условность такого разделения дискурса на

типы. В частности, Ю.Е. Прохоров справедливо замечает в этой связи: «Действительно, с одной стороны, ряд из них (видов дискурса – авт.) может быть сопоставлен в родовидовом отношении: научный дискурс может быть и юридическим, и медицинским, и педагогическим. Педагогический может быть и научным, и личностно-ориентированным – бытийным (как процесс обучения) и даже бытовым – как процесс воспитания, например, в семье. С другой стороны, даже обиходно-бытовое общение (бытовой дискурс) на практике редко не является статусно-ориентированным, как, впрочем, и дискурс бытийный. И оба они редко не содержат некоторые составляющие научного или,

например, массово-информационного дискурса» (Прохоров, 2006: 196–197).

В процитированном рассуждении мы выделили полужирным шрифтом фрагмент, непосредственно относящийся к смешению институционального и персонального в дискурсе. Действительно: что представляет собой, к примеру, педагогический дискурс (однозначно отнесенный у В.И. Карасика к институциональным видам) и чего в нем на самом деле больше: институциональных, регламентированных характеристик или же персональных, лично ориентированных?

В этой связи критически важным для целей нашего исследования представляется и дальнейшее развитие мысли В.И. Карасика, а именно следующий фрагмент: «В педагогическом дискурсе доля личностного компонента достаточно велика (она различается и в лингвокультурном отношении, например, в российских и американских школах приняты разные режимы общения учителя и ученика, в нашей стране традиционно отношения между школьниками и учителями являются более близкими, чем в США, но, с другой стороны, там менее формализованы отношения между студентами и преподавателями университетов, чем в России)» (Карасик, 2002: 11).

Во-первых, отметим здесь значительную роль «личностного компонента» в институциональном (по определению автора) педагогическом дискурсе. Во-вторых, немаловажным в свете нашей работы представляется и тот момент, что особенности образовательного дискурса могут в значительной степени варьироваться в зависимости от лингвокультурной специфики среды, в частности, русской, британской, американской, французской и других лингвокультур.

В связи с этим возникает проблема институционального статуса педагогического дискурса, степени его

опосредованности теми или иными общественными институтами, прежде всего образовательными учреждениями.

Педагогический дискурс и границы институциональности

В отечественной науке (в частности, педагогической), как и в русскоязычной лингвокультуре в целом, сложился взгляд на воспитание как неотъемлемую часть целостного учебно-воспитательного процесса. Поскольку сам процесс образования обычно рассматривается как протекающий в стенах разного рода образовательных учреждений, то и педагогический дискурс принято считать институциональным. Думается, что именно на этом основании он включен в перечень видов институционального дискурса, предложенный В.И. Карасиком.

Так, в работе Т.М. Тленкопачевой представитель социального института образования рассматривается как основной участник и инициатор педагогического дискурса. **«В ПД в роли адресанта выступают представители института образования. Адресант выступает в качестве дискурсивного экспертного общества, которое представляет собой иерархически организованную группу экспертов в сфере образования, которая владеет определенным набором жанров, посредством которых осуществляются коммуникативные цели»** (Тленкопачева, 2019: 57). Учитель, по мнению автора, выступает основным инициатором речевой ситуации: «Именно в его руках сосредоточена власть в конкретной обстановке. Его можно назвать «держателем речи», адресантом, который создает тексты по причине внутренней необходимости, потребности, желания. **Адресант наделен талантом создавать речь и воплощать ее в жизнь в виде собственного педагогического текста.** При этом целью адресанта является намеренное воздействие на адресата с желанием добиться желаемого в определенной коммуникативной ситуации» (Тленкопачева, 2019: 57–58). Итак, здесь

основными характеристиками адресанта педагогического дискурса названы принадлежность к определенному образовательному институту и особые компетенции – способность «создавать речь и воплощать ее в виде собственного педагогического текста». Все эти особенности, безусловно, подчеркивают институциональный характер педагогического дискурса. Отсюда логически следует вывод о том, что и основные виды педагогического дискурса, дидактический и воспитательный, тоже носят институциональный характер, поскольку являются частями более общего педагогического дискурса.

Основываясь на такой концепции, некоторые авторы начинают делить педагогический дискурс на виды в зависимости от их жанровых особенностей и характеристик: дидактический, академический, публицистический, социализирующий, специальный (Остражкова, 2004: 8).

Если кратко обобщить вышеизложенное, то взаимоотношение педагогического, учебного и воспитательного дискурса представляются следующим образом: педагогический дискурс представляет собой вид институционального дискурса, в котором агентами выступают представители разного рода социальных образовательных институтов, а учебный и воспитательный дискурс можно рассматривать как подвиды педагогического. В русле такого подхода получается, что не только учебный, но и воспитательный дискурс – это дискурс институциональный.

Однако современная педагогика выделяет три основных формы образования: формальную, неформальную и информальную (см., например, Меморандум непрерывного образования Европейского Союза). Из этих трех форм только первую можно считать в полной мере институциональной: формальное образование как раз и осуществляется в рамках официальных учебных заведений.

Соответственно, институциональным будет и дискурс формального образования. В неформальном образовании, которое, согласно тексту меморандума, осуществляется «в образовательных учреждениях или общественных организациях, клубах и кружках, а также во время индивидуальных занятий с репетитором или тренером», институциональная составляющая дискурса уже значительно слабее или вовсе отсутствует – как, например, в занятиях с индивидуальным тренером или репетитором. Наконец, в информальном образовании, под которым понимается «индивидуальная познавательная деятельность, сопровождающая нашу повседневную жизнь и не обязательно носящая целенаправленный характер», об институциональном дискурсе речь вести и вовсе не приходится – такое образование основано исключительно на личностно-ориентированном общении.

Более того, отмечается также отсутствие (а в последнее время – и целенаправленное стирание) четких границ между тремя этими формами образования. В частности, как отмечает Г. М. Нефедова, «именно в последнее время наблюдается движение в сторону оформления интеграции теории и практики указанных форм образовательной деятельности. Формальному образованию намеренно придаются некоторые черты неформального и наоборот, а информальное образование пронизывает как формальное, так и неформальное образование» (Нефедова, 2017: 128). Подобное размывание границ не может не отражаться на образовательном дискурсе, поскольку накладывает отпечаток на сам характер педагогического общения.

Таким образом, в неформальном обучении мы сталкиваемся с прямым нарушением основополагающего формального критерия институциональности – принадлежности ведущего участника дискурса к тому или

иному общественному институту. В самом деле, индивидуальный тренер, репетитор или наставник (например, гувернант) вовсе не обязательно представляют какую-то организацию. Еще меньше оснований усматривать признаки институциональности в дискурсе информального образования, которое по определению осуществляется в повседневной жизни, т.е. как раз в той сфере, где применяется в основном персональный, а не институциональный дискурс.

Педагогический дискурс нередко выходит за рамки институционального, что, в свою очередь, влияет на характер педагогического общения, в том числе выбор языковых средств. Институциональность воспитательного дискурса, к примеру, также не является безусловной в силу характера самого процесса воспитания, осуществляемого не только в рамках учебных заведений, но и за их стенами. Здесь прежде всего необходимо разобраться с отношением включенности данных понятий, а именно, является ли воспитательный дискурс частью педагогического или это пересекающиеся понятия, ни одно из которых не входит полностью в состав другого.

Современный энциклопедический словарь (1997) определяет образование как «целенаправленный процесс обучения и воспитания в интересах личности, общества и государства», который «ведет к овладению ценностями культуры и нравственно-эмоционального отношения к миру, опытом профессиональной и творческой деятельности, сохраняющими и развивающими духовные и материальные достижения человечества» (СЭС: Образование). При этом воспитание в том же словаре определяется как «создание условий для развития и саморазвития человека, освоения им социального опыта, культуры, ценностей и норм общества». Интересен и состав участников процесса воспитания, по мнению составителей

словаря: «В воспитании взаимодействуют личность, группы и коллективы, семья, социальные институты: учебно-воспитательные заведения, государственные общественные организации, церковь; значительное влияние на воспитание оказывают средства массовой коммуникации и др.» (СЭС: Воспитание).

Как видим, «овладение ценностями культуры» и выработка «нравственно-эмоционального отношения к миру» как раз и подразумевают воспитание. При этом в воспитании, по мнению авторов словаря, участвуют далеко не только учебно-воспитательные заведения, но и другие группы и коллективы: семья, церковь, СМИ и прочие. Но семейный дискурс нельзя назвать институциональным: он в первую очередь личностный – бытовой или бытийный. Церковь же, хоть и является социальным институтом, однако «отвечает» за другой вид дискурса, религиозный, который в вышеприведенной классификации представляет собой самостоятельный вид институционального дискурса, отличный от педагогического.

Педагогический энциклопедический словарь (2009), будучи специализированным, профильным изданием, раскрывает понятие образования шире (полужирный шрифт наш): «В О. **объединяются обучение и воспитание**, обеспечивающие готовность личности к выполнению социальных и профессиональных ролей. В О. индивид **осваивает систему ценностей**, знаний, умений и навыков, соответствующих его интересам и **общественным ожиданиям**. ... В образовательном процессе участвуют ученики и воспитанники, преподаватели и воспитатели, образовательные программы, содержание образования, методы и средства обучения, группа (класс), школа, **окружающая среда в целом и, прежде всего, культурная среда**» (ПЭС: Образование).

Авторы этого словаря указывают, что образование объединяет обучение и

воспитание. Таким образом, понятие воспитания оказывается включенным в объем более широкого понятия образования. При этом среди участников процесса образования, помимо социальных институтов, названа «окружающая среда в целом», что явно выводит понятие образовательного дискурса за рамки чисто институционального. Воспитание же в Педагогическом энциклопедическом словаре определяется следующим образом: «ВОСПИТАНИЕ, относительно осмысленное и целенаправленное возвращение человека в соответствии со спецификой целей, групп и организаций, в которых оно осуществляется. В. как относительно социально контролируемая социализация отличается от стихийной и относительно направляемой социализации тем, что в его основе лежит социальное действие, которое М. Вебер определял как направленное решение проблем, сознательно ориентированное на ответное поведение партнеров и предполагающее субъективное осмысление возможных вариантов поведения людей, с которыми человек вступает во взаимодействие. В отличие от социализации – процесса непрерывного, В. – процесс дискретный (прерывный), ибо, будучи планомерным, **осуществляется в определенных организациях**, т.е. ограничено местом и временем» (ПЭС: Воспитание).

В этом определении имеется явное несоответствие с предыдущим тезисом: если образование включает в себя воспитание и осуществляется не только в образовательных учреждениях, но и «окружающей средой в целом и, прежде всего, культурной средой», то почему воспитание ограничивается лишь рамками «определенных организаций»? Может быть, воспитание – это такая часть образования, которая может осуществляться только в определенных институтах, без которых воспитание немислимо? В этой связи попытаемся выяснить, что думают по этому поводу представители педагогической науки.

Известнейший педагог А.С. Макаренко писал: «Воспитывает все: люди, вещи, явления, но прежде всего и больше всего – люди. Из них на первом месте – родители и педагоги» (Макаренко, 2023: 11). Оказывается, воспитывать, по мнению Антона Семеновича, могут не только люди, но и неодушевленные предметы и явления. А главное – среди воспитателей-людей Макаренко выделяет не только педагогов, но и родителей, причем родителей ставит на первое место. То есть мы опять возвращаемся к повторению той, пусть и банальной, но безусловной истины, что воспитание начинается в семье. А семейный дискурс, как говорилось выше, институциональным не является. Более того, если проследить дальше за мыслью А. С. Макаренко, то можно отметить и еще более интересные вещи. «Со всем сложнейшим миром окружающей действительности ребенок входит в бесконечное число отношений, каждое из которых неизменно развивается, переплетается с другими отношениями, усложняется физическим и нравственным ростом самого ребенка.

Весь этот «хаос» не поддается как будто никакому учету, тем не менее он создает в каждый данный момент определенные изменения в личности ребенка. Направить это развитие и руководить им – задача воспитателя.

Бессмысленна и безнадежна попытка некоторых родителей извлечь ребенка из-под влияния жизни и подменить социальное воспитание индивидуальной домашней дрессировкой. Все равно это окончится неудачей: либо ребенок вырвется из домашнего застенка, либо вы воспитаеете уроды» (Макаренко, 2023: 11–12).

Как видим, в отличие от авторов Педагогического энциклопедического словаря, А.С. Макаренко вовсе не считал воспитание дискретным процессом. В его понимании это процесс непрерывный, включающий огромное количество факторов и отношений между ними, на

первый взгляд – «хаос, не поддающийся учету». Другими словами, воспитание – процесс жизненный, который лишь условно, с большими оговорками может быть разделен на какие-то этапы, виды и т.д. Отметим, что и дискурс как коммуникативное действие – тоже в известной степени «хаос, не поддающийся учету». Это жизненный процесс, и по той же причине трудно укладывается в строгую классификацию, исключая какие-то промежуточные компоненты и взаимные пересечения областей. В связи с этим ограничить воспитательный дискурс узкими институциональными рамками означало бы неоправданно сузить область его применения и, как следствие, его научного изучения.

Границы институциональности и их нарушение

Сопоставление дефиниций, исследовательской практики и нормативных документов показывают, что в отношении институциональности педагогический дискурс является одной из проблемных областей исследования с методологической точки зрения. Разумеется, вопрос соответствия классификации и практики дискурс-анализа не ограничивается одним лишь педагогическим дискурсом: при анализе дискурса любого типа могут возникнуть аналогичные бреши (когда непонятно, к какому типу дискурса относится тот или иной дискурсивный акт) или, напротив, наложения (когда один дискурсивный акт подпадает сразу под несколько пунктов классификации). Однако в вышеприведенных рассуждениях о педагогическом дискурсе нам по крайней мере не приходится говорить о сознательном нарушении правил коммуникации: вряд ли педагоги и составители словарей намеренно вводят свою аудиторию в заблуждение. Скорее всего, здесь мы имеем дело лишь со сложностью классификации, вытекающей из своего рода «размытости» самих исследуемых явлений, что является лишь

одной из сторон проблемы. Другой ее стороной является деятельность самих участников дискурса, нарушающая установленные нормы и тем самым затрудняющая классификацию и дальнейший анализ коммуникативного акта.

Как отмечают составители сборника “Institutionality: Studies of Discursive and Material (Re-)ordering”, социальные институты, вопреки распространенному представлению об этих структурах, не являются чем-то стабильным и неизменным. Они постоянно изменяются, организуя и реорганизуя власть на всем земном шаре (Porsché, 2022). Мало того, что меняются сами институты, их отдельные агенты нередко разнообразят дискурс собственными творческими находками, что делает его еще более текучим и, на первый взгляд, вовсе не поддающимся структурированию. Предлагаем в этой связи рассмотреть несколько примеров таких «институциональных сдвигов», которые не только привлекают внимание, но и, что самое главное, позволяют агентам социальных институтов вполне успешно продвигать свои интересы и добиваться целей. В качестве первого такого примера приведем фрагмент интервью от 12 апреля 2023 года, которое известный предприниматель Илон Маск, глава сразу нескольких крупных корпораций, дал корреспонденту Би-Би-Си Джеймсу Клейтону. Как мы постараемся показать в дальнейшем, сама продолжительность фрагмента также имеет довольно важное значение.

Elon Musk (00:00:33):

You see more hate speech personally?

Interviewer (00:00:35):

I would say see more hateful content in that.

Elon Musk (00:00:38):

Content you don't like, or hateful?

Describe a hateful thing.

Interviewer (00:00:42):

Yeah, I mean just content that will

solicit a reaction. Something that may include something that is slightly racist or slightly sexist. Those kinds of things.

Elon Musk (00:00:54):

So you think if it's something that is slightly sexist it should be banned?

Interviewer (00:00:57):

No, I'm not saying anything. I'm saying-

Elon Musk (00:01:00):

I'm just curious. I'm trying to understand what you mean by "hateful content." And I'm asking for specific examples. And you just said that if something is slightly sexist, that's hateful content. Does that mean that it should be banned?

Interviewer (00:01:14):

Well, you've asked me whether my feed, whether it's got less or more. I'd say it's got slightly more.

Elon Musk (00:01:20):

That's why I'm asking for examples. Can you name one example?

Interviewer (00:01:23):

I honestly don't...

Elon Musk (00:01:24):

You can't name a single example?

Interviewer (00:01:28):

I'll tell you why. Because I don't actually use that For You feed anymore. Because I just don't particularly like it. Actually a lot of people are quite similar. I only look at my following.

Elon Musk (00:01:36):

You said you've seen more hateful content, but you can't name a single example. Not even one.

Interviewer (00:01:41):

I'm not sure I've used that feed for the last three or four weeks. And I honestly couldn't-

Elon Musk (00:01:44):

Then how could you see the hateful content?

Interviewer (00:01:46):

Because I've been using it. I've been using Twitter since you've taken it over for the last six months.

Elon Musk (00:01:50):

Then you must have at some point seen the For You hateful content. I'm asking for one example.

Interviewer (00:01:54):

Right.

Elon Musk (00:01:55):

And you can't give a single one.

Interviewer (00:01:56):

And I'm saying-

Elon Musk (00:01:57):

Then I say, sir, that you don't know what you're talking about.

Interviewer (00:02:00):

Really?

Elon Musk (00:02:00):

Yes. Because you can't give a single example of hateful content. Not even one tweet. And yet you claimed that the hateful content was high. That's false.

Interviewer (00:02:12):

No. What I claimed-

Elon Musk (00:02:13):

You just lied.

Здесь сразу обращает на себя внимание смена ролей: вопросы задает И. Маск, а журналист пытается отвечать. При этом И. Маск его постоянно перебивает (на что указывают реплики репортера *I honestly don't...*, *What I claimed-*, *And I'm saying-*) и настойчиво повторяет свой вопрос снова и снова, требуя привести примеры «контента ненависти» в сети Твиттер, о котором в начале интервью упомянул журналист. Конечно, в этой связи возникает вопрос по поводу профессиональной компетенции журналиста, который поднял довольно острую тему, не удосужившись найти хотя бы несколько фактических доказательств своих утверждений. Этим немедленно воспользовался Илон Маск, перейдя в контратаку и полностью перехватив инициативу. Журналисту оставалось лишь отделаться невразумительными общими фразами про «что-то слегка расистское и слегка сексистское» и пытаться сменить тему, чего И. Маск не позволял сделать на протяжении довольно долгого времени.

Однако ответ на вопрос о

профессионализме репортера, каким бы он ни был, не приблизит нас к пониманию, зачем И. Маск сделал то, что сделал. То, что И. Маск с самого начала ломает формат интервью, можно назвать простым хулиганством или «троллингом», как это сейчас называют на сетевых ресурсах, однако не будем делать поспешных выводов.

Во-первых, если это было простой хулиганской выходкой, то она слишком затянулась. Мы привели хоть и довольно большой (2 минуты), но далеко не полный фрагмент интервью, где И. Маск препирается с журналистом, требуя предъявить примеры «контента ненависти» в своей соцсети. Все препирательство миллиардера с журналистом на эту тему, в ходе которого журналист все сильнее сдавал позиции своему визави, длилось более трех минут и закончилось, как нетрудно догадаться, торжеством Илона Маска.

Во-вторых, роли интервьюера и интервьюируемого поменялись не только на одном этом промежутке: до самого конца интервью И. Маск явно давал понять, кто доминирует в данном акте коммуникации, причем делал это открытым текстом. Приведем еще один фрагмент:

Elon Musk (00:03:49):

And what about the fact that the BBC was put under pressure by the British government to change its editorial policy? Are you aware of that?

Interviewer (00:03:56):

This is not an interview about the BBC.

Elon Musk (00:03:59):

Oh, you thought it wasn't?

Interviewer (00:04:02):

I see now why you've done Twitter Spaces. I am not a representative of the BBC's editorial policy. I want to make that clear. Let's talk about something else-

Elon Musk (00:04:08):

I'm interviewing you too.

Здесь Маск не только задает вопросы, но и сам определяет тему разговора, а

также констатирует: «Я тоже беру у вас интервью». Это уже слишком даже для злостного хулиганства.

В-третьих, и это самое главное, в анализе дискурса надо обязательно учитывать контекст, в котором звучат те или иные вопросы и ответы. И этот контекст был таким, что выставлял Илона Маска в очень невыгодном свете. Дело в том, что Маск, приобретя Твиттер, сократил более 80% его сотрудников (из 8000 осталось 1500). Именно об этом сокращении начал было спрашивать журналист, затронув проблемы модерации и упомянув злополучный «контент ненависти», за который тут же ухватился Маск. А поняв, что журналист допустил оплошность (если, конечно, все интервью с самого начала не было срежиссированной акцией) и конкретных примеров приведено не будет, Маск сразу захватил «бразды правления» в свои руки и не выпускал их до конца интервью.

Что характерно, данное интервью (а его в первый же день просмотрело более 3 миллионов человек) большинству запомнилось именно остроумными и нешаблонными репликами Маска, который «ставит зарвавшегося журналиста-манипулятора на место». Во всяком случае, содержание большинства комментариев (мы просмотрели более тысячи) именно таково. Некоторые комментаторы также отмечают непрофессионализм репортера, но никто не задается основным вопросом – массовым сокращением персонала. Во всяком случае, нам этот вопрос в комментариях не встретился ни разу. Возможно, что «неудобные» вопросы просто удаляются модераторами, но это является только лишним подтверждением того факта, что доминирующие социальные группы и институты неуклонно блюдают свои интересы, моделируя и структурируя дискурс в том числе и такими незатейливыми методами. Нам же остается лишь констатировать: данный институциональный сдвиг, смена социальных ролей или коммуникативный

сбой – можно называть как угодно – Илону Маску полностью удался, достигнув всех целей. Миллиардер не только сохранил репутацию, избежав неприятных ассоциаций, но и зарекомендовал себя остроумным, эрудированным и нешаблонно мыслящим человеком самых широких взглядов. И, что немаловажно, поборником справедливости.

И еще одно возможное возражение в связи с нашим анализом этого интервью: что если институциональность тут ни при чем, Маск просто говорил как частное лицо, выражая лишь свою точку зрения, а никак не корпоративные интересы? На этот вопрос мы даем однозначно отрицательный ответ сразу по ряду причин. Во-первых, Илон Маск, как уже говорилось, является владельцем и генеральным директором сразу нескольких очень крупных корпораций (Tesla, SpaceX, Twitter) и маловероятно, чтобы он хоть на минуту забыл об их (то есть о своих) интересах, особенно перед телекамерами. Во-вторых, вопросы в интервью были вовсе не о том, какие блюда Маск предпочитает есть на обед или какие сувениры покупает в командировках – они все были о политике и/или крупном бизнесе (а другим бизнесом Илон Маск, насколько нам известно, не занимается). В-третьих, при всей своей экстраординарности беседа Маска с репортером получилась максимально обезличенной в том смысле, что в ней выдвигаются претензии и упреки не к отдельным лицам, а к корпорациям, будь то Твиттер, Тесла или Би-Би-Си. Говорить почти исключительно о корпорациях, не подразумевая корпоративных интересов, было бы крайне странно. Таким образом, институциональность в данном интервью присутствует как прямо, так и скрыто, несмотря на кажущийся «парадокс»: интервью в целом звучит как нарушение институциональности вплоть до полного ее отрицания. Однако отрицать, на наш взгляд, следует не институциональность, а лишь формально-структурный подход к ее

описанию и объяснению.

Рассмотрим еще один дискурсивный акт тоже в формате интервью, которое состоялось буквально через несколько дней после рассмотренного выше, а именно 21 апреля 2023 года. Речь пойдет об интервью Лю Шая, посла КНР во Франции, французскому телеканалу LCI (L'interview). Цитаты из интервью были размещены на странице канала в сети Twitter, откуда довольно быстро разошлись по другим сетевым ресурсам. Но если интервью Илона Маска на канале BBC всего лишь «взорвало интернет», то с интервью посла КНР французскому телеканалу все зашло значительно дальше. На прозвучавшие в этом интервью высказывания отреагировали не только блогеры и журналисты, его не обошли вниманием и официальные государственные институты. Их задело то, что посол назвал страны бывшего СССР «не имеющими эффективного статуса в международном праве». Сразу 80 депутатов Европарламента направили министру иностранных дел Франции К. Колонна письмо, требуя объявить посла КНР персоной нон грата и называя его поведение неприемлемым и «оскорбляющим историю, культуру и целостность государств» (ТАСС, 2023). Кроме того, министерства иностранных дел нескольких бывших советских республик вызвали послов КНР «для объяснений» – по всей видимости, оскорбившись тем, что им отказали в «эффективном международном статусе» (Коммерсантъ, 2023).

Мы не будем подробно вдаваться в политический и социальный контекст данного интервью, сделав ссылку на одну из наших предыдущих работ, где все это рассмотрено более подробно (Аматов, Свищев, 2023). Тем не менее, ряд моментов следует отметить как для прояснения общей ситуации, так и в связи с главным объектом нашего внимания – институциональностью.

Во-первых, как уже упоминалось

выше, здесь как раз тот случай, где смешиваются две разновидности институционального дискурса по формальной, социолингвистической классификации: с одной стороны, это политический дискурс, с другой – медийный. Это обстоятельство в принципе могло бы и не иметь решающего значения, если бы принципы институциональности, как и в предыдущем примере, не были нарушены и журналист просто делал бы свою работу, а дипломат – свою. Однако ситуация развивалась иначе. Поэтому важно учесть, что посол и журналист изначально находятся в разных условиях, причем посол более уязвим в силу большей ответственности и более жестких рамок, которые накладывает на него дипломатический протокол.

Во-вторых, что касается контекста, то здесь необходимо учесть визит президента Франции Э. Макрона в Китай, состоявшийся незадолго до данного интервью и, по-видимому, не достигший ожидаемых целей. Именно с обсуждения этого визита дискуссия начинает приобретать напряженность, которая в конечном итоге вылилась в слова, за которые посла КНР требовали призвать к ответу. Забегая вперед, надо также отметить, что скандал не привел к отставке посла – во всяком случае, на июнь 2024 г. (т.е. спустя более чем год после событий) Лю Шай продолжал исполнять свои обязанности. Это говорит о том, что руководство внешнеполитического ведомства Китая не усмотрело «крамолы» в его интервью, в отличие от целого ряда своих европейских коллег.

Итак, посол прокомментировал визит президента Макрона в Китай и сделанные им заявления как «полностью отвечающие интересам Франции». Этот дипломатичный ответ, очевидно, не соответствовал ожиданиям репортера, который немедленно начал накалять обстановку. В разговоре происходит «скачок напряжения», что хорошо заметно по следующей реплике журналиста:

On voit à quel point cette déclaration a provoqué la colère de beaucoup en Occident qui considèrent qu'on a trop concédé à une dictature qu'est la Chine. Le petit clip que présente la république lui aussi est conçu de retour de Chine a beaucoup marqué en ce sens-là. Est-ce que vous entendez cette inquiétude des Occidentaux qui se disent on a fait là trop de concessions?

Если в предыдущем рассмотренном нами интервью Илон Маск, сменив характер дискурса и взяв на себя роль репортера, все же оставался в рамках логики и здравого смысла, то здесь мы видим набор стандартных демагогических приемов, примеры которых выделены шрифтом. Реплика начинается с подмены причины известного следствия: вспышка «общественного гнева», который переживала в те дни Франция, была связана с непопулярной пенсионной реформой, которая волновала французов гораздо больше, чем довольно ровные заявления их президента в Китае, да и акции протеста начались за несколько месяцев до визита Макрона в Китай. Следом идет пропагандистское клише про «китайскую диктатуру» и прямая ложь о том, что президент Франции якобы сделал уступки этой «диктатуре». Далее следует отсылка к общественному мнению жителей Запада, которого, разумеется, никто не спрашивал. И завершает данный пассаж еще одно упоминание об «уступках», которые никто не делал – чтобы в этом убедиться, достаточно прочитать сами заявления Макрона. Мы не будем приводить их здесь, однако найти их не представляет никакой сложности.

Дипломат на первых порах реагирует достаточно спокойно: отметив, что «вообще-то Китай не диктатура», он предлагает перейти к обсуждению насущных политических проблем. Журналист, тем не менее, продолжает давить:

La citation du général de Gaulle: Quand il parle de l'excellence du fait que vous ayez, quand il alerte tellement au datif à

vosre égard pour l'année 1964. Mao Zedong est-il toujours là? C'est un des plus grands criminels de l'histoire. Mao Zedong a fait déporter, torturer, assassiner tant de gens, des millions de gens. Est-ce que vous comprenez que les sentiments de beaucoup d'Occidentaux, que, à l'époque de général de Gaulle se trompaient. Et à l'époque, qu'aujourd'hui ceux qui veulent traiter avec la Chine peuvent se tromper.

Если в предыдущем примере «слом шаблонов» был инициативой интервьюируемого, то здесь тем же самым занимается уже сам репортер, который фактически перестает быть репортером. Он не задает вопросы с целью получить ответы, а высказывает и навязывает собственное (собственное ли?) мнение в достаточно безапелляционной форме. Снова сославшись на «общественное мнение» (точнее даже «чувства многих жителей Запада»), репортер допускает уже прямые оскорбления страны, которую представляет посол. Мао Цзэдун назван «одним из крупнейших преступников в истории», на котором «насильственные депортации, пытки и убийства миллионов человек». Вину может признать только суд, но об этом принципе права (кстати, западного) журналист почему-то забыл, поставив клеймо преступника на наиболее уважаемого в Китае руководителя. Видимо, «чувства жителей Запада», которых неизвестно каким образом касаются вымышленные преступления Мао Цзэдуна, оказываются важнее права. Очевидно, что логика и исторические факты в этот момент интересовали журналиста меньше всего. То, чем он занимался, было неприкрытой провокацией. В какой-то мере эта провокация удалась, что видно и по реакции посла, который сначала допустил некоторое снижение регистра речи (в допустимых пределах, разумеется), назвав эти домыслы «пустой болтовней» (*racontards*), а затем и те высказывания, которые вызвали бурю на Западе, не возымев в то же время никакого заметного

эффекта на Востоке. Видимо, в китайском МИД поняли, что данное интервью было провокацией. Впрочем, догадаться об этом несложно с учетом тех приемов, которые применял журналист и которые мы только что рассмотрели.

Функциональный взгляд на институциональность дискурса

Провокационные интервью, где участники (будь то журналисты или интервьюируемые) нарушают не только правила речевого этикета и/или принципы профессиональной этики, но и институциональные основы дискурса, довольно распространены и появились не вчера. Мы лишь привели наиболее резонансные из сравнительно недавних. В чем причина и цель подобных провокаций? В случае с интервью, рассмотренными выше, ответ, как ни странно, лежит на поверхности: власть. Илон Маск сразу дал понять «зарвавшемуся» журналисту, кто на самом деле представляет власть в этой студии. И, что немаловажно, добился успеха, не только сохранив имидж, но и укрепив его при помощи нестандартных коммуникативных ходов. В другом примере журналист, напротив, при помощи несложных манипуляций «поставил на место» посла, несмотря на то, что формально социальный статус дипломата гораздо выше, чем у журналиста. Стало быть, формальный статус коммуникантов (как и формальный подход к дискурсу в целом) не является определяющим при выявлении институциональных характеристик дискурса. Тогда что является?

Древние римляне рекомендовали в любой непонятной ситуации искать ответ на вопрос “*Cui prodest?*” И, как нам представляется, для любого исследователя дискурса, в особенности институционального, данный вопрос должен стать чем-то вроде девиза. Институциональный дискурс очень изменчив, может принимать разные формы, менять содержание, даже вполне допускает противоречивые тезисы вплоть

до взаимоисключающих (вспомним известный мем «Вы не понимаете, это другое»). Но есть ли в нем хоть что-то, остающееся неизменным во все времена и во всех обстоятельствах? То, от чего может (и должен, раз уж на то пошло) отталкиваться всякий добросовестный исследователь? Такая сущность есть, и это власть со своими интересами, главный из которых – оставаться властью. Соответственно, дискурс-анализ целесообразно строить не на формальных основаниях принадлежности участников дискурса к тому или иному социальному институту, а на содержательной стороне сообщений, а именно на том, чьи интересы они выражают и отстаивают. В этой связи, например, представляет интерес освещения одного и того же события разными СМИ, за которыми стоят разные политические институты (Yu, Zheng, 2022), которые и являются бенефициарами от продвижения определенной повестки. Также всевозможные случаи оговорок и в особенности намеренных подтасовок и прочих манипулятивных приемов позволяют четче рассмотреть тех, «кому выгодно».

Какие черты отличают функциональный подход к исследованию дискурса от формального? Прежде всего, сосредоточенность на содержательной, а не на формальной стороне дискурса. Скажем, в обсуждении какой-то проблемы на телевидении принимают участие журналист, политик, педагог, священник и представитель правоохранительных органов. Формальный подход, помимо трудностей с классификацией (медийный, политический, педагогический, религиозный, юридический – какой перед нами тип дискурса?), оказывается еще и малопродуктивным. Если мы, прибегая к натяжкам, даже сумеем втиснуть данный дискурсивный акт в тот или иной тип, это не приведет нас ни к каким конкретным выводам. Чтобы сделать анализ дискурса продуктивным, нам необходимо обратиться к анализу самих высказываний

и реакции на них, в том числе и оставив в стороне институциональную принадлежность участников, если потребуется.

На наш взгляд, успешный анализ дискурса должен включать (но не обязан ограничиваться ими) следующий ряд действий:

- анализ контекста (культурного, исторического, политического, экономического и т.д.);
- лингвистический анализ текста (в особенности темы и основной идеи);
- анализ открытых источников (буквально все, что можно найти в связи с данным дискурсивным актом);
- анализ обратной связи (лингвистический, психологический, социокультурный);
- анализ последствий (процедура, схожая с анализом рисков в экономике, но имеющая более широкое применение).

Институциональность, как нетрудно заметить, может проявляться (и чаще всего будет себя проявлять) на всех этапах такого анализа. Исследование контекста позволяет раскрыть, какие институты представлены (или могут быть опосредованно представлены) в том или ином дискурсивном акте. Лингвистический анализ текста дает представление о том, с каким типом текста мы имеем дело, а значит, и какой институт стоит или может стоять за теми или иными высказываниями. Анализ открытых источников позволяет получить необходимую информацию об участниках дискурса и их вовлеченности в другие дискурсивные акты. Анализ обратной связи позволяет определить отношение аудитории к тому или иному социальному институту. Наконец, анализ последствий дает возможность раскрыть бенефициаров данного коммуникативного действия, причем даже в случаях, когда цели не были ими достигнуты. При этом следует понимать, что ни один из данных шагов не представляет собой целостного дискурс-анализа, поскольку акцентирован лишь на

одной из сторон исследуемого явления. Однако последовательное применение этих методов может дать и, скорее всего, даст максимально объективную картину.

Разумеется, такой подход потребует от исследователя и более широкого кругозора в целом, и специальных знаний, и учета большего количества деталей при анализе, но ведь и результат того стоит. В конце концов, дискурс-анализ и не должен быть легким занятием, поскольку дискурс не ограничивается рамками какой-то одной дисциплины (лингвистики, социологии, политологии и т.д.), а требует учета массы факторов и применения методологической базы разных дисциплин.

Заключение

Как было показано выше, формально-структурный подход к дискурсу, несмотря на свои внешние стройность и изящество, далеко не всегда служит надежным инструментом анализа. Здесь перед исследователем встают проблемы двоякого свойства. С одной стороны, многие виды деятельности (и соответствующие им виды дискурса, который тоже деятельность) протекают как в стенах официальных институтов, так и вне их. В таком случае, как было показано на примере педагогического дискурса, определить их институциональный статус бывает довольно затруднительно. В самом деле, нельзя же всерьез полагать, что воспитательный дискурс институционален в стенах учебного заведения, но сразу становится бытовым, когда участники выходят за порог школы и перестают быть агентами и клиентами. Нет, он все также выступают в качестве воспитателей и воспитуемых, могут даже употреблять те же самые слова и речевые обороты. Что же касается эффективности, можно спорить, что воспитывает больше, семья или школа, но сам этот спор уже указывает на неоднозначность и нерешенность проблемы даже на теоретическом уровне.

Другая и, на наш взгляд, более важная особенность институционального дискурса заключается в его высокой

лабильности, способности менять форму и содержание, подстраиваясь под складывающиеся тем или иным образом обстоятельства. Такая текучесть может стать настоящим кошмаром для исследователя, потратившего массу времени и сил на выявление существенных формальных характеристик институционального дискурса только чтобы узнать, что все изменилось и выявленные им признаки уже неактуальны.

Выход из данного методологического тупика нам видится в функциональном дискурс-анализе. Дискурс, особенно институциональный, представляет собой систему, а один из путей познания системы лежит как раз через познание ее функций. Этот путь тем более актуален, если изучаемая система, как дискурс, саморазвивается, теряя одни формальные признаки и приобретая другие. Образно говоря, гусеница превращается в бабочку, однако остается одним и тем же видом насекомого, преследующим в том числе при помощи метаморфоз все ту же цель сохранения себя как вида. Аналогичным образом вопросы доминирования и подчинения никогда не теряют актуальности, если речь идет об институциональном дискурсе. Участники институционального дискурса могут представлять разные социальные институты или даже формально не представлять ни один из них, но при этом отстаивание интересов той или иной организации, социальной группы, института в широком смысле слова будет указывать на то, что дискурс институционален.

Список литературы

- Аматов А. М., Свищев Г. В. «Четвертая власть» пятой республики: дискурс на грани институциональности. / А. М. Аматов // Текст. Язык. Человек: сб. науч. тр. / УО МГПУ им. И. П. Шамякина; редкол.: С. Б. Кураш (отв. ред.) [и др.]. Мозырь: МГПУ им. И. П. Шамякина, 2023. С. 187–193.
- Карасик В. И. О типах дискурса /

В.И. Карасик // Языковая личность: институциональный и персональный дискурс: Сб. науч. тр. Волгоград: Перемена, 2000а. С. 5–20.

Карасик В. И. Языковой круг: личность, концепты, дискурс / В. И. Карасик. Волгоград: Перемена, 2002. 477 с.

Карасик В. И. Этнокультурные типы институционального дискурса / В. И. Карасик // Этнокультурная специфика речевой деятельности. М., 2000б. С. 19–31.

Макаренко А. С. Книга для родителей / А. С. Макаренко. М: АСТ, 2019. 502 с.

Нефедова Г. М. Специфика интеграции формального, неформального и информального образования / Г. М. Нефедова // Научно-педагогическое обозрение. Pedagogical Review, 2017 (3), Томск. С. 127–132.

Остражкова Н. С. Обучение пониманию содержательного компонента педагогического дискурса лекции : Английский язык, языковой вуз : автореферат дис. ... кандидата педагогических наук : 13.00.02 / Н. С. Остражкова / Тамб. гос. ун-т им. Г. Р. Державина. Тамбов, 2004. 18 с.

Прохоров Ю. Е. Действительность. Текст. Дискурс / Ю. Е. Прохоров. М.: Флинта: Наука, 2006. 224 с.

Тленкопачева М. Н. Особенности речевого поведения участников англоязычных обучающих вебинаров: дисс. ... канд. филол. наук: 10.02.04. / М. Н. Тленкопачева. – Пятигорск, 2019. – 209 с.

Brömmelstroet M., Mladenović M. N., Nikolaeva A., Gaziulusoy I., Ferreira A., Schmidt-Thome K., Ritvos R., Sousa S., Bergsma B. Identifying, Nurturing and Empowering Alternative Mobility Narratives // Journal of Urban Mobility. Volume 2. December 2022, 100031. DOI: 10.1016/j.urbmob.2022.100031

Catalano T., Waugh L. R. Critical Discourse Analysis, Critical Discourse Studies and Beyond. Springer Nature Switzerland AG, 2020. 406 p. DOI: 10.1007/978-3-030-49379-0

Dijk T. A. van. Discourse and power. NY: Palgrave Macmillan, 2008. 308 p.

Dijk T. A. van. Ideology: A Multidisciplinary Approach, London: Sage, 1998. 384 p.

Egan R. Provoking responsibility: The struggle for recognition as an everyday cyclist in Dublin City // Geoforum 127, 2021. Pp. 23–32. DOI: 10.1016/j.geoforum.2021.09.012

Fairclough N. I. Critical Discourse Analysis: The Critical Study of Language, Harlow, UK: Longman, 1995. 265 p.

Fowler R. Language in the News. Discourse and Ideology in the Press. London: Routledge & Kegan Paul, 1991. 254 p.

Freed A.F. Institutional Discourse // The International Encyclopedia of Language & Social Interaction, Karen Tracy, Cornelia Ilie, & Todd Sandel, (eds.). Boston: John Wiley & Sons, 2015. Pp. 809–826. DOI: 10.1002/9781118611463.wbielsi151

Kauhanen S., Koponen J., Komppula R. The external business environment faced by the restaurant industry: Discourse analysis of trade journal editorials in Finland in 1979-2023 // European Journal of Tourism Research. 2024. Vol. 38 3807. <https://doi.org/10.54055/ejtr.v38i.3408>

Oomen J., Hoffman J., Hajer M. A. Techniques of futuring: On how imagined futures become socially performative // European Journal of Social Theory. 2022. Vol. 25(2). Pp. 252–270. DOI: 10.1177/1368431020988826

Porsché Y., Scholz R., Singh J. N. Introducing Institutionalality // Institutionalality – Studies of Discursive and Material (Re-)ordering. Basingstoke: Palgrave Macmillan, 2022. Pp. 1–28. https://doi.org/10.1007/978-3-030-96969-1_1

Sherwani K. A. Global Review On The Past And Future Of Critical Discourse Analysis/Studies // Webology. 2021. Vol. 18, № 1. Pp. 624–648.

Wodak R., Meyer M. (Eds.) (2009). Methods of Critical Discourse Analysis. SAGE Publications, 204 p.

Yu X., Zheng H. A Critical Discourse Analysis of Different News Reports on the Same Event: Illustrated with Examples from *China Daily* and *The Guardian* // Open Journal of Social Sciences. 2022. № 10. Pp. 348–363. DOI:10.4236/jss.2022.1011023

Список использованных источников

КоммерсантЪ. Латвия, Литва и Эстония вызвали послов Китая из-за слов дипломата о «сомнительном суверенитете» постсоветских стран. 23.04.2023 URL: <https://www.kommersant.ru/doc/5952260> (дата обращения: 01.09.2024)

Педагогический энциклопедический словарь (ПЭС) / гл. ред. Б. М. Бим-Бад. 3-е изд., стер. М.: Большая российская

энциклопедия, 2009. 527 с.

Современный энциклопедический словарь (СЭС). М: Большая российская энциклопедия, 2017. 600 с.

ТАСС. Евродепутаты призвали Францию объявить посла КНР персоной нон грата за высказывания. 23.04.2023. URL: <https://tass.ru/mezhdunarodnaya-panorama/17589301> (дата обращения: 01.09.2024)

Full Elon Musk BBC Interview with Video and Timestamps 12th April 2023. URL: <https://www.youtube.com/watch?v=IflfP4XwzAI> (дата обращения: 01.09.2024)

Hornby, A. S., Cowie, A. P., Gimson, A. C. (eds.) (2007) *Oxford advanced learner's dictionary of current English*, Oxford Univ. Press, Oxford, UK. (In English)

L'interview polémique de Lu Shaye, ambassadeur de Chine en France. URL: <https://www.youtube.com/watch?v=8XYDYflgm> (дата обращения: 01.09.2024)

References

Amatov, A. M., Svishchev, G. V. (2023). Chetvertaya vlast pyatoi respublik: diskurs na grani institutsionalnosti [The Fourth Power of the Fifth Republic: Discourse on the Verge of Institutional Status], in *Tekst. Iazyk. Chelovek*, MGPU, Mozyr', Belarus, 187–193. (In Russian)

Karasik, V. I. (2000a). O tipakh diskursa, in *Yazykovaya lichnost: institutsionalny i personalny diskurs* [On Types of Discourse], Peremena, Volgograd, Russia, 5–20. (In Russian)

Karasik, V. I. (2002). Yazykovoy krug: lichnost, kontsepty, diskurs [Language Circle: Personality, Concepts, Discourse], Peremena, Volgograd, Russia. (In Russian)

Karasik, V. I. (2000b). Etnokulturnye tipy institutsionalnogo diskursa, in *Ethnocultural specificity of speech activity*, Moscow, Russia, 19–31. (In Russian)

Makarenko, A. S. (2019). Kniga dlya roditelei [The Book for Parents], AST, Moscow, Russia. (In Russian)

Nefedova, G. M. (2017). Spetsifika integratsii formalnogo, neformalnogo i informalnogo obrazovaniya, *Nauchno-pedagogicheskoe obozrenie. Pedagogical Review*, 2017 (3), Tomsk, Russia, 127–132. (In Russian)

Ostrazhkova, N. S. (2004). Teaching Understanding of the Essence of Pedagogical Discourse in a Lecture: English College Course,

Abstract of Ph.D. dissertation, Tambov, Russia. (In Russian)

Prokhorov, Yu. E. (2006). Deistvitelnost. Tekst. Diskurs [Reality. Text. Discourse], Flinta: Nauka, Moscow, Russia. (In Russian)

Tlenkopacheva, M. N. (2019). Osobnosti rechevogo povedeniya uchastnikov angloyazychnykh obuchayushchikh vebinarov [Specifics of Communication in English-language webinars], Pyatigorsk, Russia. (In Russian)

Brömmelstroet, M., Mladenović, M. N., Nikolaeva, A., Gaziulusoy, I., Ferreira, A., Schmidt-Thome, K., Ritvos, R., Sousa, S., Bergsma, B. (2022). Identifying, Nurturing and Empowering Alternative Mobility Narratives, *Journal of Urban Mobility*, 2, 100031. (In English) DOI: 10.1016/j.urbmob.2022.100031

Catalano, T., Waugh, L. R. (2020). *Critical Discourse Analysis, Critical Discourse Studies and Beyond*, Springer Nature, Switzerland AG. (In English) DOI: 10.1007/978-3-030-49379-0

Dijk, T. A. van. (2008). *Discourse and power*, Palgrave Macmillan, NY, USA. (In English)

Dijk, T. A. van. (1998). *Ideology: A Multidisciplinary Approach*, London: Sage. (In English)

Egan, R. (2021). Provoking responsibility: The struggle for recognition as an everyday cyclist in Dublin City, *Geoforum* 127, 23–32. (In English) DOI: 10.1016/j.geoforum.2021.09.012

Fairclough, N. I. (1995). *Critical Discourse Analysis: The Critical Study of Language*, Longman, Harlow, UK. (In English)

Fowler, R. (1991). *Language in the News. Discourse and Ideology in the Press*, Routledge & Kegan Paul, London, UK. (In English)

Freed, A. F. (2015). Institutional Discourse, in Karen Tracy, Cornelia Ilie, & Todd Sandel (eds.), *The International Encyclopedia of Language & Social Interaction*, John Wiley & Sons, Boston, USA, 809–826. (In English) DOI: 10.1002/9781118611463.wbielsi151

Kauhanen, S., Koponen, J., Komppula, R. (2024). The external business environment faced by the restaurant industry: Discourse analysis of trade journal editorials in Finland in 1979–2023, *European Journal of Tourism Research*, 38, 3807. (In English) <https://doi.org/10.54055/ejtr.v38i.3408>

Oomen, J., Hoffman, J., Hajer, M. A. (2021). Techniques of futuring: On how imagined futures become socially performative, *European*

Journal of Social Theory, 25(2), 252–270. (In English) DOI: 10.1177/1368431020988826

Porsché, Y., Scholz, R., Singh, J. N. (2022). Introducing Institutionalality, in Porsché, Y., Scholz, R., Singh, J. N. (eds.) *Institutionality – Studies of Discursive and Material (Re-)ordering*, Palgrave Macmillan, Basingstoke, 1–28. (In English) https://doi.org/10.1007/978-3-030-96969-1_1

Sherwani, K. A. (2021). Global Review On The Past And Future Of Critical Discourse Analysis/Studies, *Webology*, 18, 1, 624–648. (In English)

Wodak, R., Meyer, M. (eds.) (2009). *Methods of Critical Discourse Analysis*, SAGE Publications. (In English)

Yu, X., Zheng, H. (2022). A Critical Discourse Analysis of Different News Reports on the Same Event: Illustrated with Examples from *China Daily* and *The Guardian*, *Open Journal of Social Sciences*, 10, 348–363. (In English) DOI:10.4236/jss.2022.1011023

Corpus materials

Kommersant. “Latviya, Litva i Estoniya vyzvali poslov Kitaya iz-za slov diplomata o «somnitelnom suverenitete» postsovetsskikh stran.” 23.04.2023. URL: <https://www.kommersant.ru/doc/5952260> (Accessed 01 September 2024). (In Russian)

Pedagogicheskiy entsiklopedicheskiy slovar (2009) *Bolshaya rossiiskaya entsiklopediya*, Moscow, Russia. (In Russian)

Sovremenniy entsiklopedicheskiy slovar (2017) *Bolshaya rossiiskaya entsiklopediya*, Moscow, Russia. (In Russian)

TASS. Evrodeputaty prizvali Frantsiyu obyavit posla KNR personoy non grata za vyskazyvaniya. 23.04.2023. URL: <https://tass.ru/mezhdunarodnaya-panorama/17589301> (Accessed 01 September 2024). (In Russian)

Full Elon Musk BBC Interview with Video and Timestamps 12th April 2023. URL: <https://www.youtube.com/watch?v=If1fP4XwzAI> (Accessed 01 September 2024). (In English)

Hornby, A. S., Cowie, A. P., Gimson, A. C. (eds.) (2007) *Oxford advanced learner's*

dictionary of current English, Oxford Univ. Press, Oxford, UK. (In English)

L'interview polémique de Lu Shaye, ambassadeur de Chine en France. URL: <https://www.youtube.com/watch?v=8XYDYf1gmtA> (Accessed 01 September 2024). (In French)

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.

Александр Михайлович Амагов, доктор филологических наук, профессор, ФГАОУ ВО «Белгородский государственный национальный исследовательский университет», Белгород, Россия.

Alexander M. Amатов, D.Sc. (Philology), Professor, Belgorod State National Research University, Belgorod, Russia.

Сергей Бонифацевич Кураш, доктор филологических наук, профессор, УО «Мозырский государственный педагогический университет имени И. П. Шамякина», Мозырь, Беларусь.

Sergey B. Kurash, D.Sc. (Philology), Professor, I. P. Shamyakin Mozyr State Pedagogical University, Mozyr, Belarus.

Елена Сергеевна Данилова, кандидат филологических наук, доцент, ФГАОУ ВО «Белгородский государственный национальный исследовательский университет», Белгород, Россия.

Elena S. Danilova, Cand. Sc. in Philology, Associate Professor, Belgorod State National Research University, Belgorod, Russia.

UDC 81`373.2

DOI: 10.18413/2313-8912-2024-10-3-0-2

Maria N. Dubinina¹ 

**Numeric Domain Names and National Domains:
Trends in Internet Onomastics**

¹ HSE University,
11 Pokrovsky Blvd., Moscow, 109028, Russia
E-mail: mdubinina@hse.ru
ORCID: 0009-0006-3477-5543

Received 2 January 2024; accepted 15 September 2024; published 30 September 2024

Abstract: The research paper provides an analysis of the Internet domain naming system, in particular, a study on numeric domain names and national domains. In the Internet space, various signs and symbols from different semiotic systems can be used. One interesting finding is that the use of hieroglyphic, Cyrillic and numeric domain names in the Internet space is becoming a distinct trend. Hieroglyphic and numeric domain names, in particular, can play an important role in addressing the issue of Internet onomastics. From this perspective, Chinese domain names have provided a natural bridge point for Internet technology, Chinese traditional culture, modern and traditional industries. The research material consists of domain names of Internet websites. The aim of our study is to describe the development strategy of hieroglyphic and numeric domain names from a linguistic point of view. The relevance of the study is determined by the limited exploration of onomastics in virtual realities. Domain names serve as identifiers and means of communication. It is essential to keep in mind that the domain naming is a dynamic phenomenon, and trends in domain naming can change over time. Additionally, the value of a domain name is often subjective and depends on factors like brandability, memorability and market demand. The research identifies trends in Internet onomastics of domain names, that can be exemplified by the use of hieroglyphs and numerical codes in Chinese internet space. Special attention is given to domains where the creative function of language is evident in their creation.

Keywords: Internet discourse; Internet communication; Internet linguistics; Internet onomastics; Onomastics; Domain name; Hieroglyphic domain; Numeric domain; Chinese

How to cite: Dubinina, M. N. (2024). Numeric Domain Names and National Domains: Trends in Internet Onomastics, *Research Result. Theoretical and Applied Linguistics*, 10 (3), 26-41. DOI: 10.18413/2313-8912-2024-10-3-0-2

УДК 81`373.2

DOI: 10.18413/2313-8912-2024-10-3-0-2

Дубинина М. Н.¹ 

**Цифровые и национальные домены:
тенденции ономастики в интернет-коммуникации**

¹ Национальный исследовательский университет «Высшая школа экономики»
Покровский бульвар, д. 11, Москва, 109028, Россия
E-mail: mdubinina@hse.ru
ORCID: 0009-0006-3477-5543

Статья поступила 2 января 2024 г.; принята 15 сентября 2024 г.;
опубликована 30 сентября 2024 г.

Аннотация. Подобно тому, как в реальном физическом пространстве присутствуют названия зданий, вывески и адреса, уличная реклама и дорожные знаки, виртуальное пространство наполнено названиями веб-сайтов и доменными именами, образующими виртуальный лингвистический ландшафт. В данной исследовательской работе представлен анализ системы именования доменов в Интернете, в частности исследование числовых доменных имен и национальных доменов. В интернет-пространстве могут использоваться знаки и символы из разных семиотических систем, что делает эту область особенно интересной для современных лингвистических исследований. Более того, использование иероглифических и цифровых доменных имен в китайском интернет-пространстве становится определённым трендом, отражая культурные и технологические особенности региона. С этой точки зрения китайские доменные имена стали естественным связующим звеном между интернет-технологиями, китайской традиционной культурой, современными и традиционными производствами. Материалом исследования являются доменные имена интернет-сайтов. Целью нашего исследования выступило описание стратегии развития иероглифических и числовых доменных имен. Актуальность исследования определена малой изученностью ономастики виртуальных реалий. Доменные имена служат идентификаторами и средством коммуникации. Важно помнить, что доменное имя – это динамичное явление, а тенденции в области именования доменов могут со временем меняться. Кроме того, ценность доменного имени часто субъективна и зависит от таких факторов, как узнаваемость бренда, запоминаемость и рыночный спрос. В ходе исследования выявлены тенденции интернет-ономастики доменных имен, в частности, использование иероглифов и числовых кодов. Особое внимание уделяется доменным именам, в создании которых проявляется творческая функция языка.

Ключевые слова: Интернет-дискурс; Интернет-коммуникация; Интернет-лингвистика; Интернет-ономастика; Ономастика; Доменное имя; Иероглифический домен; Цифровой домен; Китайский язык

Информация для цитирования: Дубинина М. Н. Цифровые и национальные домены: тенденции ономастики в интернет-коммуникации // Научный результат. Вопросы теоретической и прикладной лингвистики. 2024. Т. 10. № 3. С. 26-41. DOI: 10.18413/2313-8912-2024-10-3-0-2

Introduction

The Internet space, inherently dominated by the organization of digits, has in recent years become a battleground for Latin, Cyrillic, and ideographic scripts. Since the birth of the Internet, due to objective reasons, the development of information products has followed a path of anglicization, and for a long time, domain names were represented solely by Latin letters and sometimes by numbers. However, the situation is changing.

Much of this change is associated with the development of alternative IT giants in countries where English is not the native language, such as Korea and China. Furthermore, multilingualism on the Internet ensures that language does not become a barrier to accessing the benefits of humanity. Indeed, the desire of nations to assert their sovereignty in virtual space becomes increasingly evident. D.S. Barinova notes that “the level of development of the national domain zone and how it is used by the state,

business, and society can serve as an indicator of the development and effectiveness of the state” (Barinova, 2010, 308). Thus, over time, Cyrillic and ideographic characters appear in the domain system, and the use of different systems on the Internet emphasizes the multipolarity of political and social attitudes. On the other hand, the desire of countries to use their national language signifies an aspiration to emphasize informational sovereignty. Moreover, there exist a number of organizations and language activism movement, such as a long-standing research project The IDN World Report¹ that believe that international domain names contribute to linguistic diversity online. Their aim is to enable users to navigate the Internet in their native language, promote multilingualism and universal access to cyberspace.

This trajectory of domain name development and the states’ determination to defend their sovereignty and interests in the Internet space highlight the relevance of researching this direction, so as O. V. Dedova and M. S. Kuprienko state, the name takes on the role of a beacon, a navigator in the online world, a landmark in electronic information space (Dedova, Kuprienko, 2013: 63). Moreover, the phenomenon of creativity in creating domain names remains a relatively understudied area of linguistics, particularly domain names registered in non-Latin scripts, known as internationalized domain names. For example, as Yu. E. Polyak carried research on internationalized Cyrillic domains, he mentioned several reasons why new domain names with native language characters can be popular. The researcher pointed out that when a new domain zone emerges, many ‘tasty’ names appear that were previously occupied – names of goods, proper names, geographical concepts etc. In addition, using native language makes it possible accurately and unambiguously indicate objects (Polyak, 2018: 116). Indeed, up to

now, far too little attention has been paid to domain names in general and to certain peculiarities in such countries as China, where giant technology companies and vast Internet market give insights into the key global trends. Thus, this paper attempts to contribute to Internet onomastics research as well as to address the issue of hieroglyphic and numeric domain names.

Methods and materials of the study

The use of national languages is mostly discussed relating to the localization of websites, where navigation sections and site content are adapted to the national-cultural specificity of users. However, researchers have recently addressed issues of the status and naming of domain names, as well as linguistic creativity and innovations in domain naming. Many of these works have served as a theoretical basis for our research, that is to say, study on phenomena of Internet communication (Khazova, 2023); research on main trends, problems and selection mechanisms of domain names (Zhiltsov, 2020); study on the development of different domain systems, including works on cyberspace management systems (specifically related to domain names and IP addresses) (Kalyatin, 2002; Zakalkin, 2022); study on the development of onomastic space of the Internet (Kersten, Lotze, 2022; Suprun, 2004); examination of the names of websites as a special phenomenon in Internet onomastics (Zubareva, 2021b); research on domain names as identifiers and means of communication (Rozhkova, 2015); study on domain name and name of the file as innovative forms of text names produced by Internet technology (Dedova, Kuprienko, 2013); analysis of linguostylistic means for creating domain names (Kuprienko, 2010); research on domain name as a unit for constructing and identifying the network landscape of the online space, a unit which forms its structure and allows some of its parts to function as microsystems, creating entire ecosystems (Ryabchenko, 2016). Furthermore, domain spaces have started to be studied as separate communities.

¹ Internationalized Domain Name World Report. URL: <https://www.idnworldreport.eu/> (Accessed 23 December 2023).

M. Z. Abdullaeva's work investigates the spiritual role of Islamic media space of domain 'UZ' (Abdullaeva, 2020); L. N. Dukhanina and A. A. Maksimenko conduct an analysis of the network oncological discourse of a three-year retrospective in the top-level domain 'RU' (Dukhanina, Maksimenko, 2020). These articles prove the fact that while conducting the research in virtual space it is reasonable to take into account geolocation, analysis of sources and distribution platforms.

In our view, there is a need to outline another research direction, namely, the study of ideographic and numeric domains as onomastic units of the Internet space. **This prospective study was designed** to investigate current trends in naming numeric and ideographic domains, and to consider linguistic means that are used to make domain name unique and memorable. To achieve this goal, several tasks need to be addressed. Firstly, consider the role of the domain name in creating a sovereign Internet. Secondly, examine the domain as an object of linguistic research. Thirdly, identify the characteristics and trends in the formation of numeric and Chinese-character domains as the brightest examples. The fundamentally new aspect is the consideration of ideographic and numeric domains as unique objects of network naming, as in some cases, domain names represent a fundamentally new phenomenon in onomastics, rather than simply duplicating offline onyms. **The practical significance** of the study lies in the fact that the obtained data can be in demand in the analysis of the linguistic development of Internet space and the development of marketing strategies.

The material for the study is based on the findings of theoretical research, providing an overview of Internet onyms and the development of onomastics in Internet discourse. Additionally, it is important to note that the material selection is not limited to publications in the Russian language. We made an attempt to analyze materials in English and Chinese.

As for **linguistic material**, more than 50 domain names were selected. Most of the examples are given from the Chinese Internet zone, though certain trends are exemplified by a few examples from other domain zones. The choice is motivated by the increasing number of numeric and character domains in the Chinese internet space and their growing significance. The employed **research methods** involve analyzing and synthesizing theoretical and practical material, classifying examples chosen for analysis, and describing the obtained data. The selected domain names were categorized based on specific criteria to facilitate analysis and interpretation. The data obtained from the analysis were described to provide a comprehensive understanding of the observed trends and patterns in the selected domain names. The inclusion of multiple languages in the analysis reflects the global nature of Internet communication and the growing significance of such domains in the Chinese Internet space.

Domain as a toponym in the virtual linguistic landscape

The study of domain names reveals the need to understand the mechanisms of Internet onomastics. Similar to how physical space is surrounded by buildings, signs, and addresses, virtual space is surrounded by website names, nicknames, domain names. What and how we see largely depends on browser design, our settings and permissions, including cookies, content, and website design.

Toponymy of virtual landscape is a relatively new field of study. According to N. A. Ryabchenko, "domain is a part of the online space, forming its structure and at the same time allowing parts of this space to function as a microsystem, creating entire ecosystems in which users are provided with the necessary functioning capabilities and access to information" (Ryabchenko, 2016: 99). Thus, domain names are a kind of toponym in the virtual Internet space, participating in the construction of virtual reality. As V. I. Suprun writes, all information on the Internet is built on an onomastic axis

and proper names permeate all texts. Onyms to be studied are the names of functioning parts and units within the Network (cit. ex: Virtual onyms in the Italian Internet environment², 2017: 18).

We share L. P. Son's opinion that "electronic onomastics can become one of its (Internet communication culture) 'building blocks', as the harmony of form and content (the concept of the sender and recipient) of an electronic proper name is culture itself" (Son, 2009: 221). Within onomastic studies, questions have been raised about the names of participants in virtual communication or anthroponyms, ergonyms, and pragmatonyms in Internet communication (Zubareva, 2021b; Son, 2009; Son, 2012). However, when it comes to anthroponyms in Internet communication, researchers often conclude that "participants wear masks" (Son, 2009: 218). On the other hand, onomastics of domain names implies different principles of functioning. As noted by A. V. Zubareva, proper names must possess such qualities as uniqueness and attractiveness (Zubareva, 2021a: 101). Meanwhile, the domain name identifies the owner, the goods. Many argue that it serves as a means of communication, a means of information exchange. So, M. Rozhkova concludes that domain names in addition to the technical function of addressing on the Internet and virtual identification began to contribute to the identification of goods, works, services of some producers, sellers and performers among similar goods, works and services "in reality" (Rozhkova, 2015: 62).

Virtual space is an environment for constructing a virtual identity. As A. V. Zubareva notes, the onomasticon of Internet communication is formed in accordance with the logic of the interaction between online (Internet environment) and

offline (the familiar "real" world) (Zubareva, 2021b: 122). We cannot but admit that partly domain naming trends are taken from physical reality. For example, in the Russian Internet space, domain names with automobile region codes corresponding to Russian regions can be found, such as 66.ru (Yekaterinburg platform), 161.ru (Rostov-on-Don platform). In addition, the site 66.мвд.рф belongs to the Main Directorate of the Ministry of Internal Affairs of Russia for the Sverdlovsk Region. To decode why exactly these numbers are used the Internet user is supposed to know automobile codes of Russian regions. Furthermore, domain names 8848.ru (The Vertical Film Festival) and 8848-altitude.ru (a ski clothing brand) refer to Mount Everest's peak. These examples confirm the opinion that the involvement of cultural data allows us to reveal the content and meaning of onyms (including Internet onyms) relying on the interaction of proper names and the culture of the people (Madieva, Suprun, 2017: 38).

We conclude that a domain name is a toponym in virtual space. They are grouped "zonally", either geographically/regionally (ru, 中国) or used by certain communities and organizations (com, net, info). When classifying domain names, they are often categorized based on their affiliation with a particular zone, national identity, or a field of use. Because of this, there is a need to understand new onomastic phenomena and the mechanisms through which they arose.

Domain name as a part of sovereign Internet

In order to understand why national and numeric domains are of great interest, the section below provides a brief overview of domain naming as an issue of concern for national security. Today, the Internet space becomes not only a zone of interaction but also a platform for asserting interests, propaganda, and lobbying. P. V. Zakalkin emphasizes that the control of cyberspace is largely monopolized by the United States, with other countries being mere consumers and thus susceptible to destructive influence.

² Давыдова Н.Н. Виртуальные онимы в итальянской интернет-среде. Выпускная квалификационная работа. Санкт-Петербургский государственный университет, 2017. URL: https://dspace.spbu.ru/bitstream/11701/8345/1/Davydova_VKR.pdf (Accessed 15 June 2023).

Therefore, for countries like Russia and China, aiming to assert their positions globally, defending their interests in the Internet environment is crucial (Ignatov, 2022; Gabdrakhmanova, Mahmutov, 2018; Shirin, 2014). Language plays a key role in this, with Cyrillic and ideographic scripts trying to find their place in Internet communication (Zakalkin, 2022).

Currently, there are national domains worldwide. According to a report (Internet Tendencies³, 2021), domain RU ranks 8th and domain CN ranks 2nd among top-level domains. In Russia there is even a special project for developing an ecosystem for supporting domain names and postal addresses in national languages, primarily in the Cyrillic domain “РФ”⁴. The development of the global domain space is closely linked to the creation of domains in national languages (IDN TLDs)⁵. According to a report (Internet Tendencies, 2021)³, there are a total of 156 IDN top-level domains worldwide, with the highest number using Chinese characters (56), followed by domains using the Arabic alphabet (33), and Cyrillic domains (17). The most demanded IDN top-level domain is .中國 with 1.8 million domain names. China has contributed a lot into development of Chinese-character domains. The Ministry of Industry and Information Technology has issued the “The 14th Five-Year Plan for the Development of the Information and Communication Industry”⁶, that purposes to

improve Chinese domain name application environment and to further promote the development and application of Chinese domain names.

S. Couture and S. Topin, discussing sovereignty, note that Russia and China insist on implementing state sovereignty on the Internet. They suggest that the development of Cyrillic and ideographic domain names is an attempt to realize digital sovereignty and resist the hegemony of the Latin and English-speaking world (Couture, Topin, 2020). However, the main restraining factor for the development of Cyrillic and Chinese-character domains is the low level of universal acceptance of such domain names and, above all, difficulties with using email addresses written in national alphabet characters.

Domain name as an object of linguistic analysis

In addition to its potential for political influence, a domain name holds commercial value, and linguistic creativity plays a crucial role in domain naming. It is already universally acknowledged that “domain names used specifically for commercial purposes become a generic identifier of goods (works, services) – like a trademark, an indication of the place of production/origin of products, a universal barcode, etc” (Rozhkova, 2015: 62). A domain name has a complex structure, including several levels: top-level, second-level, and subdomains. Concerning top-level domains, A. V. Zubareva mentions the possibility of playing with first-level domain names. For example, using .NET can create a provocative and humorous effect and even influence offline onomastics. The domain .ME (Montenegro) has undergone commercialization and is used for personal blogs and socially oriented sites. In addition, domain .AI (Anguilla) has become extremely popular in 2023.⁷ Undoubtedly, the reason is

³ Тенденции развития интернета: готовность экономики и общества к функционированию в цифровой среде: аналитический доклад / Г. И. Абдрахманова, М. Д. Ванюшина, К. О. Вишнеvский, Л. М. Гохберг и др.; АНО «Координационный центр национального домена сети Интернет»; Нац. исслед. ун-т «Высшая школа экономики». М.: НИУ ВШЭ, 2021. 248 с.

⁴ Поддерживаю.РФ URL: <https://xn--80adfafgo7bio2n.xn--p1ai/> (Accessed 23 August 2023).

⁵ Internationalized Domain Names. URL: <https://en.unesco.org/internationalized-domain-names> (Accessed 08 September 2023).

⁶ 《“十四五”信息通信行业发展规划》解读 URL: https://www.miit.gov.cn/zwgk/zcjd/art/2021/art_8f8e9

251b74940979e1f422203ceb6af.html (Accessed 24 December 2023).

⁷ Домен .AI вырос более чем на 50 тысяч регистраций за три месяца. URL:

the growing popularity of artificial intelligence and the continuous discussions of this topic in the media and social networks. Moreover, numeric domain names can also be combined with specific domain extensions (e.g., .com, .net, .org) to create unique combinations.

A. V. Kuznetsov suggests that some domain names can be considered a type of advertising text. Analyzing the linguistic features of German domain names, the researcher has revealed language play types such as the formation of neologisms, violation of language norms, use of colloquial word forms, substitution of letter combinations with numbers, stylization to foreign words, and the use of phonetic spelling (Kuznetsov, 2012). Using these means of language games helps to create novel and memorable domain names. Similar trends can be observed in the context of domain naming in other countries. Historically, domain names were created using Latin letters and numbers. Creating a domain name is limited by certain rules, such as the absence of spaces. On the other hand, a unique and user-friendly domain name needs to be created to distinguish a website from others and make it stand out for users and search engines. Thus, D. A. Zhiltsov claims that naming should correspond to the chosen general marketing strategy and be based on basic naming principles developed by Internet marketers, webmasters, site optimizers, and other digital marketing specialists over the years. The author provides several recommendations, including not to register Cyrillic domains (Zhiltsov, 2020). However, Chinese researchers convinced that native language domain names are not only easy to remember, but also easily promote and disseminate information, have a prominent native language image, which demonstrates cultural confidence. They can also create a safer and more trustworthy Internet environment for Internet users and improve

international cyberspace governance on the Internet (Ke, 2020: 166).

O. V. Dedova and M. S. Kuprienko emphasize that “on the modern Internet, the domain name is a key component of hypertext navigation and composition.” Their research reports that site traffic directly depends on the memorability and reproducibility of the domain name. Indeed, domain name becomes a mnemonic tool, reflecting the specifics of the information posted on the site, and also an element of advertising. In view of this, the researchers believe that the main strategy for increasing the effectiveness of a domain name is determined by its relevance with meaningful linguistic units, which can be achieved in various ways” (Dedova, Kuprienko, 2013: 65). On the other hand, A. V. Zubareva notes that Internet environment incorporates traditional onyms, and this process can involve the simple transfer of existing onyms into Internet communication and the creation of new proper names for virtualized objects. She emphasizes the significant influence of programming practices on domain naming trends, such as the absence of spaces, the use of underscores, and the use of uppercase letters.

Furthermore, M. S. Kuprienko mentions that “a modern domain name is intended to be not only an element of navigation, but also a mnemonic device reflecting the specifics of the information posted on the site, and an element of advertising (the site itself, the goods, services offered” (Kuprienko, 2010: 80). The author points out several ways how domain name may be formed: the domain name matches the site title or part of it, the website theme is taken as the domain name, the domain name is taken to be a word that is a hyponym in relation to the nomination site header, the domain name is formed by combining fragments of the site header, the domain was created by truncating the site name or one of its keywords, abbreviation, disabbreviation. There are also a lot of means of language game that help to make the domain name unique, i.e. the use of Latin

<https://cctld.ru/media/news/industry/35408/> (Accessed 26 December 2023).

letters that match the Russian ones, playing on the homonymy of the final part of the domain name (the so-called top-level domain, whose name is written after the dot), replacing a word or part of a word with a letter or number (Kuprienko, 2010).

Meanwhile, M. Yu. Karpenko introduces a new term siteonym and clarifies what exactly should be classified as siteonyms. According to the researcher, firstly, these are names, that users find in the address bar (for example, www.google.com), secondly, the title, that users see on the main page (Karpenko, 2012). Furthermore, Karpenko investigates the motives of nomination of websites, mainly domain names, and classifies them into essential, qualitative, locative, temporal, possessive, patronymic, memorial, ideological, symbolic, associative, nominal, situational domain names. According to the research, siteonyms with essential motivation have a significant advantage. This may be due to the fact that the site creators, in order to increase the number of potential visitors, already start with the domain name and site name, and advertise its content or functions (Karpenko, 2014). Subsequently, the study of siteonyms was continued in the works that mention siteonyms as a type of ideonyms (Chaykisova, 2017), that examine cultural and linguistic aspects of German siteonyms (Daldinova, Bardaeva, 2021; Daldinova, Buraeva, 2023; Filimonenko, 2023). These articles contribute a lot to the research of linguistic point of view. Though we believe that it is not right to examine domain names and titles of the sites as similar objects. The reason is that different means and strategies are employed to create these onyms, for example, there are certain technical restrictions, i.e. titles of the site can use different graphical styles and colors, while structure, length, use of special characters and punctuation marks in domain naming have some peculiarities. Moreover, when national top-level domain in Russia was discussed, it was decided to choose. РФ but not “py”. The reason is that domain “py” written with Cyrillic letters resembles domain

of Paraguay (“py”) written with Latin letters. Thus, instead there is Cyrillic domain “pyc”.

In summary, linguistic analysis of domain names reveals the creativity involved in their nomination, influenced by both commercial and linguistic considerations. The use of language play, adherence to marketing strategies and cultural aspects, contribute to the linguistic landscape of domain names.

Linguistic analysis of numeric and Chinese-character domain names

The analysis of Chinese data confirms one of the obvious peculiarities of the Chinese Internet space – the appeal to characters and numbers. And if character domains are a phenomenon of the last few years, then digital domains have appeared a long time ago. This is due to several reasons. Firstly, the figures are very close and understandable to the Chinese. As Zhu Guiwen notes the Chinese are accustomed to operating with numbers and using numbers to create onyms. Even e-mails most often consist of numbers rather than letters. (Zhu, 2014: 43). Secondly, the Chinese put a certain meaning into the numbers (Akhrenova, Dubinina, 2022; Han, 2022). Moving on to the consideration of numeric domains, in the contemporary Internet space in China, numbers continue to play a crucial role, especially in digital onomastics. Currently, around 9 million domain names consist solely of digits, accounting for over 3% of the total number of domains. Here are some findings from the research conducted by Domains Index on digital domains⁸: there is an increasing number of digital domains, the quantity of digital domains in new domain zones surpasses that in old ones; the highest number of digital domains is found in the COM zone, among national zones China (.cn) leads in the number of digital domains; popular domains typically consist of 3-4 digits, with the longest domains having 63 digits. As for numeric domain name’s structure, there can be found

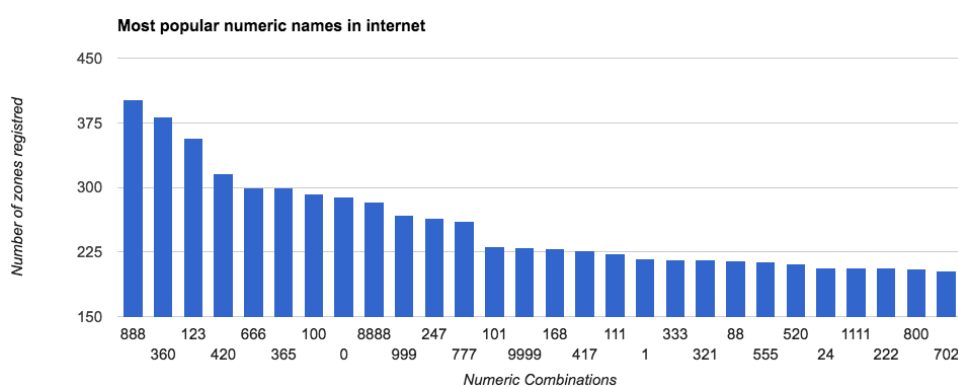
⁸ Nine Millions of Domain Names Are “Just Numbers” URL: <https://domains-index.com/nine-millions-domain-names-just-numbers/#> (Accessed 05 February 2023)

repeating numbers, that are often considered memorable; sequential numbers, that are popular for their simplicity; lucky numbers, that is numeric combinations that are considered lucky in various cultures, such as 888 in Chinese culture; premium numbers, that are considered premium, and businesses may be willing to pay a higher price for them; year-based numbers. Numbers can also be combined with letters.

The numbers in Internet communication can be considered as a kind of code, and decoding them is possible only with the information based on the speaker's culture and language, as they have national and linguistic identity (Dubinina, 2022: 26). Furthermore, An Zhiwei claims that digits have sociocultural meaning, and in China this effect is mostly based on homonymy (An, 2002).

Figure 1. The chart of most popular numeric domain names

Рисунок 1. Таблица самых популярных числовых доменных имен



Meanwhile, similar trend of popularizing numeric domains can be found not only in China. According to the chart of most popular numeric names⁸ (Figure 1), the most popular combination is 888 and it is registered in 402 domain zones. It turns out, that in different cultures it has different meaning quite often positive, including “eternity”, “limitless potential”, “power”, “success”, “prosperity” etc. As for China, the number 8 has gained success, because it is homophonic with “发”, that means to make a fortune. For example, a digital lottery trading platform uses 8791.com domain, that encourages users to get rich (加油发财). The second most popular combination 360 conveys a message of “harmony”, “integrity” etc. For example, domain 360.cn is used for security software pointing out the importance of complex safety, domain 360.yandex.ru is used for a set of services for solving every day and work problems.

Researcher Xiang Xuehua from the University of Illinois focuses on studying the linguistic and cultural characteristics of the domain names of the top fifty most-visited websites in the US and China. The researcher uses several criteria such as brevity of a domain, sound and rhythm, recourse to semiotic systems, intertextuality. She comes to the conclusion that Chinese domain names are heterogeneous, mixing semiotic systems and featuring more cultural symbolism (particularly number symbolism) and literary intertextuality (Xiang, 2012). Indeed, in China numbers not only represent mathematical concepts but also carry symbolic meanings and hold magical significance. And the abundance of homophones has led to the development of imagery and symbolism in thinking and speech. Chinese people don't like some digits, such as 4, because it is a homophone for “death”. What is more, there are some regional peculiarities. In the north, 2 is a curse and means stupid. Some people also suggest

that the best numbers are 5 and 9, because 9 is the most perfect number among single digits, 9 is the highest, and 5 is in the middle. In addition, in China number 6 is a very auspicious number, which means good luck and success.⁹ For example, 6.cn, also known as 六间房 (6 rooms), is one of the online video entertainment live broadcast platforms, providing 24-hour uninterrupted video live broadcast service. It is believed that from the perspective of traditional Chinese culture, “6 rooms” is the most auspicious combination which means good luck and happiness.¹⁰ Moreover, 6 is also homophonic with 流 which means a stream, thus pointing to specialization of the platform.

Using digital code in China is a peculiar feature of domain naming. On the other hand, though numeric virtual onomastics is rather widespread, it is not always possible to unambiguously interpret these combinations. There are some mostly common phrases that can be relatively easily identified: 51=我要, 520=我爱你, 518=我要发 etc. (Fu, 2013). However, homonymy allows to interpret numbers in different ways, and many works that examine digits and combination of digits as part of Internet user’s communication admit that digits can correlate with different words (Zhou, 2003). For instance, a leather goods store 哇咕哇咕 is located at the website 51wagu.com, where 51 can be deciphered as 我要 (I want). The same code is used in creating 51tt.com.cn (淘陶网) for a real estate and home furnishing website, 51dhw.com.cn (51订货网) for electronic business platform, a phone number and domain name 4008-517-517.com for

McDonald’s take away and delivery, and 517 can be deciphered as 我要吃 (I want to eat). Furthermore, the domain name 51job.com for platform that provides individuals with job recruitment information and enterprises with a full range of human resources services can be deciphered as “I want a job”. On the other hand, the platform is also known as 前程无忧 (a worry-free future), and 51 can be found homophonous with 无忧 (worry-free).

Let us introduce some more examples. A store for children and mothers 麦乐购 is located at the website M6go.com, where the number 6 is homophonous with the ideograph 乐. The domain name 2797.com for ticketing platform can be deciphered as 爱去就去¹¹ (if you want to go, then go). The online store 1药网 is located at the website 111.com. The use of the number 1 in the domain name may be related to the name where the digit one is also used, and the original name of the online store was 1号药店. On the other hand, the digit 1 may have the reading yāo, which is homophonous with the reading of the word medicine 药 (yào), also found in the name, creating certain associations with medicine for users. The other example is domain 91.cn, that was created for professional health information portal. It is homophonous with the phrase 就医, which means to receive medical treatment. Another well-known example is the Alibaba Wholesaler Network at the website 1688.com. The numbers 1688 are homophonous with the company’s name 阿里巴巴 (ā lǐ bā bā)

The Chinese domain name also refers to a domain names that contain Chinese characters. We share the viewpoint that the use of Chinese characters produces a form of “localised” Internet, rather than a separate Internet (Arsène, 2015). Examples include domain names like 新华网.cn (XinhuaNet.cn) containing Chinese characters, as well as

⁹ 阿飞: 对纯数字域名的一点看法 URL: <https://www.xinnet.com> (Accessed 15 November 2023)

¹⁰ 六间房 URL: <https://wenku.baidu.com/view/cedbddd4ef3a87c24028915f804d2b160a4e8679.html?wkts=1704179893909&bdQuery=%E5%85%AD%E9%97%B4%E6%88%BF%E4%B8%BA%E4%BB%80%E4%B9%88+%E7%94%A8%E5%85%AD&needWelcomeRecommand=1> (Accessed 28 December 2023)

¹¹ 四数字域名只知道“1688”, 还有这些网站, 你都用过吗? URL:

<https://baijiahao.baidu.com/s?id=1675871967586057745&wfr=spider&for=pc> (Accessed 23 August 2023)

internationalized country and region top-level domain names containing Chinese characters such as .中国 (.China) and .广东 (.Guangdong). It also includes generic Chinese domain names with suffixes such as .手机 (.mobile). In 1998, the China Internet Network Information Center (CNNIC) started the technical development of Chinese domain names, marking the integration of Chinese characters into the international Internet domain name system. In 2009, .中国 (.China) was successfully approved by the Internet Corporation for Assigned Names and Numbers (ICANN), becoming a global generic top-level domain. Since 2013, Chinese top-level domain names such as .我爱你 (.iloveyou), .手机 (.mobile), .集团 (.group), .网络 (.network), .信息" (.information), .网址 (.website), .在线 (.online), .公司 (company), .游戏(.game) and many others have successively appeared on the Internet stage. Here are some examples of domain names using Chinese characters: 中国科学院.网址, 华为.网址, 南博会.网址, 中国国际信息通信展览会.网址, 成都大运会.网址, 中央电视台.中国, QQ.中国, 腾讯.中国, 中国移动.中国.

Top-level domains are represented by both simplified and traditional characters (中國), specific punctuation marks can be used.¹² Moreover, the use of Latin letters restrict the recognition of the brand in China. Researcher Li Jingyi acknowledges the inconvenience of language use on the Internet for non-English speaking countries, citing the example of two companies whose domain names could be confused. They are 内蒙伊利集团股份有限公司(www.yili.net) and 伊丽服饰有限公司 www.yili.com (Li, 2010). Thus, Chinese-character domain names can use entirely Chinese names for

company websites, adapting to the language habits of Chinese people, meeting the needs to promote brands and quickly access information, and helping businesses better participate in Internet competition.

Conclusion

Since the birth of the Internet English has been the dominant language. However, the increasing involvement of users from different countries and nations is transforming this space. Localization of websites, games and adaptation to different phenomena are clear indicators of this evolution. Some researchers believe that multilingualism is a necessary and natural phenomenon for the Internet's evolution, emphasizing the importance of accommodating various languages. Meanwhile, Chinese experts argue that relying solely on the English language hinders the development of the Internet. That is to say, with each passing year, the Internet space continues to expand and become an integral part of virtual reality, increasingly shaping the linguistic expression of the online environment. Thus, the comprehensive exploration of current trends in naming numeric and ideographic domains, taking into account linguistic and cultural factors, is of outmost importance. The other reason is that domain names act as door numbers on the Internet.

Numeric and Chinese-character nomination serves as a part of the virtual linguistic worldview of users. The conclusion drawn is that domain names play a crucial role in shaping the virtual linguistic landscape of the Internet and are intricately linked to the cultural and linguistic preferences of users. For instance, homophonic coding demonstrates that the sound aspect of the form of a virtual name is of fundamental importance. Companies may use numeric domain names as part of their branding strategy, especially if the numbers have significance to the business. Creative language use, as demonstrated by linguistic material, showcases the inventive utilization of language in the digital realm. Moreover, the popularization of Chinese domain names

¹²

工业和信息化部关于调整中国互联网域名体系的公告 URL:
https://www.miit.gov.cn/jgsj/xgj/wjfb/art/2020/art_7b446feb028b4607a158b12530cd2b3e.html (Accessed 12 December 2023)

is beneficial to the promotion of Chinese culture on the Internet, contributing to the construction of a culturally strong and network-strong country. Thus, analyzing ideographic domains contributes to refining our understanding of the role of language in maintaining sovereignty.

References

- Abdullaeva, M. Z. (2020). Spiritual role of Islamic media space in domain “uz”, *Bulletin Social-Economic and Humanitarian Research*, 8 (10), 55-62. <https://doi.org/10.5281/zenodo.4117558> (In English)
- Akhrenova, N. A., Dubinina, M. N. (2022). The role of numbers in the Chinese cyberculture, *Topical Issues of Linguistics and Teaching Methods in Business and Professional Communication – TILTM 2022: Proceedings of the X International Research Conference Topical Issues of Linguistics and Teaching Methods in Business and Professional Communication*, 16–23. <https://doi.org/10.15405/epes.22104.3> (In English)
- An, Zh. (2002). Wǎngluò shùzì xiéyīn cíyǔ qiǎn lùn [A brief discussion on homophonic words for digital numbers on the Internet]. *Journal of Taiyuan University*. 3 (4), 36–38. (In Chinese)
- Arsène, S. (2015). Internet Domain Names in China: Articulating Local Control with Global Connectivity. *China Perspectives*, 2015 (4), 25–34. <https://doi.org/10.4000/chinaperspectives.6846> (In English)
- Barinova, D. S. (2010). National Internet domains – symbols of state borders and limitless possibilities, *MGIMO Review of International Relations*, 5 (14), 307–314. (In Russian)
- Chaikisova, A. V. (2017). Specificity of ideonyms as a special type of proper names, *Current problems of German studies, novelistic studies and Russian studies*, 3, 23–24. (In Russian)
- Couture, S., Toupin, S. (2020). What does the concept of “sovereignty” mean in the digital world?, *Bulletin of international organizations: education, science, new economy*, 15, 4, 48–69. <https://doi.org/10.17323/1996-7845-2020-04-03> (In Russian)
- Daldinova, E. O. G., Bardaeva, D. V. (2021). Characteristics of German-language siteonyms, *Integration processes in modern science: collection of scientific papers based on the materials of the XIX International Scientific-Practical Conference*, Anapa, 10–13. (In Russian)
- Daldinova, E. O. G., Buraeva, T. B. (2023). The characteristic structure of German speaking siteonym, *Bulletin of Issyk Kul University*, 54–2, 65–73. (In Russian)
- Dedova, O. V., Kuprienko, M. S. (2013). Heading complex in electronic communication, *Moscow University Bulletin. Series 9. Philology*, 1, 61–70. (In Russian)
- Dubinina, M. N. (2022). Digital codes of Internet communication (using the example of English, Russian and Chinese languages), *Current problems of German-Roman philology and methods of teaching foreign languages: Collection of scientific articles*, 23–26. (In Russian)
- Dukhanina, L. N., Maksimenko, A. A. (2020). Analysis of network oncology discourse of a three-year retrospective in the top-level domain RU, *VOPROSY ONKOLOGII*, 66 (4), 315–324. <https://doi.org/10.37469/0507-3758-2020-66-4-315-324> (In Russian)
- Filimonenko, E. N. (2023). Cultural and linguistic aspects of German-language siteonyms, *Research Center “Vector of Development”*, 19, 185–190. (In Russian)
- Fu, Y. (2013). Yīnghàn shùzì xiéyīn cí yítóng de shēncéng jiēxī [An in-depth analysis of the similarities and differences between English and Chinese numerical homophonic words]. *Language Application Research*, 9, 135–137. (In Chinese)
- Gabdrakhmanova, G. F., Mahmutov, Z. A. (2018). National Internet of Russia: problem statement revisited, *Oriental Studies*. Vol. 11, № 3(37), 142–151. <https://doi.org/10.22162/2619-0990-2018-37-3-142-151> (In Russian)
- Han, Yu. (2022). Cultural understanding of homonymy of numerals in modern Chinese, *Modern Oriental Studies*, 4 (2), 30–39. <https://doi.org/10.24412/2686-9675-2-2022-30-39> (In Russian)
- Ignatov A. (2022). The BRICS agenda on the Internet governance, *international organisations research journal*, Vol. 17, № 2, 86–109. <https://doi.org/10.17323/1996-7845-2022-02-04>
- Kalyatin, V. O. (2002). Domain names, Moscow: Information and Publishing Center of Rospatent “INIC”, 188. (In Russian)
- Karpenko, M. Yu. (2012). Identification of the basic unit of the Internet onomasticon, *Mova*, 17, 113–115. (In Ukrainian)

- Karpenko, M. Yu. (2014). Motivation of siteonyms of the English-language sector of the Internet, *Bulletin of Voronezh State University. Series: Linguistics and intercultural communication*, 2, 35–39. (In Russian)
- Ke, X. (2020). Zhōngwén yù míng de qiánshì jīnshēng [The past and present life of Chinese domain names]. *Zhongguancun Theory*. 12. 66–68. (In Chinese)
- Kersten, S., Lotze, N. (2022). Anonymity and authenticity on the web: Towards a new framework in internet onomastics. *Internet Pragmatics*, 5(1), 38–65. <https://doi.org/10.1075/ip.00074.ker>
- Khazova A. (2023). Linguistic studies of computer-mediated communication, *Voprosy Jazykoznanija*, 2, 144–156. <https://doi.org/10.31857/0373-658X.2023.2.144-156>
- Kuprienko, M. S. (2010). Linguistic stylistic analysis of methods for creating Runet domain names, In: *Hypertext as an object of linguistic research: Proceedings of the All-Russian scientific and practical conference with international participation*, 80–83. (In Russian)
- Kuznetsov, A. V. (2012). Language game in domain names, *Science and business: ways of development*, 4 (10), 42–44. (In Russian)
- Li, J. (2010). Zhōngwén yù míng yǔ jìng xià de yù míng zhù cè wèn tí xīn jiě [New solutions to domain name registration issues in the context of Chinese domain names]. *Journal of Hunan Public Security College*. 22 (4). 117–121. (In Chinese)
- Madieva, G. B., Suprun, V. I. (2017). General problems of onomastics, Theory and practice of onomastic and derivatological research, In: *Collective monograph in memory of the Honored Scientist of the Republic of Adygea and Kuban*, 11–51. (In Russian)
- Polyak, Yu. E. (2018). Internationalized Cyrillic top-level domains, In: *Modern problems and perspective directions of innovative development of science: a collection of articles based on the results of the International Scientific and Practical Conference*, 116–119. (In Russian)
- Rozhkova, M. (2015). Domain names as identifiers and means of communication, *Economy and Law*, 3(458), 55–70. (In Russian)
- Ryabchenko, N. A. (2016). Toponymy of the network landscape of online space, *Community. Control*, Vol. 17, no. 4, 98–115. (In Russian)
- Shirin, S. S. (2014). Russian initiatives in Internet governance, *MGIMO Review of International Relations*, 6(39), 73–81. (In Russian)
- Son, L. P. (2012). Virtual onomastics and non-verbal component of Internet communication, *Scientific Bulletin of the Voronezh State University of Architecture and Civil Engineering. Series: Modern linguistic and methodological-didactic research*, 18, 123–134. (In Russian)
- Son, L. P. (2009). Virtual interaction: “What’s in your name...”, *Scientific notes of the Russian State Social University*, 6 (69), 217–221. (In Russian)
- Suprun, V. I. (2004). Development of the onomastic space of the Internet, In: *Onomastics of the Volga Region. Materials of the IX International Conference on Onomastics of the Volga Region*, 53–59. (In Russian)
- Xiang, X. (2012). Linguistic and Cultural Characteristics of Domain Names of the Top Fifty Most-Visited Websites in the US and China: A Cross-Linguistic Study of Domain Names and e-Branding. *Names*, 60(4), 210–219. <https://doi.org/10.1179/0027773812Z.00000000032> (In English)
- Zakalkin, P. V. (2022). Evolution of Cyberspace Management Systems, *Cybersecurity Issues*, 1 (47), 76–86. <https://doi.org/10.21681/2311-3456-2022-1-76-86> (In Russian)
- Zhiltsov, D. A. (2020). Features of domain naming in the Internet marketing strategy of an organization, *Marketing and logistics*, 3 (29), 33–40. (In Russian)
- Zubareva, A. V. (2021a). New trends in the system of non-personal proper names in the Russian language, *Linguistic and rhetorical paradigm: theoretical and applied aspects*, 26–2, 100–102. (In Russian)
- Zubareva, A. V. (2021b). Onims in Internet communication: new phenomena and functions, *Current problems of philology and pedagogical linguistics*, 1, 120–136. <https://doi.org/10.29025/2079-6021-2021-1-120-136> (In Russian)
- Zhou, Y. (2003). Shù zì xié yīn de yǎn biàn [The evolution of number homophony], *Rhetoric Learning*, 1, 35–37. (In Chinese)
- Zhu, G. (2014). Wèi shé me zhōng guó hù lián wǎng dì zhǐ piān hào shù zì yù míng [Why do Chinese Internet addresses prefer digital domain

names?], *Computers and Networks*, 1, 43–44. (In Chinese)

Список литературы

Abdullaeva M. Z. Spiritual role of Islamic media space in domain “uz” / M. Z. Abdullaeva // *Bulletin Social-Economic and Humanitarian Research*. 2020. № 8 (10). Pp. 55–62. <https://doi.org/10.5281/zenodo.4117558>

Akhrenova N. A. The role of numbers in the Chinese cyberculture / N. A. Akhrenova, M. N. Dubinina // *Topical Issues of Linguistics and Teaching Methods in Business and Professional Communication – TILTM 2022: Proceedings of the X International Research Conference Topical Issues of Linguistics and Teaching Methods in Business and Professional Communication, TILTM 2022, Moscow, 22–23 апреля 2022 года. – Moscow: ISO LONDON LIMITED European Publisher, 2022. Pp. 16–23. <https://doi.org/10.15405/epes.22104.3>*

An Zh. Wǎngluò shùzì xiéyīn cíyǔ qiǎn lùn [A brief discussion on homophonic words for digital numbers on the Internet]/ Zh. An // *Journal of Taiyuan University*. 2002. № 3 (4). Pp. 36–38.

Arsène S. Shaping the Chinese Internet Special Feature Internet Domain Names in China Articulating Local Control with Global Connectivity / S. Arsène // *China Perspectives*. 2015. № (4). Pp. 25–34. <https://doi.org/10.4000/chinaperspectives.6846>

Барина Д. С. Национальные домены интернета – символы государственных границ и безграничных возможностей / Д. С. Барина // *Вестник МГИМО Университета*. 2010. № 5 (14). С. 307–314.

Чайкисова А. В. Специфика идеонимов как особого типа имен собственных / А. В. Чайкисова // *Актуальные проблемы германистики, романистики и русистики*. 2017. № 3. С. 23–24.

Кутюр С. Что означает понятие «суверенитет» в цифровом мире? / С. Кутюр, С. Топин // *Вестник международных организаций: образование, наука, новая экономика*. 2020. Т. 15. № 4. С. 48–69. <https://doi.org/10.17323/1996-7845-2020-04-03>

Дальдинова Э. О. Г. Характеристика немецкоязычных сайтонимов / Э. О. Г. Дальдинова, Д. В. Бардаева // *Интеграционные процессы в современной науке: сборник научных трудов по материалам XIX Международной научно-практической*

конференции, Анапа, 26 апреля 2021 года. – Анапа: Общество с ограниченной ответственностью «Научно-исследовательский центр экономических и социальных процессов» в Южном Федеральном округе, 2021. С. 10–13.

Дальдинова Э. О. Г. Структурная характеристика немецкоязычных сайтонимов / Э. О. Г. Дальдинова, Т. В. Бураева // *Вестник Исык-Кульского университета*. 2023. № 54-2. – С. 65–73.

Дедова О. В. Заголовочный комплекс в электронной коммуникации / О. В. Дедова, М. С. Куприенко // *Вестник Московского университета. Серия 9: Филология*. 2013. № 1. С. 61–70.

Дубинина М. Н. Цифровые коды интернет-коммуникации (на примере английского, русского и китайского языков) / М. Н. Дубинина // *Актуальные проблемы германо-романской филологии и методики преподавания иностранных языков: Сборник научных статей, Коломна, 09 февраля 2022 года. М.: Некоммерческое партнерство «Национальное общество прикладной лингвистики», 2022. С. 23–26.*

Духанина Л. Н. Анализ сетевого онко-дискурса трехлетней ретроспективы в домене верхнего уровня RU / Л. Н. Духанина, А. А. Максименко // *Вопросы онкологии*. 2020. Т. 66. № 4. С. 315–324. <https://doi.org/10.37469/0507-3758-2020-66-4-315-324>

Филимоненко Е. Н. Культурные и лингвистические аспекты немецкоязычных сайтонимов / Е. Н. Филимоненко // *Научно-исследовательский центр "Вектор развития"*. 2023. № 19. С. 185–190.

Fu Y. Yīnghàn shùzì xiéyīn cí yìtóng de shēncéng jiēxī [An in-depth analysis of the similarities and differences between English and Chinese numerical homophonic words]/Y. Fu // *Language Application Research*. 2013. № 9. Pp. 135–137.

Габдрахманова Г. Ф. Национальный интернет России: к постановке проблемы / Г. Ф. Габдрахманова, З. А. Махмутов // *Oriental Studies*. 2018. Т. 11. № 3 (37). С. 142–151. <https://doi.org/10.22162/2619-0990-2018-37-3-142-151>

Хань Ю. Культурное осмысление омонимии числительных в современном китайском языке / Ю. Хань // *Современные*

востоковедческие исследования. 2022. Т. 4. № 2. С. 30–39. <https://doi.org/10.24412/2686-9675-2-2022-30-39>

Игнатов А. А. Управление Интернетом в повестке БРИКС / А. А. Игнатов // Вестник международных организаций: образование, наука, новая экономика. 2022. Т. 17, № 2. С. 86–109. <https://doi.org/10.17323/1996-7845-2022-02-04>

Калятин В. О. Доменные имена / В. О. Калятин. М.: Информационно-издательский центр Роспатента «ИНИЦ», 2002. 188 с.

Карпенко М. Ю. Выделение базовой единицы Интернет-ономастикона / М. Ю. Карпенко // Мова. 2012. № 17. С. 113–115.

Карпенко М. Ю. Мотивация сайтонимов англоязычного сектора Интернета / М. Ю. Карпенко // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2014. № 2. С. 35–39.

Ke X. Zhōngwén yù míng de qiánshì jīnshēng [The past and present life of Chinese domain names] / X. Ke // Zhongguancun Theory. 2020. № 12. Pp. 66–68.

Kersten S., Lotze N. Anonymity and authenticity on the web: Towards a new framework in internet onomastics / S. Kersten, N. Lotze // Internet Pragmatics. 2022. № 5 (1). Pp. 38–65. <https://doi.org/10.1075/ip.00074.ker>

Хазова, А. Б. Изучение компьютерно-опосредованной коммуникации в лингвистическом аспекте / А. Б. Хазова // Вопросы языкознания. 2023. № 2. С. 144–156. – <https://doi.org/10.31857/0373-658X.2023.2.144-156>

Куприенко М. С. Лингвостилистический анализ приемов создания доменных имен Рунета / М. С. Куприенко // Гипертекст как объект лингвистического исследования : Материалы Всероссийской научно-практической конференции с международным участием, Самара, 15–16 марта 2010 года / Ответственный редактор: С. А. Стройков. – Самара: Самарский государственный социально-педагогический университет, 2010. С. 80–83.

Кузнецов А. В. Языковая игра в названиях доменов / А. В. Кузнецов // Наука и бизнес: пути развития. 2012. № 4 (10). С. 42–44.

Li J. Zhōngwén yù míng yǔ jìng xià de yù míng zhù cè wèn tí xīn jiě [New solutions to domain name registration issues in the context of Chinese domain names] / J. Li // Journal of Hunan Public Security College. 2010. № 22 (4). Pp. 117–121.

Мадиева Г. Б. Общие проблемы ономастики / Г. Б. Мадиева, В. И. Супрун // Теория и практика ономастических и дериватологических исследований: Коллективная монография Памяти заслуженного деятеля науки Республики Адыгея и Кубани, профессора Розы Юсуфовны Намитокковой / Научные редакторы В. И. Супрун, С. В. Ильясова. Майкоп: Издательство «Магарин Олег Григорьевич», 2017. С. 11–51.

Порляк Ю. Е. Интернационализованные кириллические домены верхнего уровня / Ю. Е. Поляк // Современные проблемы и перспективные направления инновационного развития науки: сборник статей по итогам Международной научно-практической конференции, Новосибирск, 12 марта 2018 года. Том 2. Часть 2. Новосибирск: Общество с ограниченной ответственностью «Агентство международных исследований», 2018. С. 116–119.

Рожкова М. Доменные имена как идентификаторы и средства коммуникации / М. Рожкова // Хозяйство и право. 2015. № 3(458). С. 55–70.

Рябченко Н. А. Топонимика сетевого ландшафта online-пространства / Н. А. Рябченко // Человек. Сообщество. Управление. 2016. Т. 17. № 4. С. 98–115.

Ширин С. С. Российские инициативы по вопросам управления Интернетом / С. С. Ширин // Вестник МГИМО Университета. 2014. № 6 (39). С. 73–81.

Сон Л. П. Виртуальная ономастика и невербальный компонент Интернет-коммуникации / Л. П. Сон // Научный вестник Воронежского государственного архитектурно-строительного университета. Серия: Современные лингвистические и методико-дидактические исследования. 2012. № 18. С. 123–134.

Сон Л. П. Виртуальная интеракция: "Что в имени моем..." / Л. П. Сон // Ученые записки Российского государственного социального университета. 2009. № 6 (69). С. 217–221.

Супрун В. И. Развитие ономастического пространства Интернета / В. И. Супрун // Ономастика Поволжья: материалы IX Международной конференции по ономастике Поволжья, Волгоград, 09–12 сентября 2002 года. Волгоград: Институт этнологии и антропологии им. Н.Н. Миклухо-Маклая РАН, 2004. С. 53–59.

Xiang X. (2012, December). Linguistic and Cultural Characteristics of Domain Names of the Top Fifty Most-Visited Websites in the US and China: A Cross-Linguistic Study of Domain Names and e-Branding/ X. Xiang// Names. № 60 (4). Pp. 210–219.
<https://doi.org/10.1179/0027773812Z.00000000032>

Закалкин П. В. Эволюция Систем Управления Киберпространством / П. В. Закалкин // Вопросы кибербезопасности. 2022. № 1 (47). С. 76–86.
<https://doi.org/10.21681/2311-3456-2012-1-76-86>

Жильцов Д. А. Особенности доменного нейминга в стратегии интернет-маркетинга организации / Д. А. Жильцов // Маркетинг и логистика. 2020. № 3 (29). С. 33–40.

Зубарева А. В. Новые тенденции в системе неличных собственных имен в русском языке / А. В. Зубарева // Лингвориторическая парадигма: теоретические и прикладные аспекты. 2021. № 26-2. С. 100–102.

Зубарева А. В. Онимы в интернет-коммуникации: новые явления и функции /

А. В. Зубарева // Актуальные проблемы филологии и педагогической лингвистики. 2021. № 1. С. 120–136.
<https://doi.org/10.29025/2079-6021-2021-1-120-136>

Zhou Y. Shùzì xiéyīn de yǎnbiàn [The evolution of number homophony] / Y. Zhou // Rhetoric Learning. 2003. № 1. Pp. 35–37.

Zhu G. Wèishéme zhōngguó hùliánwǎng dìzhǐ piānhào shùzì yùnmíng [Why do Chinese Internet addresses prefer digital domain names?] / G. Zhu // Computers and Networks. 2014. № 1. Pp. 43–44.

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.

Maria N. Dubinina, PhD, Associate Professor, School of Foreign Languages, HSE University, Moscow, Russia.

Мария Николаевна Дубинина, PhD, доцент, Школа иностранных языков, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия.

УДК 81'42

DOI: 10.18413/2313-8912-2024-10-3-0-3

Савельева И. В.¹ 
Мельник Н. В.² 

Рецептивно-продуктивная деятельность субъектов
непрофессионального интернет-дискурса

¹ Кемеровский государственный университет
ул. Красная, 6, Кемерово, 650000, Россия
E-mail: saviren1973@mail.ru
ORCID: 0000-0003-4437-8489

² Кемеровский государственный университет
ул. Красная, 6, Кемерово, 650000, Россия
E-mail: saikova@mail.ru
ORCID: 0000-0001-9390-5842

*Статья поступила 13 мая 2024 г.; принята 15 сентября 2024 г.;
опубликована 30 сентября 2024 г.*

Аннотация: В работе исследуется непрофессиональный интернет-дискурс, порождаемый авторами интернет-комментариев. Он рассматривается сквозь призму категорий интенциональности, как характеристики текста, обладающего способностью выражать то или иное намерение его автора, и коммуникативных интенций, реализуемых личностью, которая выполняет две коммуникативные роли: читателя и автора одновременно. Произведенная выборка текстов демонстрирует, что в интернет-комментировании непрофессиональная личность в пределах одной текстовой единицы реализует и рецепцию, и продукцию. Реализация этих глобальных интенций непрофессионального интернет-дискурса осуществляется на разных уровнях: синтаксическом, семантическом и прагматическом. Анализ текстовых маркеров рецептивных и продуктивных интенций позволяет установить закономерности в их реализации через три основных модели: сцепление, диффузия и наложение. При порождении текста по модели «сцепление» авторы комментариев преимущественно последовательно и эксплицитно демонстрируют переход от рецепции к продукции, на что указывают дискурсивные слова, синтаксическая организация текста интернет-комментария. Более сложным вариантом реализации рецепции и продукции в пределах одного интернет-текста является модель «диффузия». Здесь рецепция и продукция реализуются одновременно, за счет тесно переплетающихся в тексте маркеров реакции и продукции на уровнях синтаксиса, семантики и прагматики. Рецепция устанавливается на уровне нелинейных связей текста и семантико-прагматических характеристик интернет-комментария. В третьем варианте происходит реализация рецепции и продукции с помощью одних и тех же текстовых средств. Поэтому указанная модель получает название «наложение». В интернет-комментариях, где наблюдается наложение, рецептивно-продуктивная деятельность непрофессиональной личности эксплицируется на уровне вертикальных связей текста и его контекстуальных характеристик. Количественный анализ текстовых типов по выделенным критериям показывает, что в русскоязычном дискурсе наиболее распространенный вариант реализации рецепции и продукции – сцепление. В англоязычном дискурсе чаще представлена

модель диффузии. В целом двойной коммуникативный статус авторов интернет-комментариев выявлен в 19.5 % англоязычных текстов и 27.4 % русскоязычных текстов. В перспективах исследования проведение лонгитюдного экспериментального исследования интенциональности непрофессионального интернет-дискурса.

Ключевые слова: Непрофессиональный интернет-дискурс; Коммуникативный статус; Автор; Адресат; Глобальная интенция; Интенциональность; Рецепция; Продукция; Интернет-комментарий.

Информация для цитирования: Савельева И. В., Мельник Н. В. Рецептивно-продуктивная деятельность субъектов непрофессионального интернет-дискурса // Научный результат. Вопросы теоретической и прикладной лингвистики. 2024. Т. 10. № 3. С. 42-69. DOI: 10.18413/2313-8912-2024-10-3-0-3

UDC 81'42

DOI: 10.18413/2313-8912-2024-10-3-0-3

Irina V. Savelyeva¹ 
Natalya V. Mel'nik² 

Receptive and productive activity of non-professional Internet
discourse participants

¹ Kemerovo State University,
6 Krasnaya St., Kemerovo, 650000, Russia
E-mail: saviren1973@mail.ru
ORCID: 0000-0003-4437-8489

² Kemerovo State University,
6 Krasnaya St., Kemerovo, 650000, Russia
E-mail: saikova@mail.ru
ORCID: 0000-0001-9390-5842

Received 13 May 2024; accepted 15 September 2024; published 30 September 2024

Abstract: The paper examines the non-professional Internet discourse generated by the authors of Internet comments. It is viewed through the prism of the categories of intentionality, as characteristics of a text that has the capacity to express one or another intention of its author, and as communicative intentions realised by a person who simultaneously plays two communicative roles: the reader and the author. The textual data demonstrate that in Internet commenting, an unprofessional personality implements both reception and production within one text unit. The implementation of these global intentions of non-professional Internet discourse is carried out at different levels: syntactic, semantic and pragmatic. The analysis of textual markers of receptive and productive intentions allows us to establish patterns in their implementation through three main text types: cohesion, diffusion and overlapping. Each of the text types has several variants. Remarkably, when generating a text according to the cohesion model, the authors of the comments consistently and explicitly demonstrate the transition from reception to production, as indicated by discursive words, marking the syntactic organization of the commentary. A more complex variant of the implementation of reception and production within the same Internet text is the text type called “diffusion”. In this model, reception and production are realized simultaneously, due to the closely intertwined markers of reaction and production in the text at the levels of syntax, semantics and pragmatics. Reception is established at

the level of non-linear text liaisons as well as semantic and pragmatic characteristics of an Internet comment. In the third model, reception and products are realized using the same textual means. Therefore, the specified text type is called "overlapping". In the commentary constructed according to the overlapping principle, the receptive and productive activity of a non-professional personality is explicated at the level of vertical connections of the text and its contextual characteristics. A quantitative analysis of text types according to the selected criteria shows that in Russian-language discourse, the most common variant of the implementation of reception and products is cohesion. The diffusion model is more often presented in English-language discourse. In general, the dual communicative status of the authors of Internet comments was revealed in 19.5% of English-language texts and 27.4% of Russian-language texts. The prospects of the study include conducting a longitudinal experimental study of the intentionality of non-professional Internet discourse.

Keywords: Non-professional Internet discourse; Communicative status; Author; Addressee; Global intention; Intentionality; Reception; Production; Internet commentary.

How to cite: Savelyeva, I. V., Mel'nik, N. V. (2024). Receptive and productive activity of non-professional Internet discourse participants, *Research Result. Theoretical and Applied Linguistics*, 10 (3), 42-69. DOI: 10.18413/2313-8912-2024-10-3-0-3

Введение (INTRODUCTION)

На современном этапе развития коммуникации влияние медиа на все сферы жизни и стремительное развитие коммуникативных технологий меняют статус адресата текстов СМИ. Сегодня читатель – это участник речевого взаимодействия, обсуждающий события внешней действительности, особенно те, которые включаются СМИ в повестку дня (McCombs, 2005; Tsfati, 2014; Wettstein, Wirth, 2017; Wojcieszak, Mutz, 2009). Читатель новостных материалов не просто стремится внести вклад в «шумящий информационный рой» (Дейк, 2013: 10), он вербализирует свою позицию, критически реагируя на новостные публикации и формируя сообщество «новых читателей» (Ковалевская, 2022). «Новые каналы коммуникации позволяют читателю включиться в оценку и производство информации, что оказывается особенно актуально в моменты импульсных событий и кризисов» (Лукина, 2006: 54).

Дискурс, порождаемый как ответная реакция на медиаконтент в сетевом пространстве, сочетает характеристики медийного и персонального типов. Рассматривая публикации к статьям

политической тематики, мы определяем тексты, создаваемые любым пользователем, рядовым носителем языка, как непрофессиональный дискурс Интернета. Автор анонимного интернет-комментария вербализирует результат восприятия новости на интересующую его тематику, не эксплицируя при этом свою профессиональную принадлежность, или социальный статус. Он варьирует вербальные реакции от короткого отклика до полноценного текста, тем самым внося вклад в порождение дискурса непрофессионального типа. Например, рассуждая о политике в ленте комментариев, комментаторы в разной степени являются экспертами, но большая их часть – обычные читатели, который смотрят на политическую сферу с позиции внешнего наблюдателя.

Такая прагматическая характеристика непрофессионального дискурса, как вторичность текстовой деятельности накладывает свой отпечаток на его структуру. С одной стороны, НД характеризуется спонтанным текстопорождением как результатом рецепции медийного дискурса, а с другой, ставит перед непрофессиональной

личностью (НЛ) две задачи: она и реципиент, и продуцент информации одновременно (Арутюнова, 1981).

Репрезентация непрофессиональной личностью себя одобряется технологическими возможностями глобальной сети и новыми медиа, порождающими креативного читателя. Глен Крибер и Мартин Ройстон пишут об этом следующее: «...the increased interactivity of New Media generally allows audiences to play around with and make their own composite identities from various and sometimes even contradictory sources. This process is referred to <...> as ‘DIY citizenship’, the notion that the media now allows us to all create our own complex, diverse and many faceted notions of personal identity» (возросшая интерактивность новых медиа, как правило, позволяет аудитории экспериментировать и конструировать свою собственную комплексную идентичность из различных, а иногда даже противоречивых источников. Этот процесс называется <...> “создавай свой гражданский контент своими руками”, подразумевающим, что средства массовой информации теперь позволяют нам всем создавать свои собственные сложные, разнообразные и многогранные представления о личной идентичности) (Creeber, Royston, 2008: 18).

В то же время, одним из ведущих принципов формирования современного медиaprостранства является следующий: «концепция будущего текста (медiateкста – И.С., Н.М.) и его цель формируются как реакция на запросы аудитории и как стимул к ее ответу в дальнейшем»¹.

Эти факторы обуславливают формирование медiateкстов, в совокупности относимых к сфере «гражданской» журналистики (Wall, 2017).

Одной из разрастающихся дискурсивных практик Интернета является

комментирование новостей в СМИ. Значительная доля комментариев, публикуемых к новостным и информационно-аналитическим статьям, представляет собой тексты, в которых читатель проявляет себя и как адресат, и как автор. Субъект непрофессионального дискурса как адресат реализует глобальную интенцию когнитивно-рецептивного характера (Клушина, 2008). В процессе кодирования/декодирования полученной информации он решает следующие прагматические задачи: оценивать/легитимировать политические события и поведение политических акторов; оценивать/легитимировать подачу новостной повестки и дискурсивные события медиаканалами; обнаруживать/устанавливать скрытые смыслы в предлагаемых СМИ версиях событий (Иванова, 2023; Мельник, 2021; Романтовский, 2024; Савельева, 2019; Сар, 2008). Реализация продуктивной интенции субъекта непрофессионального дискурса детерминируется как лингвистическими факторами (влияние языка СМИ как особой сферы коммуникации на производство комментариев, с одной стороны, и речевых характеристик личности, с другой), так и экстралингвистическими (влияние особенностей менталитета личности, социальных условий, политической ситуации). Иными словами, в роли автора непрофессиональная личность реализовывает замысел своего микротекста на уровнях синтаксиса, семантики и прагматики.

Основная часть (MAIN PART)

Обзор литературы (Literature review)

Сложность системы отношений двух антропоцентров – автора и адресата – в дискурсе и тексте давно привлекает внимание языковедов. Вопросу разграничения дихотомий «говорящий/слушающий», «пишущий/читающий», «адресант/адресат», «автор/читатель»,

¹ Текст. Теоретические основания и принципы анализа: учебное пособие / под ред. К. А. Роговой. Златоуст, 2011. С. 22.

«реципиент/продуцент»,
«utterer/interpreter»
(говорящий/интерпретирующий – англ.) посвящено немало исследований в области философии языка, семиотики, прагмалингвистики, психолингвистики, литературоведения, дискурс-анализа и других областей знания, исследующих различные аспекты текста и дискурса (Арутюнова, 1981; Виноградов, 1961; Винокур, 1993; Барт, 1994; Эко, 1998; Verschueren, 1999). Способы актуализации адресат-адресантных отношений в дискурсе и тексте показывают, что существует объективная разница между реальными статусами автора/адресата и данными феноменами как составляющими категориальный аппарат лингвистики.

Антропометафора, предложенная М. М. Бахтиным и включенная в дискурсивный контекст Ю. Кристевой, отражает комплексность и неоднозначность описываемых отношений: «Слово-дискурс разлетается на “тысячу осколков” оттого, что попадает во множество контекстов; в контекст дискурсов, в межтекстовое пространство, где расщепляется и рассеивается не только говорящий, но и слушающий субъект, то есть мы сами, а именно, то “единство”, которое Бахтин именуется “человеком”, но которое, по его собственным словам, восстает против всякого овнешняющего определения, потому что “человек никогда не совпадает с самим собой”» (Кристева, 2004: 15).

В настоящем исследовании нас интересуют модели реализации непрофессиональной личностью реактивной и репрезентативной интенций в пределах одного текста, которые обусловлены доминированием в дискурсе коммуникативных статусов комментатора как читателя или как автора, а также дискурсивная реализация случаев перехода адресата в адресанта в непрофессиональном дискурсе, субъекты которого вовлечены в практику оценивания и обсуждения медиопубликаций.

Категория адресата медиадискурса в последние декады развития лингвистики претерпевает значительные изменения, имеющие объективную обусловленность экстралингвистической ситуацией: формирование гражданской журналистики и активное участие рядового носителя языка в производстве медиаконтента (Дускаева, 2012; Клушина, 2008; Пром, 2020; Цветова, 2012; Чепкина, 2000; Domingo, 2008; Wall, 2017).

В медиадискурсе рецептивные и когнитивные возможности адресата планируются автором произведения и включаются в авторский замысел. Система текстовых координат, нацеленная на взаимодействие профессионального автора медиатекстов со своим читателем, в качестве облигаторного компонента включает систему средств, в конечном итоге формирующих образ адресата. «Адресат может служить таким же генерирующим началом, он играет ту же роль, которую В. В. Виноградов отводит образу автора в художественном тексте: он обуславливает способы построения этого текста и способы интерпретации фактов действительности» (Каминская, 2008: 316).

Однако коммуникативные роли читателя и автора могут предполагать одного и того же субъекта. Неделимость коммуникативной диады «адресант-адресат» изучается в рамках коммуникативно-прагматического подхода к исследованию текста. Важность функционирования адресата, его роли в восприятии и интерпретации сообщения и диффузность двух коммуникативных ипостасей, отмечается Н. Д. Арутюновой: «Переход к речевому реагированию превращает получателя речи в говорящего» (Арутюнова, 1981: 359).

Ситуации, где нет строго зафиксированного статуса адресата и адресанта, описываются в работе Дж. Вершуерена следующим образом: «not only can utterer and interpreter coincide at certain moments, there are also stages in certain types of speech events when there is

no interpreter outside the one mentally constructed by the utterer; in those cases, and at that stage, the interpreter role is simply incorporated into the world of the utterer \diamond in fact, I can be telescoped into U, but not the other way around, since there is always is or has been an utterer» (мало того, что говорящий и слушающий могут совпадать в определенные моменты, в определенных типах речевых событий также существуют этапы, когда у собеседника нет слушающего, кроме того, которые мысленно сконструированы самим говорящим; в этих случаях и на таких этапах роль слушающего просто включена в мир говорящего \diamond на самом деле, Г (говорящий) может быть преобразован в С (слушающего), но не наоборот, поскольку всегда есть или был тот, кто произносит) (Vershueren, 1999: 76).

Иначе рассматривает ситуацию смены коммуникативных ролей Н. А. Пром. С ее точки зрения, случаи пребывания субъекта дискурса в двойной коммуникативной роли часто предполагают, что интенция адресата предшествует интенции автора, «поскольку, прежде чем сообщить о чем-либо, человек получает информацию об этом из внешней реальности — объективной (от окружающего мира), субъективной (от другого человека) либо медиареальности» (Пром, 2020: 97).

Эта позиция созвучна результатам исследований в области лингвистической прагматики, вторичного текстопорождения, интерпретации текста, семиотики. У. Эко говорил: «Текст, физически конечный и предельный, может интерпретироваться бесконечными способами, или, скажем, очень многими способами, но – не любыми способами» (Эко, 1998: 13).

Результатом ответной речевой деятельности субъекта дискурса являются особые речевые структуры, в лингвистической прагматике рассматриваемые как реактивные речевые акты (Арутюнова, 1981; Путина, Балакин,

2022). Данный вид высказываний более характерен для диалогического взаимодействия. Однако сегодня такой вид дискурса, как дискурс обратной связи, распространяется в письменной форме через онлайн технологии. Реактивные ответы пользователей показывают, насколько сильное коммуникативное воздействие на читателя оказывает новостной текст (Савельева, 2021). Интернет-комментарий как макроречевой акт имеет сложную природу, поскольку в текстах происходит совмещение процессов рецепции и продукции текста именно в такой последовательности: интериоризация фактов внешней действительности (физически или дискурсивно представленной), а затем порождение нового дискурса как результата осмысления полученного знания (Ким, 2012; Клушина, 2008; Мельник, Савельева, 2017; Пром, 2020).

Интенциональный подход в медиаисследованиях позволяет указать на группу реализуемых языковой личностью когнитивно-реактивных интенций, которые являются базой для реализации глобальных стратегий. Одна из глобальных стратегий публицистического дискурса, по мнению Н. И. Клушиной, стратегия убеждения. Ее реализация и, следовательно, интенциональность текста невозможны без реакции читателя – реципиента информации. По словам Н. И. Клушиной, «чтобы глобальная стратегия – убеждение – осуществилась, недостаточно только лингвистических способов ее реализации, необходима ответная реакция со стороны читателя как равноправного участника коммуникации. Без понимания предложенного читателю авторского текста коммуникация не может состояться» (Клушина, 2008: 21).

Комплексная коммуникативно-прагматическая организация медийного текста, обусловленная действием целого спектра коммуникативных интенций автора, в медиаисследованиях понимается как полиинтенциональность (Дускаева,

Цветова, 2012). Данный феномен характеризует и непрофессиональный дискурс, понимаемый как совокупность текстов, порождаемых личностью, не проявляющей свой профессиональный статус. Тексты такой личности рождаются как результат действия следующих глобальных интенций: когнитивно-реактивной, репрезентативно-оценочной, коммуникативно-интерактивной (Савельева, 2021).

Повсеместное распространение медиа и их воздействие на формирование языковой способности личности превращает ее в личность информационно-медийного типа. В зависимости от принадлежности или непринадлежности к профессиональной сфере текстовой деятельности такая личность может выступать в следующих ипостасях: информационно-медийная личность (Болотнов, 2015), целевая аудитория (Каминская, 2008), непрофессиональный ретранслятор (Пром, 2020), участник дискурса «второго круга» (Fetzer, Lauerbach, 2007).

Адресат СМИ, до некоторого времени справедливо считавшийся размытым, непроявленным, в условиях виртуальной коммуникации остается им лишь в некоторой степени, учитывающей анонимность публикаций. Он часто играет роль не субъекта дискурса, но наблюдателя, «участника второго круга» (Fetzer, Lauerbach, 2007). Эта специфическая черта незримого присутствия адресата в любом отрезке текста сохраняется и в исследуемом нами материале: одни авторы выбирают Я-повествование, приводят факты из личной жизни, а другие производят фактологический, нейтральный по тональности дискурс, дистанцируя себя как физическую личность от себя как субъекта дискурса.

В политическом дискурсе отношения участников также связаны с распространением политического дискурса на медиаплатформах. Позиция

наблюдателя, приписываемая языковой личности рядового носителя языка как представителю общества, типична в ситуации политической коммуникации. Так, О. Л. Михалева отмечает, что среди адресатов разного типа в политическом дискурсе массы, наблюдающие за тем, как разворачиваются политические события (предвыборные кампании, политические переговоры, противостояние политических элит, медийное воплощение политической риторики и. т.п.), как правило, занимают позицию стороннего наблюдателя (Михалева, 2008).

В классификации, предлагаемой А. Фетцер, выделяются участники политического дискурса первого круга (*first-frame coparticipants*), подразумеваются сами представители медиа, как если бы они были единственными получателями сообщений. Участники второго круга (*second-frame coparticipants*) – реципиенты, выступающие в статусе аудитории, получившей прямой доступ к информации (*the ratified generic audience*) и являющиеся интерпретаторами сообщений с позиции ратифицированных слушателей. Статус последних дает участникам второго круга (то есть непрофессиональному адресату) право оценивать сообщения, предназначенные для журналистов, и, что более важно, оценивать сообщения медиа и стратегии, реализуемые участниками первого круга (то есть медиа) при производстве и интерпретации политического дискурса (Fetzer, Lauerbach, 2007: 147).

Цель работы (Research Objectives)

Цель данной статьи – смоделировать интенциональную структуру дискурса интернет-комментариев, поместив в основание механизмы реализации интернет-комментатором двух глобальных интенций – когнитивно-рецептивной и продуктивно-репрезентативной в пределах одного текста.

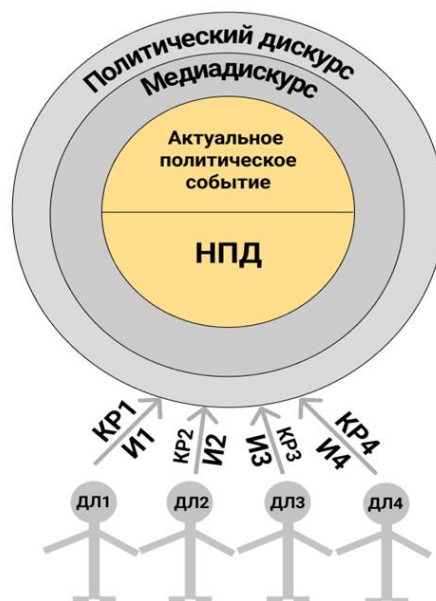
Материалы и методы исследования (Materials and Methods)

Настоящее исследование следует традициям коммуникативно-прагматического подхода к исследованию текста и дискурса, лингвопрагматических и лингвоперсонологических исследований дискурса, с одной стороны, а с другой, выполняется в русле разработанной

модели непрофессионального политического дискурса (НПД), в основе которой лежит понятие глобальных интенций, реализуемых личностью при производстве НПД (Савельева, 2021). Данная модель схематично представлена на рисунке.

Рисунок 1. Интенциональная модель НПД

Figure 1. Intentional model of non-professional political discourse



В модели ключевыми понятиями являются коммуникативные роли участников дискурса и реализуемые ими глобальные интенции, которые в конкретных текстах варьируются и распадаются на частные интенциональные сценарии. Эти интенции могут быть разделены на четыре группы: И 1 – когнитивно-рецептивные (или реактивные), И 2 – продуктивно-репрезентативные, И 3 – политико-идеологические и И 4 – интерактивно-диалоговые. Соответственно, коммуникативный статус личности варьируется в зависимости от преобладания в тексте интернет-комментария средств, маркирующих реализацию одной из четырех глобальных интенций. Дискурсивная личность автора интернет-комментария (ДЛ) также зависит

от коммуникативного статуса, в котором позиционирует себя НЛ.

С учетом перечисленных концепций представляется обоснованным полагать, что интенциональность интернет-комментария как текстовая категория, в первую очередь, устанавливается на основе выделения средств, привлекаемых для реализации двух коммуникативных интенций: когнитивно-рецептивной и продуктивно-репрезентативной, имеющих глобальный характер в аспекте порождения интернет-дискурса. Случаи их реализации в пределах одного текста рождают многочисленные комбинации. Однако дискурсивное пространство, создаваемое непрофессиональной личностью, возможно подвергнуть систематизации и на основе типичных сочетаний текстовых реализаций выявить устойчивые варианты рецептивно-

продуктивной деятельности непрофессиональной личности.

Моделирование типов текстовой реализации НЛ статусов читателя и автора в своей основе содержит прагматические категории интенции субъекта и интенциональности производимого им дискурса. Данные категории прослеживаются в порождении конкретных текстовых продуктов – интернет-комментариев и определяют синтаксис, семантику и прагматику текста как семиотического продукта и как результата порождения дискурса НЛ. Для осуществления цели исследования и выявления средств реализации авторских интенций применялись методики лингвопрагматического и лингвостилистического анализа текста, методика количественного подсчета выделенных вариантов текстовых единиц. Эмпирической базой исследования послужили интернет-комментарии к политическим медиаматериалам, опубликованным «РИА Новости» и «MSN-US» в 2022–2024 годах. Общее количество текстов составило 1357 и 2579 единиц соответственно.

Результаты исследования и их обсуждение (Results and Discussion)

Текстовые маркеры когнитивно-рецептивной интенции

В результате рецепции комментатор создает отклики, проявляющие зависимость от исходного текста, демонстрирующие ее на разных его уровнях. Синтаксический уровень текста характеризуется наличием метатекстовых показателей, дискурсивных слов, помогающих передать отношения между единицами текста, средств устно-диалогического общения. Семантический план текста характеризуется тематической общностью отклика и стимула, наличием тождественных референтов. Прагматический уровень комментария, как результата вербализации рецепции исходной новости характеризуется

включением анафорических и дейктических элементов.

Подчас, реализация рецептивной интенции может быть отлита в форму осознанного реагирования, где доминантой рецепции является формальный план текста, а актуализатором рецепции – метатекстовые маркеры, как в следующем примере:

Очень важное дополнение!!! Я цитирую сайт РИА Новости. Первая польская комиссия под руководством экс-главы МВД Ежи Миллера тогда пришла к аналогичным выводам, назвав причиной катастрофы снижение самолета ниже дозволенного минимума в условиях тумана." **Надо было сказать!!! - что возле аэродрома в Смоленске есть глубокий овраг! Лётчик шел в тумане по показаниям высотомера. высотомер показывал Неправильную высоту! – от дна оврага! Поэтому самолёт шел слишком низко и врезался в деревья возле аэродрома...².**

Автор комментария сам определяет функцию своего текста как комплементарную. Реактивная интенция актуализируется за счет включения в комментарий метатекстовых показателей *дополнение*, *я цитирую*, *надо было сказать*. Как читатель, комментатор обнаружил недосказанность новостного источника и обратил внимание на нарушение максимы количества (говори достаточно для того, чтобы тебя поняли). Переходя от рецепции к продукции с помощью метатекстового маркера *надо было сказать* с прагматикой «экспертного мнения», автор становится на более высокую коммуникативную позицию и привлекает медийный контекст, а также выражает упрек через использование директивной конструкции с модальным словом *надо*. Автор включает в текст последовательное описание этапов

² В Польше заявили о фальсификации экспертизы по смоленской катастрофе (2023). URL: <https://ria.ru/20231215/ekspertiza-1916095785.html> (дата обращения: 26.04.2024).

события (авиакатастрофы над Смоленском), чем показывает, что он более сведущ в способах корректной подачи информации. На уровне графической экспликации автор также выражает свое возмущение сокрытием известных широкой общественности фактов.

В следующем примере прагматика реактивности проявляется через введение в комментарий и неоднократное использование автором анафоры:

*Авторы этой чудовищной доктрины желают, чтобы на земле остался только “золотой” миллиард. Одно только обстоятельство затмевает эту идею, как-то уж больно он с***дит, этот “золотой миллиард”³.*

Актуализатором рецептивной интенции в тексте является его прагматический план, построенный на основе анафоры. Системное использование указательных местоимений *этой, эту, этот* создает эффект эха и указывает на деривационную связь как внутри интернет-комментария, так и вне его – связь с дискурсом-стимулом. Как читатель, комментатор устанавливает импликации первичного текста, которые оказали следующий когнитивный эффект: актуализировали сферу знаний, называемую «теория заговора». Использование эпитета *чудовищный* отражает возбужденное эмоциональное состояние автора относительно тиражируемой в СМИ и в непрофессиональном дискурсе версии политического поведения элит. Самостоятельное текстообразование (продукция) опирается на метафорический перенос качества с одного объекта на другой. Смысловой уровень также показывает, что денотат фразы «золотой миллиард» качественно противоречит его сигнификату. Использование инверсии *уж больно он с***т* усиливает негативный

смысл идеологически нагруженного выражения «золотой миллиард».

В третьем примере прагматика адресата наиболее явно эксплицируется на уровне референциальных связей текста:

Нам не надо озвучивать причины снижения курса рубля, нам нужно знать, что ЦБ делает что бы курс рубля укрепился!!!! То есть ЦБ должен осуществлять функции возложенные на него КОНСТИТУЦИЕЙ!!!⁴.

Оценка читателем деятельности Центробанка России производится через актуализацию системы отношений участников политического дискурса «власть – народ», которая в данном тексте репрезентирует позицию адресата через использование косвенной формы местоимения «Мы». Синтактика повтора *нам, ... нам* и семантика контраста действий власти воле народа «власть озвучивает причины – народ хочет знать, что власть делает» отражают переход от позиции читателя к позиции автора. Если читатель счел информацию нерелевантной, то автор призывает к соблюдению властью (в частном примере ЦБ – центральным банком) своих конституционных прав. Деонтическая модальность заключительной фразы *ЦБ должен осуществлять* прямым образом указывает на политическую позицию автора комментария как субъекта политического дискурса.

Текстовые маркеры продуктивной интенции

Выступая в роли автора, комментаторы стараются отмежеваться от обсуждаемой новости, «отпочковать» свой дискурс от медиаконтента, создать иную дискурсивную реальность и сконструировать свою политическую действительность. В текстах комментариев это проявляется следующим образом: на уровне синтаксиса реализация

³ Минобороны: США создали вариант коронавируса с 80-процентной смертностью (2024). URL: <https://ria.ru/20221126/koronavirus-1834450813.html> (дата обращения: 26.04.2024).

⁴ Первый зампред ЦБ назвала причину снижения курса рубля (2023). URL: <https://ria.ru/20230704/rubl-1882171857.html> (дата обращения: 26.04.2024).

продуктивной интенции отражается на композиции текста, на форме повествования (рассказ от первого лица, включение автобиографических фактов); на уровне семантики – выбор смежной/периферийной темы, привлечение дополнительной информации как из политической, так и из других сфер жизни; на уровне прагматики – конструирование своей идентичности, представление себя как эксперта по обсуждаемому вопросу. Обратим внимание на следующий пример:

Почему вы так думаете? Комментарии даются *не *стадом, идущим, куда пастух поведёт**, а людьми, каждый из которых мыслит вполне самостоятельно. У меня (к счастью? или к сожалению?) присутствует неистребимая **привычка думать**. Дана была в советской школе. Говорят, самое лучшее образование в то время было (и с этим не поспорить). **Привычка думать** осталась на всю жизнь...⁵.

Структура данного комментария эксплицирует наличие средств реализации реактивной интенции через реактивный зачин устно-диалогического типа: *Почему вы так думаете?* В следующем высказывании наблюдается отрицание мнения собеседника через противительную синтактику (*не..., а...*) и аксиологическую семантику «хорошо – плохо» посредством противопоставления лексем *люди* и *стадо*. Затем происходит переход от рецепции к продукции. Автор привлекает Я-историю, композиционно оформленную фреймом и содержательно отклоняющуюся от корневой ветки обсуждений. Как читатель, комментатор находит слова собеседника оскорбительными для сообщества, а как автор, он представляет в тексте фрагмент личной биографии, композиционно отделив свою Я-историю от рецептивной части. Показателем начала продуктивной

⁵ СМИ: Израиль поручил посольствам давить на дипломатов из-за иска ЮАР (2024). URL: <https://ria.ru/20240106/izrail-1919916287.html> (дата обращения: 26.04.2024).

части текста является пропозиция *У меня... привычка думать*, а ее завершение содержит ее модификацию *привычка думать осталась...*, где рематический компонент перемещается на позицию тематического (известного компонента).

«Новый» читатель осознает полиинтенциональную природу своего дискурса и подвергает рефлексии коммуникативную роль продуцента, подчас делая это в упрощенно-обиходной манере, характерной для бытового дискурса рядовых носителей языка:

*Не старался вас обидеть. **Весь чат на публику**, как и **вся лента РИА Новости***⁶.

*У нас **всегда так**. **Вчера** был главным вирусологом, **сегодня** главным военным экспертом, а **завтра** главным финансистом страны! И **все это** один человек на диване в мамкиной кофме!*⁷.

*I've been **reading through the comments** and I realized something. All the liberals commenting either don't know or don't care that the United States has a **Mutual Security Treaty with Israel**. Under that **treaty**, we are **LEGALLY REQUIRED** to provide aid and support to Israel against its enemies. Don't like it, contact your Senators and ask them to **nullify that treaty**. Good luck...because it takes a Super Majority Vote to **nullify a treaty***⁸.

Представленные отклики демонстрируют рефлексивность участников непрофессионального дискурса относительно коммуникативно-прагматических задач, решаемых ими в дискурсивной практике данного типа. Так,

⁶ Дугин А. Убить дракона: мир стоит на пороге глобальной войны (2024). URL: https://ria.ru/20240117/globalnaya_voyna-1921723243.html (дата обращения: 26.04.2024).

⁷ Там же.

⁸ Kika T. Democrats Issue Warning to Joe Biden (2024). URL: <https://www.msn.com/en-us/news/world/democrats-issue-warning-to-joe-biden/ar-BB1kpSQL?ocid=hpmsn&cvid=4ce8b0e674a54f2a85d4c0720dc2f697&ei=76#comments> (дата обращения: 26.04.2024).

в первом комментарии автор приносит свои извинения собеседнику, ссылаясь на следование принципу работы *на публику*, то есть определяя театральность как ведущую стратегию дискурса комментариев. Выражая авторскую позицию, он как раз причисляет себя к сообществу тех, кто реализуют двойную роль.

Во втором отклике реактивная часть, эксплицирующая статус читателя (реактивный зачин текста), оформлена семантическим приемом генерализации: *У нас всегда так*. За обобщением как вербальным проявлением рецепции следует ирония, привлекаемая как отражение позиции автора. Для ее репрезентации автор использует синтактику известного слогана *вчера, сегодня, завтра*. В понимании автора коммуникативное поведение и экспертные мнения собеседников заслуживают осмеяния. Завершение текста на семантическом уровне маркируется генерализацией (*И все это*), на синтаксическом – кольцевой композицией.

Третий отклик интересен эксплицитным стилем текстообразования с использованием метакоммуникативных средств. В тексте комментатор открыто рефлексировать относительно своей рецептивной деятельности как читателя. Первое высказывание выражает интенцию реципиента. Однако начиная со следующего высказывания, реципиент переходит в статус автора, желающего выразить свое, оформившееся в процессе восприятия новости, видение вопроса, активно обсуждаемого сообществом MSN-US. Продуктивная часть открывается генерализатором *all*, который отражает склонность комментатора к обобщениям. Как и в предыдущих примерах на русском языке, автор англоязычного комментария использует лексические повторы. Так, лексема *treaty* повторяется четыре раза, два раза – в одинаковом синтаксическом оформлении. Кроме вводной обобщающей конструкции, автор комментария

использует обобщающее «мы», тем самым «укрупняя» свою фигуру и обозначая принадлежность к гражданскому обществу, обладающему демократическими правами.

Текстовые варианты реализации рецептивной и продуктивной интенций непрофессиональной личностью

Возможность давать обратную связь и репрезентировать себя и свои речемыслительные способности приводит к появлению текстов, сочетающих реактивную инициацию как способ реализации обратной связи и проявленную авторскую позицию, как способ самопрезентации. Реакция и самопрезентация как стимулы производства дискурса в тексте вступают в некую систему отношений. Проследим специфику и имеющиеся варианты таких отношений с точки зрения синтаксиса, семантики и прагматики текста.

Сцепление. Синтаксический уровень текста отражает последовательную реализацию реакции и продукции, что коррелирует с принципом линейности текстовых связей. Переход от рецепции к продукции совершается эксплицитно. Комментарий открывает реактивный зачин – реплика комплементарного характера, преимущественно имеющая признаки устного диалогического дискурса и демонстрирующая зависимость текста от новости-стимула. После зачина следует выражение позиции автора текста, как субъекта непрофессионального дискурса.

Сцепление с помощью дискурсивных слов. Комментарии данного типа по композиции имеют выраженный реактивный компонент в начале текста, маркирующий коммуникативную роль адресата. Так, автор следующего комментария выбирает способ реакции по устно-диалогическому типу, показывая, что он и адресат, и в то же время собеседник:

К сожалению, да. Дело в том, что Западные политики понимают, что на данный момент капитуляция Киева, это прямая дорога к Нюрнбергу, где придется

отвечать за каждое политическое слово, каждую подпись, каждый приказ. **Поэтому** они будут **максимально** **затягивать** конфликт, **при этом** **стращать** европейских обывателей войной с Россией, скрести по сусекам деньги налогоплательщиков и менять их на агонию украинского народа, **по сути** обрекая его на истребление, на осознанный геноцид славян, в угоду политической трусости, которая **в совокупности** с непомерной алчностью, создает условия для большой войны⁹.

Реактивная реплика *К сожалению, да* – вербализатор реакции читателя новостей. Коннектив *Дело в том* сцепляет реакцию на стимул и продукцию как репрезентацию политической позиции комментатора. Эта вводная конструкция с семантикой экспланаторности одновременно является маркером и реактивной интенции (толкование и косвенное несогласие), и продуктивной, так как вводит аргумент автора. Употребление определения *западные* с использованием верхнего регистра, уподобляющего лексему топониму *Запад*, также рассматривается нами как окказиональный маркер репрезентации в тексте авторского «Я». Следование автора англоязычному варианту орфографической нормы создает образ вступления, открывающего самостоятельное текстовое произведение автора комментария. Его текст содержит развернутое рассуждение и функционирует как самостоятельное целое. За исключением двух первых реактивных речевых конструкций. Сюжет текста также прозрачен, как и точка зрения автора. Речемышлительный вектор направлен на рациональное описание политической ситуации. Внутренняя форма текста согласуется с рациональной стратегией порождения, однако имеются и некоторые экспрессивные единицы. К ним относятся лексические повторы (местоимение *каждый*), интенсификаторы смысла

⁹ Там же.

(наречие *максимально*), глаголы, характерные для разговорного стиля (*стращать*). Среди дискурсивных формул преобладают единицы негативной оценки, характерные для языка СМИ: *затягивать конфликт, стращать европейских обывателей войной, непомерная алчность, осознанный геноцид, политическая трусость*.

Представим пример англоязычного комментария с аналогичным вариантом сцепления реакции и продукции. Обозначим контекст обсуждения, процитировав лид медиапубликации: «A Texas newspaper issued a warning on Saturday, which is also Texas Independence Day, about the state seceding from the United States as a number of Republican candidates for the Texas House of Representatives and other offices are showing support for it» (В субботу, в День независимости Техаса, техасская газета опубликовала предупреждение о выходе штата из состава Соединенных Штатов, поскольку ряд кандидатов-республиканцев в Палату представителей штата Техас и другие учреждения выражают поддержку данному решению). В одном из комментариев читаем:

Succession¹⁰ is fine with me. However, I'll need compensation to move out of the state. This means that I'll need to be reimbursed for full market value for my house and moving expenses, plus \$100k for general living expenses¹¹.

Первая реплика, как и в предыдущем примере, служит вербализатором реакции адресата на публикацию. Присоединение средств, прагматическое значение которых соответствует интенции продукции, происходит с привлечением коннектива

¹⁰ Лексема содержит ошибку. Имеется в виду английское слово *secession* (отделение).

¹¹ Venegas N. Texas Newspaper Issues Warning About Secession (2024). URL: <https://www.msn.com/en-us/news/us/texas-newspaper-issues-warning-about-secession/ar-BB1jeAYp?cvid=a73f4918ee194218f15bdf1107a1298&ei=21#comments> (дата обращения: 26.04.2024).

however, в данном случае указывающего не на отношения противопоставления, а на отношения «общее – частное». Как читатель, субъект дискурса одобряет решение о выходе штата, а как автор – излагает свои условия. Об этом свидетельствуют модальность его дискурса – модальность необходимости. Комментатором привлекаются приемы синтаксического повтора *I'll need* и повтора предложных конструкций *for...*, *for...*. Эти приемы маркируют его персональный дискурс.

Сцепление с помощью вопросно-ответной конструкции. Названные случаи сцепления преимущественно представлены в текстах, где объектом реакции является единичная реплика собеседника. Реактивный зачин отражает когнитивную деятельность читателя, связанную с желанием осмыслить прочитанное, понять как содержание, так и цель публикации. Поэтому в комментариях читатель-автор склонен задавать вопросы риторического типа и сам на них отвечать. Для демонстрации средств реализации рецептивной интенции находим необходимым привести предваряющую интеракцию реплику собеседника, комментирующего публикацию о конфликте на Ближнем Востоке:

*Почему все игнорируют террористический акт на фестивале? Там людей прикладами автоматов забивали на смерть*¹².

На данную реплику опубликована следующая реакция:

Кто игнорирует? Все всё знают, знают, что теперь эта военищина хочет выставить это как оправдание геноциду. Но это не может быть оправданием геноциду, если ты претендуешь на то, что

*ты сам не дикарь. Во времена Шмуэля¹³ такое проходило, но сейчас XXI век, представления о морали сменились, теперь нельзя просто пойти и вырезать целый город, и чтоб потом тебя уважали за это*¹⁴.

Зачин текста содержит глагол *игнорировать*, послуживший стимулом для вербальной реакции адресата. Данная словоформа оказала перлокутивный эффект: она вызвала реакцию читателя на высказанную предыдущим автором альтернативную точку зрения. Вопросительная конструкция согласно иллюкутивной цели – это реплика, выражающая несогласие, что подтверждается последующими высказываниями комментатора, встающего на позицию автора. Поскольку большинство читателей в ленте к публикации «СМИ: Израиль поручил посольствам давить на дипломатов из-за иска ЮАР» принимают сторону обвинения *Израиля* в конфликте, один из читателей пытается противопоставить свое мнение большинству, инициировав серию реакций. Как видим, второй комментатор легитимирует мнение большинства. Он использует пропозицию с семантикой обобщения *Все всё знают*, которая еще раз доказывает, что он высказывается в поддержку мнений большинства. Переход от реакции к продукции выражен в форме вопросно-ответного комплекса. Далее автор конструирует обвинительный дискурс, в смысловой центр которого помещена идеологема *геноцид*. Создание образа мирового преступника производится через эмоционально-экспрессивные единицы, характерные для дискурса информационной войны *военищина, дикарь, вырезать город*. Для усиления Я-позиции и более четкого выражения авторства в продуктивной

¹² СМИ: Израиль поручил посольствам давить на дипломатов из-за иска ЮАР (2024). URL: <https://ria.ru/20240106/izrail-1919916287.html> (дата обращения: 26.04.2024).

¹³ Шмуэль, Бен-Арци (1914 – 2011) – израильский педагог и писатель, отец жены премьер-министра Израиля Бенджамина Нетаньяху.

¹⁴ Там же.

части текста комментатор приводит исторические факты, создавая контраст между *диким* обществом *во времена Шмуэля* и *21м веком*, символизирующим в данном контексте более высокий уровень развития человечества.

Сцепление с эмфазой. Реакция на стимул имеет аффективный характер. Ее вербализатором становится реплика, оформленная с помощью эмфатической конструкции. Затем коммуникативная роль читателя сменяется ролью продуцента, однако между данными компонентами комментария отсутствуют какие-либо вводные средства, маркирующие смену интенций.

*What a bunch of misinformation. The total annual increase for my \$1.7 million homeowners insurance, unberalla, 3 autos and earthquake insurance is \$200 or a little over \$18 a month*¹⁵.

Вступительное высказывание демонстрирует холистическую, аффективную реакцию читателя на ближний контекст. Далее автор приводит конкретные цифры, и тем самым продуцирует фактологический дискурс, контрастирующий с иррациональной реакцией, продемонстрированной им как читателем. При этом данные статистики, как и в предыдущих текстах, привлекаются для экспликации фрагментов частной жизни автора текста: *my \$1.7 million homeowners insurance*. Чтобы снивелировать контраст иррационального начала и рационального продолжения, автор комментария вместо корректного использования восклицательного знака в конце первой реплики ставит точку *What a bunch of misinformation(!)*. Таким образом, в данном комментарии сцепление построено на соположении текстовых средств, отражающих многоуровневое,

¹⁵ Carbonaro G. Florida Residents Decry Insurance 'Nightmare' as They Flee State (2023). URL: <https://www.msn.com/en-us/money/realestate/florida-residents-decry-insurance-nightmare-as-they-flee-state/ar-AA1kvdcy#comments> (дата обращения: 26.04.2024).

комплексное восприятие действительности: общее – частное, иррациональное – рациональное.

Сцепление с эффектом возврата. Реактивная часть текста представлена и в начале, и в конце комментария. Интенция реакции переходит в продуктивное порождение, а затем автор вновь возвращается к позиции читателя:

*Понимаете, в чём дело. Все эти мировые сдвиги меня интересуют только в одной плоскости: повышение благосостояния *всех* граждан – прежде всего, граждан РФ, конечно. Не олигархов с чиновниками и "артистами", не среднего или малого бизнеса, какого-нибудь "среднего класса", а именно всех без исключения. Включая тех, кого-таки интересуют цены на яйца (и кто уже забыл, когда последний раз ел говядину). А их немало, если вы откроете комментарии в той ветке, в которую направляете меня. Но "философствования" г-на Дугина этого не предполагают. У него там сплошь империи с императорами и благодарный "норот" у подножия трона*¹⁶.

Композиция данного комментария основана на рамочном принципе. Семантический аспект – идея, высказанная в начале, повторяется в конце, совпадает с прагматическим – актуализаторы реактивной интенции вводятся и в начале, и в конце текста. Реплика-обращение сменяется Я-контекстом (*меня интересуют*), маркирующим продуктивный характер порождения текста. Продуктивная часть текста имеет выраженную смысловую завершенность, так как автор обозначает свою самостоятельную позицию и конструкцией *меня интересует*, и предложной фразой *в одной плоскости*, которая акцентирует аспект рассмотрения вопроса. Повторение номинаций ключевого объекта

¹⁶ Дугин А. Убить дракона: мир стоит на пороге глобальной войны (2024). URL: https://ria.ru/20240117/globalnaya_voyna-1921723243.html (дата обращения: 26.04.2024).

комментария **всех* граждан, граждан РФ, всех без исключения* также вносит вклад в семантику равноправия в гражданском обществе, а также в прагматику самопрезентации. Однако в заключительной части мы видим возврат к реагированию и оценке семантико-прагматического плана исходной статьи. Таким образом, данный комментарий построен по принципу смены интенций: реакция – продукция – реакция.

Сцепление с элементами диффузии (переходный вариант). Представленный вариант сцепления отличается использованием полной синтаксической конструкции, характерной для публицистического дискурса. Новостные тексты, не имеющие указания на конкретного автора, реже становятся стимулом для реакций подобного типа. Тексты, построенные по данной модели, чаще публикуются в аналитических материалах. В данном случае комментарий написан к публикации Александра Дугина «Убить дракона: мир стоит на пороге глобальной войны»:

Я не согласен с уважаемым автором. За период после распада СССР идеи глобализма глубоко проникли в головы и "карманы" политических элит всех стран даже таких как Афганистан или Сомали. Философия и культура потребительского общества сегодня не имеет альтернативы. То, что мы видим сегодня есть ничто иное как борьба за лидерство в ГЛОБАЛИСТСКОМ МИРЕ. Сегодня под видом многополярности возникают различные "союзные образования" и их не 6-7 как указал автор их намного больше, просто все эти союзы являются временными и пока не очень устойчивые... Население планеты психологически не готово от отказа глобализма потребительства, а политические элиты так до конца и не понимают что по факту представляет из себя многополярный мир. Мир не созрел для многополярности...¹⁷.

Реактивная реплика читателя открывает интернет-комментарий. Второе высказывание вводит аргумент автора на предмет *идеи глобализма*, которая и становится смысловым центром его авторского текста. Развитие аргумента происходит с введением более широкого исторического и политического контекста. Наряду с диахроническим подходом автор несколько раз использует обстоятельство *сегодня*. Оно использована для акцента на результате. Далее по тексту автор снова обращается к источнику обсуждения и возвращается на позицию читателя, который счел нужным скорректировать данные из источника.

Приведенная ниже таблица (см. Таблицу 1) демонстрирует результаты количественного подсчета текстов, в которых реализуется рецептивно-продуктивная модель сцепления, где В 1 – сцепление с дискурсивными словами; В 2 – сцепление с вопросно-ответной конструкцией; В 3 – сцепление с эмфазой; В 4 – сцепление с эффектом возврата и В 5 – сцепление с элементами диффузии.

Как видим, сцепление как прагматический сценарий рецептивно-продуктивной деятельности НЛ чаще наблюдается в текстах интернет-комментариев, публикуемых на портале «РИА Новости». Преобладающим вариантом из пяти предложенных в данном исследовании является вариант В1 (сцепление с использованием дискурсивных слов), предусматривающий последовательную реализацию рецепции и продукции, и проявляющийся на синтаксическом уровне текста комментария. Данный вариант используется в 5 % текстов на русском языке, что значительно превышает другие значения. Согласно представленным данным, сцепление как эксплицитно выраженный переход от реакции на прочитанный медиатекст к репрезентации своей позиции более характерно для дискурса русскоязычной НД.

¹⁷ Там же.

Таблица 1. Сцепление как вариант рецептивно-продуктивной деятельности непрофессиональной личности

Table 1. Cohesion as a type of receptive and productive activity of an unprofessional personality

Варианты сцепления	Кол-во, ед.	% от общего количества текстов	Кол-во, ед.	% от общего кол-ва текстов
Общее количество текстов	На английском языке – 2579 ед.		На русском языке – 1357 ед.	
В 1	39	1.5 %	68	5 %
В 2	33	1.3 %	38	2.8 %
В 3	51	2 %	42	3 %
В 4	25	1 %	18	1.3 %
В 5	20	0.8 %	16	1.2 %
Всего:	168	6.5 %	182	13.4 %

Диффузия. Семантико-прагматические отношения в тексте отражают взаимопроникновение средств реализации реакции и продукции. Переход от рецепции к продукции на уровне синтактики отсутствует. Этот текстотип эксплицирует горизонтально, но нелинейно сконструированный тип отношений. К данной группе относятся тексты, в которых переход от реакции к продукции не столь очевиден, размыт, и на протяжении всего текста видны вкрапления средств, сигнализирующих о коммуникативном статусе читателя. Актуализаторы реакции и продукции тесно переплетены в текстовом «теле». Как правило, в таких комментариях представлены общие референты, а также сеть анафорических средств, вкраплений, маркирующих связь единичного комментария с дискурсом данной ленты и текстом-стимулом.

Диффузия с дейксисом. Текст содержит прагматические единицы, которые соотносят его с коммуникативной ситуацией обсуждения медиапубликации, а также события, находящегося в центре обсуждения. Дейктические элементы текста не позволяют наблюдателю установить его полную автономность. Наоборот, они соотносят комментарий с

ситуацией обсуждения и с публикацией как стимулом для его появления.

*We already agree that abortion is **no one else's** business except your own. Now we need to agree that the home you live in is **nobody else's** business except your own. My house, my money, my life, my choice and **nobody else's** business. Then you could live in a small, affordable home that is much easier to replace. If something happened to my singlewife I could just live in a travel trailer for a while because we ended the HOA **out here** about 55 miles NE of Cincinnati. I took several years and going to court many times. The whole country needs their housing freedom restored. We'd have a better country. Freedom leads to prosperity. Freedom does not mean the freedom to take away **someone else's** freedom¹⁸.*

Дейктические элементы представленного комментария соотносят данный текст не только с общей политико-экономической ситуацией, которая происходит в момент обсуждения условий жизни граждан США, но также и с

¹⁸ Carbonaro G. California's Insurance Crisis Is Spiraling Out of Control (2024). URL: <https://www.msn.com/en-us/money/markets/california-s-insurance-crisis-is-spiraling-out-of-control/ar-BB1jaHoU?ocid=hpmsn&cvid=b36a6b673e5343da9f3fa1265a058eab&ei=11#comments> (дата обращения: 26.04.2024).

правовой ситуацией, о чем свидетельствуют упоминание в первом высказывании сложного с позиции частного права понятия *abortion*. С одной стороны, автор с самого начала независимо мыслит и практически не опирается на исходную новость. Однако постановка наречия *уже* способствует отсылке к обсуждаемой ситуации, оно служит дейктическим центром текста. Обратим внимание, что продуктивная часть текста содержит четыре случая притяжательной конструкции *nobody else's*, которая также служит дейксисом, помогая автору отмежеваться от тех, кто посягает на его права. С помощью этой местоименной конструкции читатель осуществляет обособление себя, выражает свою индивидуалистскую позицию, высказывает желание оградить свое право на частную жизнь. Дискурсивная конструкция с однородными членами предложения *my choice and nobody else's business*, передающая прагматику обособленности, сопровождается указанием на конкретный локус – американский штат Огайо. Ссылка на локацию обозначается дейксисом *out here*, а конкретизируется обстоятельственным оборотом *about 55 miles NE of Cincinnati*.

Диффузия с общими референтами.

Наличие в тексте референтов из стимула также позволяет говорить о диффузии текстовых средств реализации реакции и продукции. Через многократные повторы лексических единиц, означающих ключевых референтов новости, комментаторы показывают, что они в первую очередь адресаты, и только во вторую – авторы:

Insurance companies left Louisiana in 2022. There is the state and a couple of companies now that cover homeowner's insurance. Car insurance is high also. What about Florida? They are experiencing the same problem. I knew it was going to happen to California's wildfire and flood zones. And it

*won't just be California. What about the people in tornado alley?*¹⁹.

Автор интернет-комментария включается в контекст обсуждения страховых компаний на правах адресата и многократно использует ключевое понятие *insurance*. Разноаспектность рассмотрения проблемы страхового бизнеса, обозначенная оборотами с общим словом *homeowner's insurance* и *car insurance*, помимо прочего, может быть проинтерпретирована и как позиция автора. Авторская интенция выразить мнение не простого обывателя, но знатока реального положения дел прослеживается на синтаксическом уровне текста, представляющего определенную последовательность полных предложений, построенных по принципу соблюдения прямого порядка слов. Однако они чередуются с вопросительными конструкциями неполного, эллиптического типа: *What about Florida? What about the people in tornado alley?* Выбор автором сочетания нейтральных и экспрессивных высказываний указывает на стратегию самопрезентации, которая также эксплицируется в изложении ситуации в Калифорнии от первого лица: *I knew it was going to happen to California's wildfire*.

Диффузия с анафорой.

Анафорические элементы в текстах адресатов медиадискурса помогают им сфокусироваться на обсуждении ключевой новости. Некоторые тексты изобилуют указательными местоимениями, которые, по-видимому, позволяют комментаторам показать свою принадлежность к сообществу и поддержать мнение единомышленников. Следующий текст служит примером сказанному:

This is happening to many other states, red and blue. Making this all about CA is just a political stunt and a way to divide the country and not actually address the issue. How about a federal law that stares Insurance is just that. It should take in the

¹⁹ Там же.

*risks. Make companies have to operate in ALL states if they want to stay in business*²⁰.

Введение в текст комментария анафорических элементов также указывает на связь текста с первичным материалом. В цитируемом примере такая связь присутствует в словосочетаниях указательных местоимений с обобщающими словами и дискурсивными частицами *this all, just that, ALL*. Реактивное начало комментария перекликается с его анафорическими компонентами и создает эффект регулярной отсылки автора к своему статусу читателя. Как автор, комментатор экстраполирует обсуждаемую проблему на другие объекты, такие как *other states, red and blue*. Перенос с частного на общее маркируется обобщающим выражением *ALL states*, в котором капитализация обозначает фокусный компонент смыслового целого и подчеркивает точку зрения автора текста. Однако неоднократное возвращение автора к уже

названным пунктам и вопросам с помощью вышеперечисленных единиц образует эффект диффузного сочетания реакции и продукции.

Приведенная ниже таблица (см. Таблицу 2) демонстрирует результаты количественного подсчета текстов, в которых реализуется рецептивно-продуктивная модель диффузии, где В 1 – диффузия с дейксисом; В 2 – диффузия с общими референтами и В 3 – диффузия с анафорой.

Приведенные количественные данные показывают, что в процентном соотношении порождение текстов, сочетающих реакцию и продукцию, по модели «диффузия» примерно одинаковое в дискурсах российского и американского комментирующего сообщества. В текстах на русском языке показатели немного выше, что дает более высокий средний результат – 8 % русскоязычных текстов по сравнению с 7.3 % англоязычных комментариев.

Таблица 2. Диффузия как вариант рецептивно-продуктивной деятельности непрофессиональной личности

Table 2. Diffusion as a type of receptive and productive activity of an unprofessional personality

Варианты сцепления	Кол-во, ед.	% от общего кол-ва текстов	Кол-во, ед.	% от общего кол-ва текстов
Общее количество текстов	На английском языке – 2579 ед.		На русском языке – 1357 ед.	
В 1	79	3 %	44	3.2 %
В 2	69	2.7 %	43	3.2 %
В 3	40	1.6 %	22	1.6 %
Всего:	188	7.3 %	109	8 %

²⁰ Там же.

Наложение. Семантико-прагматический план текста отражает одновременно и обратную связь как ведущий способ реализации рецепции, и презентацию Я-позиции как способ реализации продукции. Отношения читателя и автора нелинейны, имплицитны. Тип отношений по принципу наложения прослеживается в текстах, в которых реакция не маркирована в начале или в конце. Можно утверждать, что средства, актуализирующие интенции реакции и продукции, совпадают. Продукция и есть реакция, но она требует привлечения дополнительных фактов и мнений. Такие тексты существуют как автономные единицы, они синтаксически и семантически самостоятельны. На связь с сообществом указывает референциальная общность и прагматический план, состоящий в намерении автора оценить новость и внести свой вклад в дискуссию.

Наложение с привлечением политического контекста. Продуктивная интенция реализуется за счет репрезентации в комментариях фрагмента обыденного политического сознания авторов. Как правило, модус таких интернет-комментариев характеризуется повышенной эмоциональностью. Эмоциональная палитра в текстах разнообразна, она включает опасения, сожаления, сарказм, сочувствие, разочарование и т.д. Так, обсуждению подлежат человеческие качества политических лидеров, они рассматриваются авторами политических комментариев как контекстуально значимые:

Сонный Джо это то, что надо Америке! – чтобы проспать начало третьей мировой войны! России надо активнее работать с адекватными западными политиками и гражданами- и тогда множество мелких потрясений и

конфликтов не перерастут в глобальную ядерную войну...²¹.

В представленном комментарии есть самостоятельный замысел и топик: президент США в аспекте задач России в текущей политической обстановке. Акцент ставится на одном из качеств лидера, которое систематически высвечивается прессой. Отсутствуют явные отсылки к исходному тексту статьи. Семантика текста конструируется с помощью политической лексики военной тематики: *третья мировая война, ядерная война, потрясения, конфликты*. Прагматический уровень определяется склонностью автора к использованию директивов *надо Америке, России надо активнее работать*. Данные модальные фразы выдают более высокую коммуникативную позицию читателя и автора-«эксперта».

Наложение с привлечением персонального контекста. Дискурс НЛ о политике порождается согласно стратегиям обратной связи и самопрезентации, поэтому политика в данной дискурсивной практике – это один из поводов поговорить о собственных проблемах и реализовать потребность в самопрезентации. Выражение НЛ субъективной позиции часто сопровождается примерами из повседневной жизни граждан.

Оформление следующего комментария как личностного нарратива показывает, что автор описывает фрагмент своей жизни как пример, подходящий в качестве аргумента в дискуссии о кризисе страхования. Текст комментария одновременно есть и реакция, и продукция. Однако это фрагмент не из политического знания, а из внешней действительности – мира реальных субъектов и объектов:

Have a friend moving to Florida in the near future. He's keeping his condo up north and only plans to live in an apartment with

²¹ Дугин А. Убить дракона: мир стоит на пороге глобальной войны (2024). URL: https://ria.ru/20240117/globalnaya_voyna-1921723243.html (дата обращения: 26.04.2024).

*the bare minimum minus any insurance instead of buying a house because of the insurance crisis and hurricane potential*²².

Ссылка автора комментария на эпизод из жизни близкого человека (*Have a friend*) дает понять, что он извлекает из памяти случай, который соотносится с темой обсуждения. Однако эксплицитных реактивных выражений комментатор не использует. Он формулирует отдельный фрагмент из жизни приятеля и тем самым проявляет одновременно двойной статус. Рецепция способствовала актуализации конкретной ситуации из жизни. Поэтому комментарий – это частный пример, который, с точки зрения автора комментария, подтверждает наличие проблемы – кризиса в области страхования недвижимости в США (*minus any insurance, the insurance crisis*).

Наложение с привлечением широкого исторического/экзистенциального контекста. Проблема политической жизни общества, несомненно, в сознании НЛ сопрягается с вопросами существования человека как вида, с поиском закономерностей в его эволюции. Поэтому многие комментаторы вписывают конкретный факт в более глобальный контекст, анализируя витки развития человеческой цивилизации и соотнося их с текущим моментом. Необходимость упорядоченного образа жизни устанавливается автором и в следующем тексте:

Everyone needs to follow the rules, people, and nations. Without them, there would be chaos. People have been oppressed for so long in Palestine what do you think could happen, they would revolt somehow. After more than 50 years of talking and talking and no resolution, the Security

²² Carbonaro G. Florida Residents Decry Insurance 'Nightmare' as They Flee State (2023). URL: <https://www.msn.com/en-us/money/realestate/florida-residents-decry-insurance-nightmare-as-they-flee-state/ar-AA1kvdcy#comments> (дата обращения: 26.04.2024).

*Council needs to abolish permanent members*²³.

Комментарий открывается прескриптивной модальностью, указывая на морализаторский тон автора: *needs to follow the rules*. Однако во второй части прескрипция сменяется мнением в защиту действий одного из малых народов Ближнего Востока, которое на синтаксическом уровне выражается условными конструкциями с грамматическим значением возможного альтернативного положения вещей *there would be chaos, could happen, would revolt*. Третье высказывание снова демонстрирует иллюкцию рекомендации и прескрипции.

Наложение с привлечением медийного контекста. Данный вариант модели «наложение» отражает заинтересованность и проактивный характер непрофессиональной личности, которая следит за медиаверсиями событий. Например, при комментировании на американских сайтах авторы традиционно ссылаются на медиа, как на основной источник и объективного политического контента, и манипулятивного. Сегодня оценочное отношение к медиа как к «четвертой власти» отражается во многих откликах. Проиллюстрируем привлечение НЛ медийного контекста на примере развернутого мнения на публикацию «Texas Newspaper Issues Warning About Secession» (Техасская газета опубликовала предупреждение о возможном отделении). Поскольку объем цитируемого материала высокий, мы приводим его в сокращенном виде:

*A few things that I noted from the editorial and the TNM's response*²⁴. 1) In

²³ Kika T. Democrats Issue Warning to Joe Biden (2024). URL: <https://www.msn.com/en-us/news/world/democrats-issue-warning-to-joe-biden/ar-BB1kpSQL?ocid=hpmsn&cvid=4ce8b0e674a54f2a85d4c0720dc2f697&ei=76#comments> (дата обращения: 26.04.2024).

²⁴ TNM – Texas Nationalist Movement or Texas independence movement (Textit) – организация,

mentioning the loss of 30% of their state budget, they probably should have mentioned how big a cut individuals would 'save' in not paying federal taxes. Then they should have compared that to how large an increase in state taxes there would be. People are already complaining about increases in property taxes and that Abbott's cut in property taxes has done little to nothing to alleviate the problem.

2) *I can't say the editorial 'neglects a core issue'. All they are pointing out is that there will be fiscal ramifications that apparently haven't been discussed even in generalities, let alone any detail. by the secessionists.*

3) *The TNM mentions Texas is not MERELY a participant in the federal system out of financial necessity but... contributed significantly to unions prosperity. <...>*

8) *"The suggestion that Texas could not navigate the complexities of independence..." But you need to start figuring out those complexities NOW. If you're serious about presenting secession to the voters you better be able to give detailed explanations of all the effects, not just platitudes about self-governance, frustration with federal overreach, etc.; but the REAL WORLD implications²⁵.*

Выделенные нами в тексте единицы отражают связь представленного текста с медийным контекстом. Автор ставит своей задачей подробно проанализировать обсуждаемый вопрос. Он показывает, что пристально следит за публикациями редактора и деятельностью организации TNM, и теперь в состоянии сформировать цельную картину политики региональной газеты в отношении вопроса выхода Техаса из состава Соединенных Штатов. Самостоятельность и авторская позиция – это эксплицитная структура текста,

выступающая за выход штата Техас из состава Соединенных Штатов Америки. Источник: URL: https://en.wikipedia.org/wiki/Texas_secession_movements (дата обращения: 24.04.2024).

²⁵ Venegas N. Texas Newspaper Issues Warning About Secession (2024). URL: <https://www.msn.com/en-us/news/us/texas-newspaper-issues-warning-about-secession/ar-BB1jeAYp?cvid=a73f4918ee194218f15bdf1107a1298&ei=21#comments> (дата обращения: 26.04.2024).

списочный принцип синтаксического оформления помогает систематизировать мысли и облегчает восприятие текста адресатом. Семантика текста сопряжена с оценкой и анализом вопроса отделения штата, а также с конструированием медийной реальности посредством регулярных публикаций по теме. Прагматика преимущественно состоит в выражении автором персонального отношения. Он вскрывает манипулятивный характер деятельности медиа и апеллирует к тому, чтобы читатели его текста, другие комментаторы, смотрели между строк и видели «реальный мир», а не тот, который создается в СМИ. Любопытно, что изложение от третьего лица в начале сменяется обращением с использованием второго лица в последнем пункте. Однако директивная иллокуция выдерживается. От высказываний с семантикой упрека (*they should have mentioned, they should have compared*) она смягчается до менее категоричной с семантикой совета (*you need, you better be able*). Это отражает текущую тенденцию, наблюдаемую в непрофессиональном дискурсе (гражданской журналистике). НЛ сегодня все более уверенно позиционирует себя как легитимирующее лицо, получившее право проводить аудит политической и медийной реальности. Ее метакоммуникативная характеристика как «медийного менеджера» может стать перспективной для исследований непрофессионального интернет-дискурса.

Результаты количественного подсчета описываемых типов текстов, построенных по модели наложения (см. Таблицу 3) показывают частотность использования вариантов реализации комментаторами двойной коммуникативной роли – читателя и автора – В 1 – наложение с привлечением политического контекста; В 2 – наложение с привлечением персонального контекста; В 3 – наложение с привлечением широкого исторического/экзистенциального контекста; В 4 – наложение с привлечением медийного контекста.

Таблица 3. Наложение как тип рецептивно-продуктивной деятельности непрофессиональной личности

Table 3. Overlapping as a type of receptive and productive activity of an unprofessional personality

Варианты сцепления	Кол-во, ед.	% от общего количества текстов	Кол-во, ед.	% от общего кол-ва текстов
Общее количество текстов	На английском языке – 2579 ед.		На русском языке – 1357 ед.	
В 1	50	1.9 %	30	2.2 %
В 2	48	1.9 %	17	1.3 %
В 3	32	1.2 %	20	1.5 %
В 4	18	0.7 %	14	1 %
Всего:	148	5.7 %	81	6 %

Исходя из представленных количественных данных, можно сказать, что процент автономных текстов достаточно скромен по сравнению с предыдущими вариантами. Соотношение моделей указывает на наибольшую распространенность первого и второго

вариантов в англоязычных комментариях, в то время как русскоязычных более распространены варианты 1 и 3.

Все полученные данные были сведены в общую таблицу (см. Таблицу 4) и представлены в количественном и процентном виде.

Таблица 4. Количественное соотношение вариантов рецептивно-продуктивной деятельности НЛ на русском и английском языках

Table 4. Quantitative ratio of variants of receptive and productive activity of unprofessional personality in Russian and English

	Сцепление		Диффузия		Наложение		Всего	
	Кол-во, ед.	%	Кол-во, ед.	%	Кол-во, ед.	%	Кол-во, ед.	%
Комментарии на «MSN-US» (2579)	168	6.5 %	188	7.3 %	148	5.7 %	504	19.5 %
Комментарии на «РИА Новости» (1357)	182	13.4 %	109	8 %	81	6 %	372	27.4 %

Судя по полученным результатам, показатели количественного соотношения демонстрируют умеренную разницу в производстве непрофессионального дискурса согласно описанным моделям. Как видно из таблицы, 19.5 % пользователей в англоязычном дискурсе и 27.4 % комментаторов в русскоязычном не

только реагируют на публикацию короткой репликой-откликом, но и реализуют себя в статусе непрофессионального автора политического дискурса. Соотношение рецептивно-продуктивного типа прагматических моделей на русском и английском языках показывает относительно небольшой разрыв в

значениях, что говорит о сохранении тенденции к реагированию в форме коротких, односложных реплик, указывающих на рецептивную интенцию как доминирующую при порождении непрофессионального интернет-дискурса. Тем не менее, от одной пятой до одной четверти реципиентов (19.5 % и 27.4 %) стремятся обозначить себя и свою политическую картину мира в непрофессиональном дискурсе, выполняя роль продуцента. На фоне общего формально зафиксированного количества текстовых единиц это значимый показатель.

Заключение (CONCLUSIONS)

Рассмотрение сочетания в одном тексте двух коммуникативных ролей, выполняемых современным читателем новостей в СМИ, в лингвопрагматическом ключе дает интересные результаты. Прежде всего, читательская функция оценки сохраняется далеко не во всех комментариях. Оценка все чаще переходит в имплицитные слои текста-отклика, растекаясь по «телу» текста и принимая характер эмотивно-аналитического текстопроизводства. Свобода выбора формально-содержательной организации текстового целого, проявляющаяся в непрофессиональном дискурсе, обуславливает его широкий вариативный потенциал как с позиции синтаксической организации, так и в семантико-прагматическом аспекте. Однако при всем разнообразии лингвистических средств актуализации результатов рецепции и продукции в комментариях выделяются типизированные прагматические сценарии. Во-первых, это сочетание в одном тексте лаконичных реплик, имитирующих устно-диалогический дискурс, и продуктивных речевых средств, характерных для публицистического дискурса. Во-вторых, выделяются линейный и нелинейный типы сочетания реакции и продукции. В-третьих, если реакция как правило семантически соотносится с оценкой формально-содержательного плана

исходного текста и откликов (стратегия обратной связи), то продукция чаще всего связана с прагматикой представления себя и своей картины мира. Тенденция к самопрезентации, наблюдаемая в комментариях, приближает непрофессиональный интернет-дискурс к жанрам, в которых самопрезентация – это ведущая стратегия, детерминирующая текстообразование. Это, в частности, жанры блога, дневника и т.п., где главными мотивами к порождению текста становятся самопрезентация и представление личного опыта.

Типизирование репертуара лингвистических средств, актуализирующих реакцию и продукцию на уровне синтактики, семантики и прагматики непрофессионального дискурса дает возможность предложить три основных типа сочетаемости перехода статуса читателя в статус автора: сцепление, диффузия и наложение. Наиболее эксплицитно сочетание двух ипостасей нового читателя проявляется в текстах первого типа. Второй тип представляет собой промежуточный вариант реализации амбивалентной позиции автора в жанре интернет-комментария. Наименее эксплицитно и нелинейно амбивалентный характер производства непрофессионального дискурса представлен в текстах третьего типа. В комментариях с прагматическим вариантом наложения роли читателя и роли автора наилучшим образом отражается двойственная природа текстовой деятельности личности. Если пользоваться словами М. М. Бахтина, языковая личность «никогда не совпадает сама с собой». Ее дискурсы – это непрерывное сочетание реактивно-продуктивной речемыслительной деятельности, приводящее к конструированию особого ментального мира, который, хотя и проявляет коллективные признаки дискурсивной деятельности, всегда индивидуален, уникален и неповторим.

В перспективе исследования – дальнейшее изучение прагматики сетевых интернет-комментариев на предмет их интенциональности с применением экспериментальных методов, с учетом динамики в реализации интернет-комментаторами коммуникативных интенций. Данные, полученные в результате исследования прагматической структуры непрофессионального дискурса, могут быть востребованы специалистами в сфере политической и медийной коммуникации.

Список литературы

- Арутюнова Н. Д. Фактор адресата // Известия Академии наук СССР. Серия литературы и языка. М.: Наука, 1981. Т. 40. № 4. С. 356–367.
- Барт Р. Смерть автора // Барт Р. Избранные работы: Семиотика. Поэтика. М., 1994. С. 384–391.
- Болотнов А. В. Идиостиль информационно-медийной языковой личности: коммуникативно-когнитивные аспекты исследования: дисс. ... д-ра филол. н. Томск, 2015. 405 с.
- Виноградов В. В. Проблема авторства и теория стилей. М.: Государственное издательство художественной литературы, 1961. 615 с.
- Винокур Т. Г. Говорящий и слушающий. Варианты речевого поведения. М.: Наука, 1993. 172 с.
- Дейк Т. ван. Дискурс и власть: Репрезентация доминирования в языке и коммуникации. Пер. с англ. Е. А. Кожемякина, Е. В. Переверзева, А. М. Амадова. М.: Книжный дом «ЛИБРОКОМ», 2013. 344 с.
- Дускаева Л. Р. Интенциональность речевой деятельности журналиста: онтология и структура // Вестник Санкт-Петербургского университета. Серия 9. Филология. Востоковедение. Журналистика. 2012. № 2. С. 253–260.
- Иванова И. Векторная вариативность и характер реакций читателей на сообщения в онлайн-медиа // Медиалингвистика. 2023. № 10 (1). С. 27–46. DOI: 10.21638/spbu22.2023.102
- Каминская Т. Л. Автор и адресат в современных медиатекстах // Вестник Санкт-Петербургского университета. Серия 9. Филология. Востоковедение. Журналистика. 2008. № 2–2. С. 314–319.
- Ким Л. Г., Голев Н. Д. Об отношениях адресата, автора и текста в парадигме лингвистического интерпретационизма // Сибирский филологический журнал. 2008. № 1. С. 144–153.
- Ким Л. Г. Единицы описания интерпретационного функционирования текста в пространстве адресата // Вестник Кемеровского государственного университета. 2012. № 1 (49). С. 186–190.
- Клушина Н. И. Интенциональные категории публицистического текста (на материале периодических изданий 2000–2008 гг.): Автореф. дис. ... д-ра филол. наук. Москва, 2008. 57 с.
- Ковалевская Н. И. Цифровое чтение: формирование нового типа читателя // Труды БГТУ. Серия 4: Принт- и медиатехнологии. 2022. № 1 (255). С. 159–165. DOI: 10.52065/2520-6729-2022-255-1-159-165
- Кристева Ю. Избранные труды: Разрушение поэтики / Пер. с франц. Г. К. Косикова, Б. П. Нарумова. М.: «Российская политическая энциклопедия» (РОССПЭН), 2004. 656 с.
- Лукина М. М. Новая жизнь старых газет: как пресса приспосабливается к новому читателю // Вестник Московского университета. Серия 10: Журналистика. 2006. № 3. С. 52–59.
- Медиатекст как полиинтенциональная система: сб. статей / отв. ред. Л. Р. Дускаева, Н. С. Цветова. СПб.: С.-Петербург. гос. ун-т, 2012. 250 с.
- Мельник Н. В., Савельева И. В. Лингвоперсонологические стратегии восприятия текста (на материале интернет-комментариев к политическим статьям) // Вестник Кемеровского государственного университета. 2017. № 4 (72). С. 197–204. DOI: 10.21603/2078-8975-2017-4-197-204
- Михалева О. Л. Политический дискурс. Специфика манипулятивного воздействия: монография. М.: Либроком, 2008. 256 с.
- Проблема легитимизации в политическом дискурсе: лингвоперсонологический аспект: коллективная монография. Под ред. Мельник Н. В. М.: Издательство «Перо», 2021. 176 с.

Пром Н. А. Типы адресата в медиадискурсе // Медиалингвистика. 2020. Т. 7 (1). № 1. С. 95-103. DOI: 10.21638/spbu22.2020.108

Путина О. Н., Балакин С. В. Речевая коммуникация: прагматический аспект // Евразийский гуманитарный журнал. 2022. №1. С. 4–9.

Романтовский А. В. Нарративизация в жанре интернет-комментария при общении в социальных сетях // Коммуникативные исследования. 2024. Т. 11. № 1. С. 155–170. DOI: 10.24147/2413-6182.2024.11(1).155-170

Савельева И. В. Механизмы легитимизации в медиадискурсе (на материале интернет-СМИ) // Вестник Новосибирского государственного университета. Серия: История, филология. 2019. Т. 18. № 6. С. 188–198. DOI: 10.25205/1818-7919-2019-18-6-188-198

Савельева И. В. Непрофессиональный политический дискурс: лингвопрагматический и лингвоперсонологический аспекты: монография. Санкт-Петербург: Издательство «Наукоемкие технологии», 2021. 139 с.

Цветова Н. С. Категория автора в интенциональном поле медиатекста // Медиатекст как полиинтенциональная система: сб. статей / отв. ред. Л. Р. Дускаева, Н. С. Цветова. СПб., 2012. С. 17–23.

Чепкина Э. В. Русский журналистский дискурс: текстопорождающие практики и коды (1995–2000): монография. Екатеринбург, 2000. 279 с.

Эко У. От Интернета к Гутенбергу: текст и гипертекст. Отрывки из публичной лекции Умберто Эко на экономическом факультете МГУ 20 мая 1998. URL: <https://studfile.net/preview/1714056/> (дата обращения: 12.03.2024).

Сар Р. Towards the proximization model of the analysis of legitimization in political discourse // Journal of Pragmatics. 2008. № 40. Pp. 17–41. DOI: 10.1016/j.pragma.2007.10.002

Creeber G., Royston M. Digital Culture: Understanding New Media. McGraw-Hill Education (UK), 2008. 224 p.

Domingo D. Participatory journalism practices in the media and beyond: an international comparative study of initiatives in online newspapers / Domingo D., Quandt T., Heinonen A., Paulussen S., Singer J. B., Vujnovic M. // Journalism Practice. 2008.

№ 2 (3). Pp. 326–342. DOI: 10.1080/17512780802281065

Fetzer A., Lauerbach G. E. Political Discourse in the Media: Cross-cultural perspectives. Amsterdam: John Benjamins Publishing Company, 2007. 393 p. DOI: 10.1075/PBNS.160

McCombs M. E. A look at agenda-setting: Past, present and future // Journalism Studies. 2005. № 6. P. 543–557. DOI: 10.1080/14616700500250438

Tsfati Y. Uses and perceptions of political media // Political Communication: Handbooks of Communication Science / Reinemann C. (Ed.). V. 18. Berlin: de Gruyter, 2014. Pp. 489–506. DOI: 10.1515/9783110238174.489

Verschueren J. Understanding pragmatics. London: Edward Arnold / New York: Oxford University Press, 1999. 309 p.

Wall M. Mapping Citizen and Participatory Journalism: In newsrooms, classrooms and beyond // Journalism Practice. 2017. № 11 (2–3). Pp. 134–141. DOI: 10.1080/17512786.2016.1245890

Wettstein M., Wirth, W. Media Effects: How Media Influence Voters // Swiss Political Science Review. 2017. № 23 (3). Pp. 262–269. DOI: 10.1111/spsr.12263

Wojcieszak M., Mutz D. C. Online Groups and Political Discourse: Do Online Discussion Spaces Facilitate Exposure to Political Disagreement? // Journal of Communication. 2009. № 59 (1). Pp. 40–56. DOI: 10.1111/j.1460-2466.2008.01403.x

References

Arutyunova, N. D. (1981). Addressee factor, *Izvestiya Akademii nauk SSSR. Seriya literatury i yazyka* [Proceedings of the USSR Academy of Sciences. Literature and Language Series], Nauka, Moscow, Russia, 40 (4), 356–367. (*In Russian*)

Bart, R. (1994). The author's death, in *Izbrannye raboty: Semiotika. Poetika* [Selected papers. Semiotics. Poetics], Moscow, Russia, 384–391. (*In Russian*)

Bolotnov, A. V. (2015). *The idiosyncrasy of an information-media linguistic personality: communicative and cognitive aspects of research*, D. Sc. Thesis, Tomsk, Russia, Tomsk State University. (*In Russian*)

Vinogradov, V. V. (1961). *Problema avtorstva i teoriya stilei* [The problem of

authorship and the theory of styles], Gosudarstvennoe izdatelstvo khudozhestvennoy literatury, Moscow, Russia. (In Russian)

Vinokur, T. G. (1993). *Govoryashchiy i slushayushchiy. Varianty rechevogo povedeniya* [The speaker and the listener. Variants of speech behavior], Nauka, Moscow, Russia. (In Russian)

Dijk, T. van. (2013). *Diskurs i vlast: Rerezentatsiya dominirovaniya v yazyke i kommunikatsii* [Discourse and power], transl. by Kozhemyakin E. A., Pereverzev E. V. and Amatov A. M., Publishing House «LIBROKOM», Moscow, Russia. (In Russian)

Duskaeva, L. R. (2012). The intentionality of a journalist's speech activity: ontology and structure, *Saint Petersburg University Bulletin, Series 9, Philology, Oriental Studies, Journalism*, 2, 253–260. (In Russian)

Ivanova, I. (2023). Vector variability and the nature of readers' reactions to messages in online media, *Medialingvistika*, 10 (1), 27–46. DOI: 10.21638/spbu22.2023.102 (In Russian)

Kaminskaya, T. L. (2008). Author and addressee in modern media texts, *Saint Petersburg University Bulletin, Series 9, Philology, Oriental Studies, Journalism*, 2 (2), 314–319. (In Russian)

Kim, L. G. and Golev, N. D. (2008). On the relationship of the addressee, the author and the text in the paradigm of linguistic interpretationism, *Sibirskiy filologicheskii zhurnal*, 1, 144–153. (In Russian)

Kim, L. G. (2012). Units of description of the interpretative functioning of the text in the addressee's space // *Kemerovo State University Bulletin*, 1 (49), 186–190. (In Russian)

Klushina, N. I. (2008). *Intentional categories of journalistic text (based on the material of periodicals 2000–2008)*, Abstract of D. Sc. Dissertation, Journalism, Lomonosov Moscow State University, Moscow, Russia. (In Russian)

Kovalevskaya, N. I. (2022). Digital reading: the formation of a new type of reader, *Trudy BGTU, Seriya 4, Print- i mediatekhnologii*, 1 (255), 159–165. DOI: 10.52065/2520-6729-2022-255-1-159-165 (In Russian)

Kristeva, Yu. (2004). *Izbrannye trudy: Razrushenie poetiki* [Selected works: The Destruction of Poetics], G. K. Kosikov (Ed.), transl. by Kosikov, G. K. and Narumov B. P., Rossiyskaya politicheskaya entsiklopediya (ROSSPEN), Moscow, Russia. (In Russian)

Lukina, M. M. (2006). The new life of old newspapers: how the press adapts to a new reader, *Moscow University Bulletin, Series 10, Journalism*, 3, 52–59. (In Russian)

Duskaeva, L. R. and Tsvetova, N. S. (Eds.) (2012). *Mediatekst kak poliintencionalnaya sistema* [Media text as a multi-functional system], the collection of papers, Saint Petersburg State University, Saint Petersburg, Russia. (In Russian)

Melnik, N. V. and Saveleva, I. V. (2017). Linguopersonological strategies of text perception (based on the material of online comments on political articles), *Bulletin of Kemerovo State University*, 4 (72), 197–204. DOI: 10.21603/2078-8975-2017-4-197-204 (In Russian)

Melnik, N. V. (Ed.) (2021). *Problema legitimizatsii v politicheskom diskurse: lingvopersonologicheskii aspekt* [The problem of legitimization in political discourse: linguistic and phonological aspect], Izdatelstvo «Pero», Moscow, Russia. (In Russian)

Mikhaleva, O. L. (2008). *Politicheskii diskurs. Spetsifika manipulyativnogo vozdeistviya* [Political discourse. The specifics of manipulative influence], Librokom, Moscow, Russia. (In Russian)

Prom, N. A. (2020). The types of addressees in media discourse, *Medialingvistika*, 7 (1), 1, 95–103. DOI: 10.21638/spbu22.2020.108 (In Russian)

Putina O. N., Balakin S. V. (2022). Speech communication: pragmatic aspect, *Eurasian Humanitarian Journal*, 1, 4–9. (In Russian)

Romantovskiy, A. V. (2024). Narrativization in the genre of Internet commentary when communicating on social networks, *Kommunikativnye issledovaniya*, 11 (1), 155–170. DOI: 10.24147/2413-6182.2024.11(1).155–170. (In Russian)

Savelieva, I. V. (2019). Mechanisms of legitimization in the media discourse (a case study of the Internet media), *Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Istoriya, filologiya*, 18 (6), 188–198. DOI: 10.25205/1818-7919-2019-18-6-188-198 (In Russian)

Savelieva, I. V. (2021). *Neprofessionalny politicheskii diskurs: lingvopragmaticheskii i lingvopersonologicheskii aspekty* [Non-professional political discourse: linguopragmatic and linguopersonological aspects], Naukoemkie tekhnologii, Sankt-Peterburg, Russia, 139 p. (In Russian)

Tsvetova, N. S. (2012). The author's category in the intentional field of the media text, in Duskaeva, L. R. and Tsvetova, N. S. (Eds.) *Mediatekst kak poliintentsionalnaya sistema* [Media text as a multi-functional system], Saint Petersburg State University, Saint Petersburg, Russia, 17–23. (In Russian)

Chepkina, E. V. (2000). *Russkiy zhurnalistskiy diskurs: tekstoporozhdayushchie praktiki i kody (1995–2000)* [Russian Journalistic Discourse: Text-generating Practices and Codes (1995–2000)], Yekaterinburg, Russia. (In Russian)

Eco, U. (1998). From the Internet to Gutenberg: Text and Hypertext Excerpts from a public lecture by Umberto Eco at the Faculty of Economics of Moscow State University on May 20, 1998 [Online]. Available at: <https://studfile.net/preview/1714056/> (accessed: 12.03.2024). (In Russian)

Cap, P. (2008). Towards the proximization model of the analysis of legitimization in political discourse, *Journal of Pragmatics*, 40, 17–41. DOI: 10.1016/j.pragma.2007.10.002 (In English)

Creeber, G. and Royston, M. (2008). *Digital Culture: Understanding New Media*, UK, McGraw-Hill Education. (In English)

Domingo, D., Quandt, T., Heinonen, A., Paulussen, S., Singer, J. B. and Vujnovic, M. (2008). Participatory journalism practices in the media and beyond: an international comparative study of initiatives in online newspapers, *Journalism Practice*, 2 (3), 326–342. DOI: 10.1080/17512780802281065 (In English)

Fetzer, A. and Lauerbach, G. E. (2007). *Political Discourse in the Media: Cross-cultural perspectives*, Amsterdam, John Benjamins Publishing Company. DOI: 10.1075/PBNS.160 (In English)

McCombs, M. E. (2005). A look at agenda-setting: Past, present and future, *Journalism Studies*, 6, 543–557. DOI: 10.1080/14616700500250438 (In English)

Tsfati, Y. (2014). Uses and perceptions of political media, in Reinemann, C. (Ed.). *Political Communication: Handbooks of Communication Science*, Volume 18, Berlin, de Gruyter, 489–506. DOI: 10.1515/9783110238174.489 (In English)

Verschueren, J. (1999). *Understanding pragmatics*, London, Edward Arnold, New York, Oxford University Press. (In English)

Wall, M. (2017). Mapping Citizen and

Participatory Journalism: In newsrooms, classrooms and beyond, *Journalism Practice*, 11 (2–3), 134–141. DOI: 10.1080/17512786.2016.1245890 (In English)

Wettstein, M. and Wirth, W. (2017). Media Effects: How Media Influence Voters, *Swiss Political Science Review*, 23 (3), 262–269. DOI: 10.1111/spsr.12263 (In English)

Wojcieszak, M. and Mutz, D. C. (2009). Online Groups and Political Discourse: Do Online Discussion Spaces Facilitate Exposure to Political Disagreement?, *Journal of Communication*, 59 (1), 40–56. DOI: 10.1111/j.1460-2466.2008.01403.x (In English)

Все авторы прочитали и одобрили окончательный вариант рукописи.

All authors have read and approved the final manuscript.

Конфликты интересов: у авторов нет конфликта интересов для декларации.

Conflicts of Interest: the author has no conflict of interest to declare.

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, D. Sc. (Philology), Associate Professor, Professor of the Department of Foreign Languages, Institute of Philology, Foreign Languages and Media Communications, Kemerovo State University, Kemerovo, Russia.




Мельник Наталья Владимировна, доктор филологических наук, профессор, профессор кафедры стилистики и риторики Института филологии, иностранных языков и медиакоммуникаций, Кемеровский государственный университет, Кемерово, Россия.

Natalya V. Mel'nik, D. Sc. (Philology), Full Professor, Professor of the Department of Stylistics and Rhetoric, Institute of Philology, Foreign Languages and Media Communications, Kemerovo State University, Kemerovo, Russia.

РАЗДЕЛ II. ПРИКЛАДНАЯ ЛИНГВИСТИКА
SECTION II. APPLIED LINGUISTICS

УДК 004.8:811.1/.2

DOI: 10.18413/2313-8912-2024-10-3-0-4

Гатиятуллина Г.М.¹ 
Гафиятова Э.В.² 
Данилов А. В.³ 

Динамика уровня сложности как функция
морфологических параметров текста на русском языке

¹ НИЛ «Текстовая аналитика», Казанский (Приволжский) федеральный университет
ул. Кремлевская, 18, Казань, 420008, Россия
E-mail: ggaliya-m@mail.ru
ORCID: 0000-0001-6752-2100

² Казанский (Приволжский) федеральный университет
ул. Кремлевская, 18, Казань, 420008, Россия
E-mail: rg-777@yandex.ru
ORCID: 0000-0003-3190-4566

³ НИЛ «Текстовая аналитика», Казанский (Приволжский) федеральный университет
ул. Кремлевская, 18, Казань, 420008, Россия
E-mail: tukai@yandex.ru
ORCID: 0000-0002-2358-1157

*Статья поступила 21 июля 2024 г.; принята 15 сентября 2024 г.;
сентября 30 июня 2024 г.*

Источник финансирования: Исследование выполнено при поддержке гранта РНФ 24-28-01355 «Жанрово-дискурсивные характеристики текста как функция лексического диапазона».

Аннотация. Представленное исследование нацелено на (1) выявление предикторов морфологической сложности и (2) идентификацию маркеров, дифференцирующих предметные области учебного текста на русском языке. Общий объем корпуса, включающий двадцать учебников по биологии и обществознанию трех уровней сложности, соответствующих 6 – 7, 8 – 9, 10 – 11 классам общеобразовательной российской школы, составил 941963 словоформы. Лингвистическая сложность текстов рассчитывалась при помощи модифицированной для русского языка формулы читабельности Флеша–Кинкейда, а взаимозависимость параметров оценивалась на основе корреляционного анализа с использованием программы STATISTICA. Расчеты значений лингвистических параметров, включающих дистрибуцию имени существительного, прилагательного, глагола, а также индекс читабельности, производились при помощи текстового профайлера для русского языка RuLingva (rulex.kpfu.ru/), а метрики частотности девербативов и деадъективов имен существительных были выделены авторами исследования вручную.




Дистрибутивный анализ частотности метрик субстантивации производился с нормализацией текста на 10000 словоупотреблений. Последовательное контрастирование индекса читабельности указанных текстов с метриками «дистрибуция имени существительных», «лексическая плотность», «девербация», «деадъективация» продемонстрировало их линейную взаимозависимость. Читабельность изучаемых текстов находится в обратно пропорциональной зависимости от дистрибуции глагола. Морфологический анализ подтвердил высокий уровень номинативности текстов и стабильный рост субстантивов, объективируемый в увеличении доли суффиксов девербации и деадъективации от 6 к 11 классу. Метрики лексической плотности, дистрибуции имени прилагательного и отдельных суффиксов субстантивов также выступают в качестве дискриминант, дифференцирующих учебные тексты по биологии и обществознанию указанных уровней сложности. Результаты исследования будут полезны специалистам в области текстовой аналитики, компьютерной лингвистики, жанроведения, а также разработчикам тестов и авторам учебно-методических материалов. Перспектива исследования видится в изучении долей сложных терминов-лексем, образованных от латинских и греческих корней, в учебных текстах. Предполагается, что данный параметр имеет потенциал выступать в качестве предиктора лингвистической сложности и дискриминанта предметной области.

Ключевые слова: Учебные тексты; Читабельность; Русский язык; Учебник; Морфологические параметры; Номинализация.

Информация для цитирования: Гатиятуллина Г. М., Гафиятова Э. В., Данилов А. В. Динамика уровня сложности как функция морфологических параметров текста на русском языке // Научный результат. Вопросы теоретической и прикладной лингвистики. 2024. Т. 10. № 3. С. 70-90. DOI: 10.18413/2313-8912-2024-10-3-0-4

UDC 004.8:811.1/2

DOI: 10.18413/2313-8912-2024-10-3-0-4

Galiya M. Gatiyatullina¹ 
Elzara V. Gafiyatova² 
Andrei V. Danilov³ 

Text complexity increase in Russian texts as a function of morphological changes

¹ Text Analytics Laboratory, Kazan Federal University,
18 Kremlevskaya St., Kazan, 420008, Russia
E-mail: galiya-m@mail.ru
ORCID: 0000-0001-6752-2100

Kazan Federal University,
² 18 Kremlevskaya St., Kazan, 420008, Russia
E-mail: rg-777@yandex.ru
ORCID: 0000-0003-3190-4566

³ Text Analytics Laboratory, Kazan Federal University,
18 Kremlevskaya St., Kazan, 420008, Russia
E-mail: tukai@yandex.ru
ORCID: 0000-0002-2358-1157

Received 21 July 2024; accepted 15 September 2024; published 30 September 2024

Acknowledgments: The research was supported by the RSF grant 24-28-01355 “Genre-discourse characteristics of the text as a function of lexical range”.

Abstract: The study aims to identify (1) morphological complexity predictors and (2) domain inherent markers able to differentiate subject areas of academic text in Russian. The total size of the corpus, including textbooks on biology and social studies of three levels of complexity, corresponding to 6-7, 8-9, and 10-11 grades of the Russian school, amounted to 941963 tokens. The linguistic complexity of the texts was assessed using the Flesch-Kincaid readability formula modified for the Russian language, and the interdependence of the parameters was measured based on the correlation analysis conducted with STATISTICA. Calculation of linguistic parameters values, including distribution of nouns, adjectives, verbs, and readability index, were performed using RuLingva (rutex.kpfu.ru/), a text profiler for the Russian language, while the frequency metrics of deverbatives and deadjectives were identified by the contributors manually. To ensure comparability of the metrics, the distributional analysis of deverbatives and deadjectives was performed in the corpus normalized to 10000 tokens. Metrics “noun distribution”, “lexical density”, “deverbation”, “deadjectivation” demonstrated linear interdependence with readability and as such can be viewed as complexity predictors. Inverse correlation was revealed between text readability and verb distribution. Morphological analysis confirmed a high level of texts nominativity and a stable growth of substantives frequency. The latter explicates in an increase in the frequency of deverbation and deadjectivation suffixes in texts from the 6th to the 11th grade. Metrics of lexical density, adjective distribution and substantive suffixes demonstrate ability to discriminate academic texts domains. The research findings are applicable in text analytics, computational linguistics, genre studies, and can be useful for test developers and textbook writers. The authors view the research prospect in the study of compounds of Latin and Greek origin in academic texts. The identified parameters may be used as linguistic complexity predictors and domain discriminants.

Key words: Academic texts; Readability; The Russian language; Textbooks; Morphological parameters; Nominalization.

How to cite: Gatiyatullina, G. M., Gafiyatova, E. V., Danilov, A. V. (2024). Text complexity increase in Russian texts as a function of morphological changes, *Research Result. Theoretical and Applied Linguistics*, 10 (3), 70-90. DOI: 10.18413/2313-8912-2024-10-3-0-4

Введение

Одной из детерминант успеха чтения и понимания текста является его сложность или читабельность. Сложность текстов на английском языке стала объектом исследования почти 70 лет тому назад (Flesch 1949; McLaughlin 1969; Kincaid et al. 1975; Just and Carpenter 1987; Alderson 2000; Koda 2005; Crossley et al. 2008), а исследования сложности текстов на русском языке начали активно изучаться

с 1970-х гг. и в настоящее время продолжают находиться в фокусе современных исследователей (Микк 1970, Мацковский 1976, Оборнева 2006, Шпаковский 2007, Solnyshkina and Kiselnikov 2015).

Для оценки читабельности текста существуют более 20 формул (DuBay, 2007). Однако преимущественное большинство этих формул являются функциями двух метрик – «длина

предложения» и «длина слова». Наиболее распространенной является формула удобочитаемости по Р. Флешу, разработанная для английского языка: УФ (формула удобочитаемости) (англ) = $206.835 - (1.015 \times \text{средняя длина предложения}) - (84.6 \times \text{среднее число слогов})$. Для русских художественных текстов формула Р. Флеша была адаптирована И.В. Оборневой (2006): УФ (рус) = $206,836 - (1,52 \times \text{средняя длина предложения}) - (65,14 \times \text{среднее число слогов})$. Оценки сложности научных и учебно-научных текстов на русском языке производится в настоящее время по формуле $ФК_{\text{Мод}}$ (модифицированная формула Флеша-Кинкейда) = $0.36 \times \text{средняя длина предложения} + 5.76 \times \text{среднее число слогов в слове} - 11.97$ (Solnyshkina et al. 2018).

Известно, что формулы читабельности даже в случае их использования для текстов соответствующих функциональных стилей демонстрируют весьма ограниченную достоверность. Именно поэтому, по признанию отечественных и зарубежных ученых, работающих в рамках дискурсивной комплексологии, вопрос о спектре параметров, влияющих на сложность текста, остается исследовательской нишей (Vahrusheva et al., 2023).

Представленное исследование осуществлено с целью тестирования двух **гипотез**: (1) о взаимозависимости читабельности учебно-научного текста, с одной стороны, и следующих морфологических параметров, с другой стороны: дистрибуция существительных, прилагательных и глагола, лексическая плотность, доля существительных, образованных при помощи девербации и деадъективации; (2) о способности морфологических параметров дифференцировать предметные области.

Обзор литературы

Письменные научные тексты, в том числе лингвистические характеристики

вузовских и школьных учебников не раз становились предметом исследований (Biber 1989; Biber et al. 1999; Biber and Gray 2013, 2016; Halliday and Martin 2003; Banks 2005, 2008; Fang, Schleppegrell and Cox 2006). Для учебных и научных текстов на английском языке выявлена высокая дистрибуция имен существительных и именных словосочетаний. Выраженная предметность (термин Б. Н. Головина), именной характер (Жеребило 2010) или номинативность (Валгина, 2003) научного стиля резко контрастирует с разговорной речью, в которой существенно преобладают глаголы и придаточные конструкции (Biber 1989; Biber and Gray 2013, 2016).

Выявленные зависимости подтверждаются и результатами исследований развития речи: именно номинативные группы представляют особую трудность для понимания учениками начальной школы, поскольку в привычной устной речи существительные используются преимущественно для обозначения предметов и людей, а глаголы – для обозначения действий. Способность заменять придаточные предложения номинативными группами появляется у носителей языка значительно позднее способности формировать простые предложения. Например, фраза *the rate of glass crack growth* (скорость роста трещин стекла) значительно труднее для понимания, чем *how quickly cracks in glass grow* (как быстро разрастаются трещины на стекле) (Halliday and Martin, 2003: 82). Учебный дискурс предполагает высокую степень субстантивации глаголов, поэтому, по мнению М. Халлидея, вопрос восприятия дискурса учебником является отчасти вопросом взросления: даже отучившись пару лет в средних классах ребенок может испытывать трудности с пониманием текста, особенно текстов по точным и техническим предметам, так как именно в них наиболее часто встречаются субстантивированные глаголы и отглагольные существительные (Halliday

and Martin, 2003: 82). Таким образом М. Халлидей фактически предлагает список предикторов сложности, которые в значительной степени обуславливают трудности восприятия учебного дискурса.

По мнению ряда ученых научный дискурс характеризуется двумя активными процессами: девербацией, т.е. образованием отглагольных существительных со значением отвлеченного действия (Biber 2013), и деадъективацией, образованием имен существительных со значением отвлеченного качества (Martin 1991). По мнению О. Д. Митрофановой, такие суффиксы как *-ость*, *-ние* (*-ение*, *-ание*) в общелитературном языке не имеют форм множественного числа и предполагают высокую абстрактность, однако под влиянием языка науки данные номинативы могут обнаруживать множественное число для повышения конкретизации и индивидуализации (Митрофанова 1973: 62-63).

Номинализация или субстантивация, т.е. процесс перехода лексем из класса глаголов, прилагательных, предлогов, союзов в класс существительных, позволяет вместить или «упаковать» больше информации на предложение или сделать предложения более содержательными (Eggins 2004: 96). Это связано с потенциалом «номинативной группы», т.е. части клаузы, которая состоит из существительных и слов, служащих модификатором к имени существительному. С. Эггинс (2004: 95) также пишет, что номинализация обладает двумя текстовыми преимуществами: риторической организацией текста и лексической плотностью. Высокая степень лексической плотности, с одной стороны, наполняет текст содержанием, но с другой стороны, приобретает особую значимость

при оценке трудности восприятия текста. Оценка лексической плотности текста рассчитывается как отношение суммы словоупотреблений знаменательных частей речи, т.е. имени существительного, прилагательного, глагола и наречия, ко всему объему текста в словоупотреблениях (Biber et al. 2021; To et al. 2013).

Модели субстантивации, выявленные для русского языка немногочисленны, но весьма продуктивны (см. Петрова 2009; Мамонтов, Мамонтова 2012). Девербация имеет место при образовании существительных, например, *чтение*, *чихание* (V+ а(е)ни-е); *заятие*, *принятие* (V+ яти-е); *бахвальство*, *баловство* (V+ ств-о); *акция*, *дивизия* (V+ц/зи-я). Существительные *героизм*, *эгоизм* (Adj+изм); *выносливость*, *заносчивость* (Adj+ость); *широта*, *доброта* (Adj+ота); *глубина*, *кривизна* (Adj+ и(з)на) образованы по словообразовательным моделям деадъективации.

Материал и методы исследования

Материал исследования.

Исследование осуществлено на материале российских учебников по биологии и обществознания. Общий объем подкорпуса текстов по биологии составил 399555 словоупотреблений (см. Таблицу 1 и Список иллюстративных источников).

Гуманитарный блок исследования представлен двумя линейками российских учебников по обществознанию для 6-11 классов: (а) под редакцией А. Ф. Никитина и (б) под редакцией Л. Н. Боголюбова. Объем подкорпуса российских учебников по обществознанию составил 542408 словоупотреблений (см. Таблицу 1 и Список иллюстративных источников).

Все учебники в составе корпуса были разделены на три уровня сложности: уровень I – 6-7 классы, уровень II – 8-9 классы, уровень III – 10-11 классы.

Таблица 1. Объем корпуса исследования
Table 1. Size of the research corpus

Уровень	Учебник	Количество словоупотреблений	Учебник	Количество словоупотреблений	Итого
Биология					
I	БС 6 *	26870	БП 6**	37164	64034
	БС 7	27956	БП 7	47969	75925
II	БС 8	30136	БП 8	52173	82309
	БС 9	30010	БП 9	50927	80937
III	БС 10-11	27408	БП 10-11	68942	96350
Итого					399555
Обществознание					
I	ОН 6	16843	ОБ 6 ****	16444	33287
	ОН 7	23111	ОБ 7	23855	46966
II	ОН 8	50492	ОБ 8	40986	91478
	ОН 9	44508	ОБ 9	43231	87739
III	ОН 10	101716	ОБ10	102140	203856
	ОН 11	39289	ОБ11	39793	79082
Итого					542408

*Буквами БС маркированы учебники по биологии Л. Н. Сухоруковой и В. С. Кучменко. Цифра указывает класс.

** БП маркируют учебники по биологии В. В. Пасечника и А. А. Каменского.

***ОН маркируют учебники по обществознанию под редакцией А. Ф. Никитина.

**** ОБ маркируют учебники по обществознанию под редакцией Л. Н. Боголюбова.

Методы и алгоритм исследования.

1. Расчеты метрик морфологических параметров (дистрибуция имени существительного, прилагательного, глагола) и индекса читабельности (ФКмод) осуществлялись при помощи текстового анализатора RuLingva [rulex.kpfu.ru/]. Индекс читабельность ранжирует тексты по годам обучения, т.е. текст, имеющий читабельность 6 (ФКмод=6), соответствует лингвистическим и когнитивным способностям школьников 6 класса и т.д. RuLingva производит морфологический анализ текста и позволяет выгружать списки знаменательных частей речи. Для последующего сравнения каждая метрика была нормализована на 1000 словоупотреблений (см. Таблицу 2).

2. В группе имен существительных каждого текста вручную были выделены девербативы и деадъективы.

3. Для расчета дистрибуции количество каждого суффикса было нормализовано на 10 тыс. словоупотреблений

4. Расчеты среднестатистических значений, корреляции читабельности с морфологическими параметрами (дистрибуция имени существительного, прилагательного, глагола, лексическая плотность, степень субстантивации), а также оценка статистически значимых различий в показателях текстов изучаемых дисциплин осуществлялись в программе Statistica [statsoft.com/]. Для визуализации динамики параметров, коррелирующих с читабельностью текста, были построены

гистограммы и графики (см. Рис. 1-5). Все графики и гистограммы строились на основе метрик каждого из сегментов (глав) в составе изучаемых текстов учебников. Количество изученных сегментов указано в Таблице 2. Например, в учебниках обществознания уровня I

проанализировано 203 отрывка, на уровне II – 143 и т.д.

Результаты

Рассчитанные при помощи текстового анализатора Rulingva метрики изучаемых лингвистических параметров демонстрируют сильную корреляцию с индексом читабельности (см. Табл. 2).

Таблица 2. Лингвистические параметры текстов трех образовательных уровней (I – III)
Table 2. Linguistic features of texts of three grade levels (I – III)

параметр	Уровень I (6-7 классы)			Уровень II (8-9 классы)			Уровень III (10-11 классы)		
	Ср. стат. био (N=203)	Ср. стат. общ (N=76)	р-значение, U-критерий Манна-Уитни	Mean био (N=143)	Mean общ (N=112)	р-значение, U-критерий Манна-Уитни	Mean био (N=145)	Mean общ (N=110)	р-значение, U-критерий Манна-Уитни
класс	6,458	6,526	0,4	8,6	8,473	0,1	10,1	10,5	0,8
Количество слов (Tokens)	702,6	1074,5	<.001*	934,3	1588,4	<.001*	670,3	2355,9	<.001*
Ср. количество слов в предложении	11,4	12,7	<.001*	13,8	14,5	.003*	14,6	16,6	<.001*
Ср. количество слогов в слове	2,6	2,4	<.001*	2,8	2,657	<.001*	2,8	2,8	0,2
Сущ\1000	397,8	341,4	<.001*	406,9	379,9	<.001*	403,1	391,2	<.001*
Прил\1000	151,2	118,0	<.001*	153,7	141,1	<.001*	145,9	163,6	<.001*
Гл.\1000	130	145,7	<.001*	124,8	127,7	0,4	123,7	115,9	<.001*
лекс.плотность\1000	718	657,7	<.001*	725,5	692,4	<.001*	715,2	713,3	0,6
Местоим\1000	54	103	<.001*	55,9	79,7	<.001*	59,5	68,7	<.001*
Индекс ФК (мод) SIS	7,2	6,8	<.001*	8,9	9,04	<.001*	9,3	8,8	<.001*
Наст.в.\1000	90,4	62,9	<.001*	83,0	63,5	<.001*	74,7	64,1	<.001*
Прош.в.\1000	24	41,9	<.001*	24,8	34,1	<.001*	33,0	29,2	0,7

* $p < .05$ – статистически значимые различия

Рассмотрим каждый из параметров.

Читабельность.

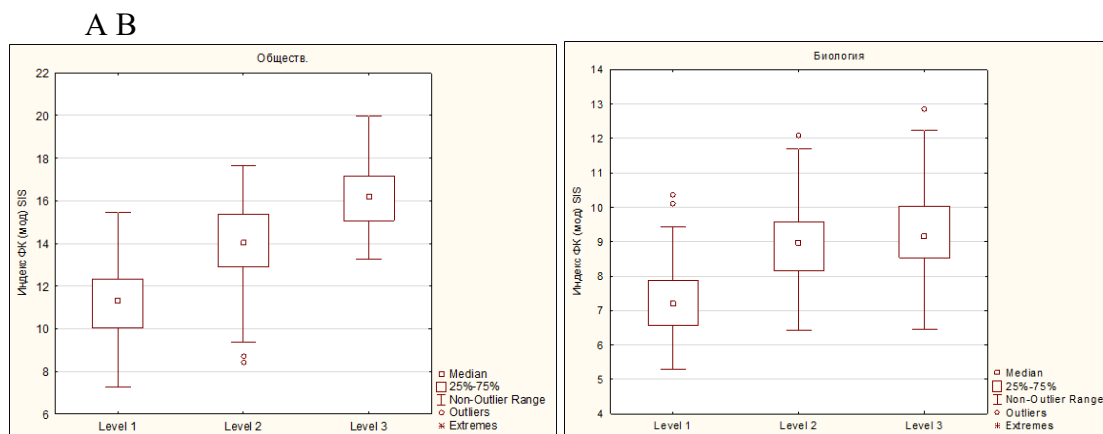
Анализ текстов показал, что для текстов обеих дисциплин характерна динамика роста от уровня I к уровню II, однако читабельность текстов по

обществознанию к уровню III снижается на 0,6, а тексты по биологии, наоборот, незначительно растут к уровню III. Читабельность текстов по

обществознанию для 6-7 классов (уровень I) составляет 6,8, для 8-9 классов (уровень II) читабельность вырастает до 9,04, а для

10-11 классов (уровень III) снижается до 8,8 (рис. 1 А).

Рисунок 1. Динамика индекса читабельности текста по обществознанию (А) и биологии (В)
Figure 1. Readability rate in Russian texts of textbooks on Social Studies (A) and Biology (B)



Метрики, на основании которых осуществляется расчет читабельности, демонстрируют положительную динамику: длина слова на уровне I равна 2,4 слога, на уровне II – 2,7 слога, а на уровне III – 2,8 слога. Длина предложения в текстах по обществознанию составляет 12,7 слов на уровне I, в текстах уровня II среднестатистическая длина предложения составляет 14,5 слов, а на уровне III – 16,6 слов.

Читабельность текстов по биологии показывает положительную динамику: в текстах по биологии уровня I читабельность составляет 7,2, в текстах уровня II – 9, а на уровне III – 9,3 (рис. 1 В). Аналогичным образом линейно растут метрики параметров, формирующих читабельность. Длина слова на I и II уровнях выше, чем в текстах по обществознанию: на уровне I равна 2,6

слога, на уровне II – 2,8 слога, а на уровне III – 2,8 слога. Длина предложения в текстах по биологии в целом ниже, чем в текстах обществознанию, и составляет 11,4 слов на уровне I, на уровне II длина предложения составляет 13,8 слов, а на уровне III – 14,6 слов.

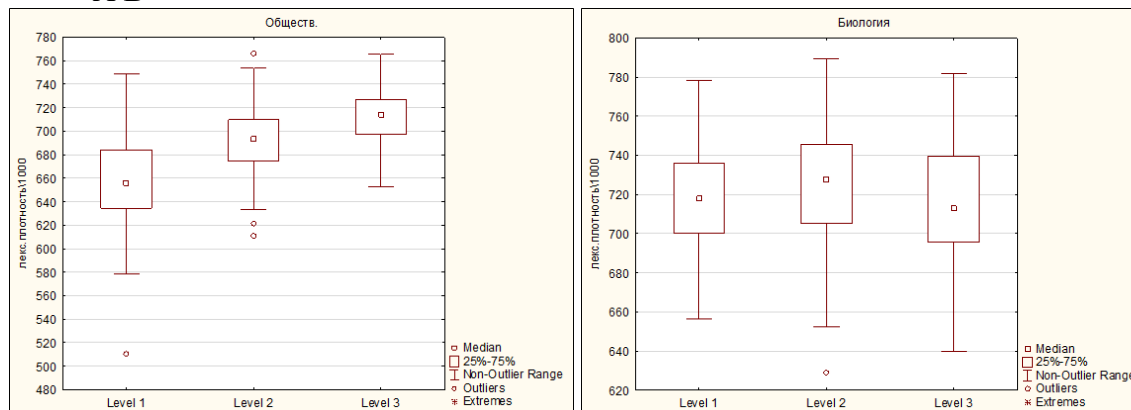
Лексическая плотность.

Динамика степеней лексической плотности в текстах обеих дисциплин отличается в значительной степени. В текстах по обществознанию лексическая плотность растет от уровня к уровню: на уровне I – 658 знаменательных частей речи на 1000 словоупотреблений, в текстах уровня II лексическая плотность выросла до 692 слов знаменательных частей речи на 1000 словоупотреблений, а на уровне III лексическая плотность составила 713,3 (рис. 2 А).

Рисунок 2. А) Лексическая плотность на 1000 слов (Обществознание); В) Лексическая плотность на 1000 слов (Биология)

Figure 2. А) Lexical density per 1000 words (Social Studies); В) Lexical density per 1000 words (Biology)

А В



В текстах по биологии метрики лексической плотности высоки в текстах всех уровней. Однако стабильный рост от уровня к уровню не выявлен. В текстах уровня I лексическая плотность составляет 718, далее растет в текстах уровня II, но на уровне III падает ниже уровня I и составляет 715,2 слов знаменательных частей речи на 1000 словоупотреблений (рис. 2 В). При этом необходимо отметить, что даже наиболее минимальное значение лексической плотности в текстах по биологии выше, чем максимальное значение лексической плотности в текстах по обществознанию.

Дистрибуция частей речи.

Анализ дистрибуции частей речи показал существенное преобладание имен существительных в текстах обеих дисциплин. Тем не менее, в текстах по

обществознанию динамика роста существительных наиболее выражена. В текстах уровня I дистрибуция имен существительных составляет 341 вхождение на 1000 словоупотреблений, в текстах уровня II этот параметр имеет значение 380 вхождений на 1000 словоупотреблений, а в текстах по обществознанию уровня III дистрибуция имен существительных составляет 391,2 вхождение на 1000 словоупотреблений (см. Рисунок 3).

В текстах по биологии дистрибуция существительных в целом выше, однако динамика роста неравномерна: на уровне I – 397,8 вхождений на 1000 словоупотребления, на уровне II – 407 вхождений, а в текстах уровня III – 403 вхождения (см. Рисунок 4).

Рисунок 3. Динамика дистрибуции частей речи в текстах по обществознанию
Figure 3. Part of Speech distribution rate in Social Studies texts

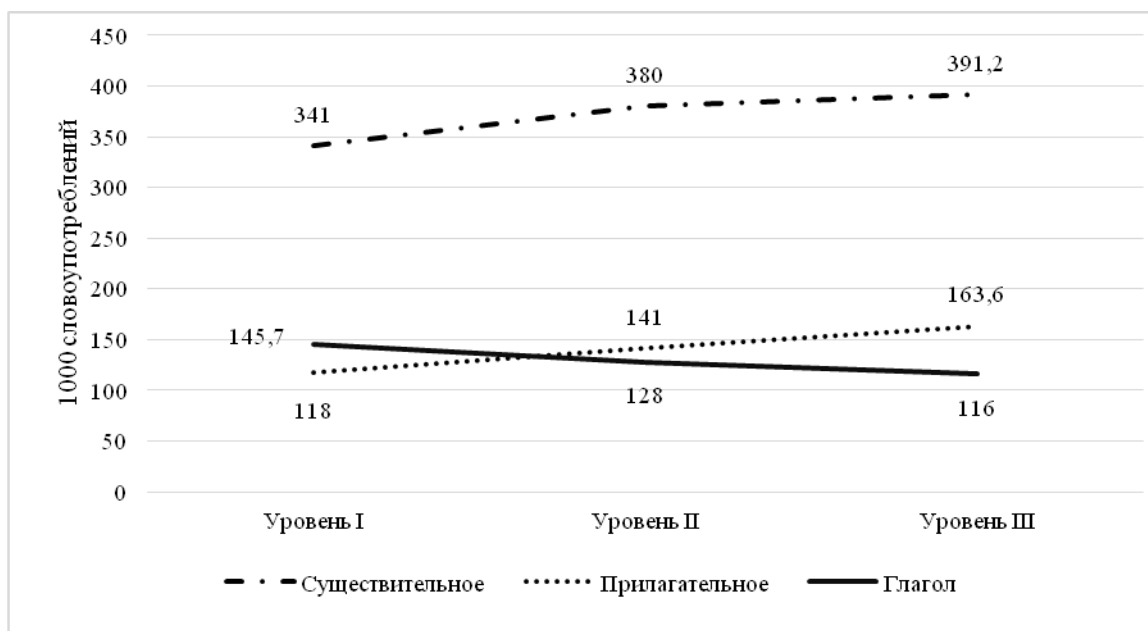
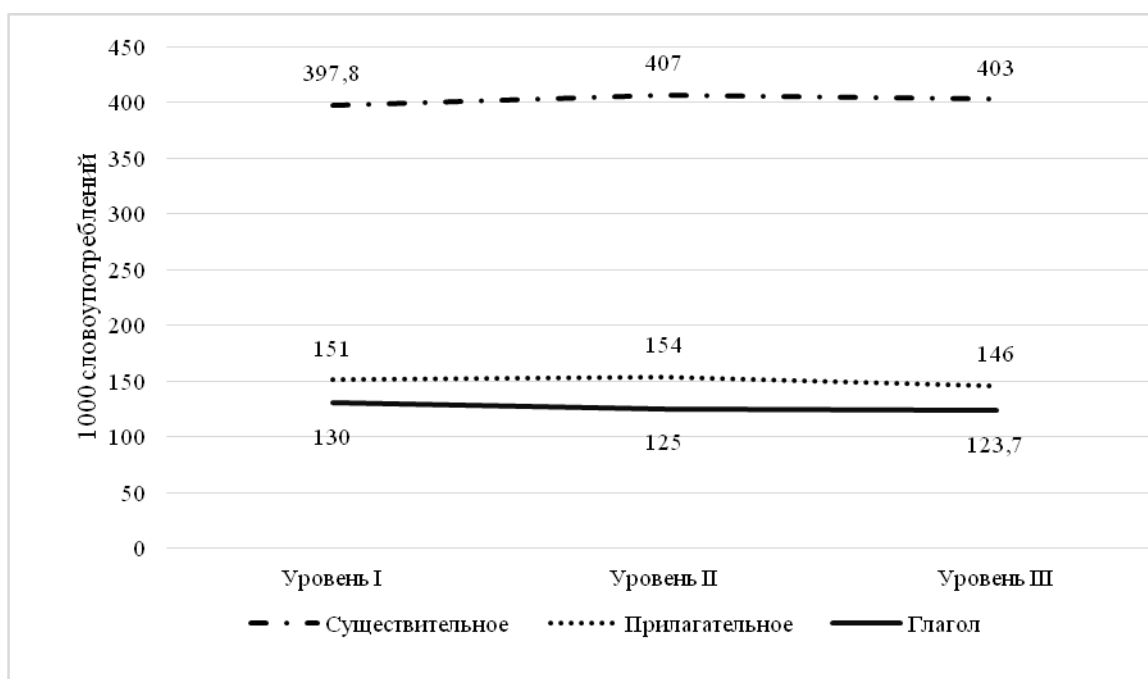


Рисунок 4. Динамика дистрибуции частей речи в текстах по биологии
Figure 4. Part of Speech distribution rate in Biology texts



Дистрибуция имен прилагательных в изучаемых текстах несколько ниже дистрибуции имен существительных: в текстах по обществознанию динамика роста дистрибуции прилагательных равномерна: 118 (I уровень) → 141 (II

уровень) → 163, 6 (III уровень) на 1000 словоупотреблений. Для сравнения в текстах биологии: 151 (I уровень) → 154 (II уровень) → 146 (III уровень) на 1000 словоупотреблений.

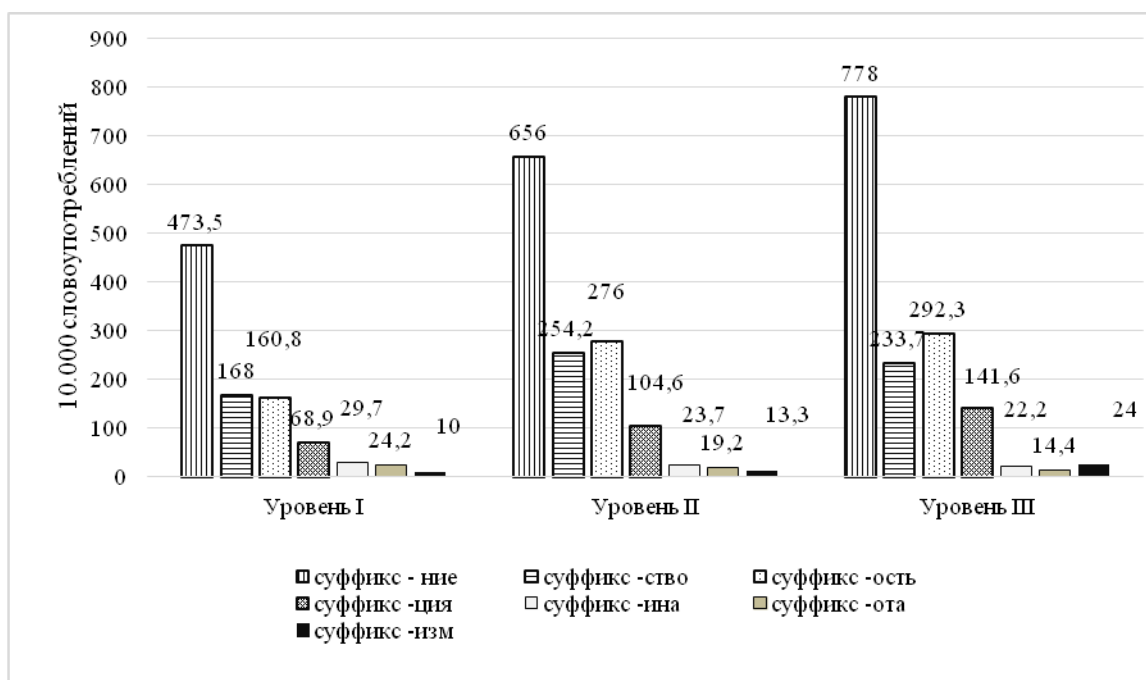
В текстах обеих дисциплин динамика роста дистрибуции глаголов стабильно отрицательная. Обществознание: 145,7 (I уровень) → 128 (II уровень) → 116 (III уровень) на 1000 словоупотреблений. В текстах по биологии: 130 (I уровень) → 125 (II уровень) → 123,7 (III уровень) на 1000 словоупотреблений.

Номинативность текстов.

В текстах обеих дисциплин выявлен активный процесс субстантивации. В текстах по обществознанию совокупность суффиксов субстантивации в текстах уровня I составляет 935,1 вхождений на 10 тыс. словоупотреблений, в текстах уровня II дистрибуция суффиксов субстантивации

растет до 1347 вхождений на 10 тыс. словоупотреблений, а в текстах по обществознанию уровня III суффиксы субстантивации вырастает до 1506,2 вхождений на 10 тыс. словоупотреблений. В текстах по биологии дистрибуция суффиксов субстантивации на всех уровнях в целом ниже, чем в текстах по обществознанию: на уровне I – 929 вхождений, в текстах уровня II наблюдается рост суффиксов субстантивации до 1285,4 на 10 тыс. словоупотреблений, а на уровне III – снижение до 1273,2 вхождений (см. Рисунок 5).

Рисунок 5. Динамика суффиксов субстантивации в текстах по обществознанию
Figure 5. Distribution rate of nominalizations in Social Studies school textbooks



Динамика роста дистрибуции суффикса *-ние* является наиболее стабильной: в текстах по обществознанию уровня I – 473,5 вхождений на 10 тыс. словоупотреблений, уровня II – 656 вхождений, а в текстах уровня III его частотность возрастает до 778 вхождений (см. Рисунок 5).

Вторую позицию по частотности среди суффиксов субстантивации в

изучаемых текстах по обществознанию занимают суффиксы *-ство* и *-ость*. Суффикс *-ство* демонстрирует значительный рост от уровня I до уровня II: 168 → 254,2. При этом дистрибуция суффикса *-тво* незначительно падает в текстах уровня III и составляет 233,7 вхождений на 10 тыс. словоупотреблений. Динамика роста суффикса *-ость* стабильна на всех уровнях: 160,8 вхождений на

уровне I, 276 вхождений – на уровне II и 292,3 – на уровне III.

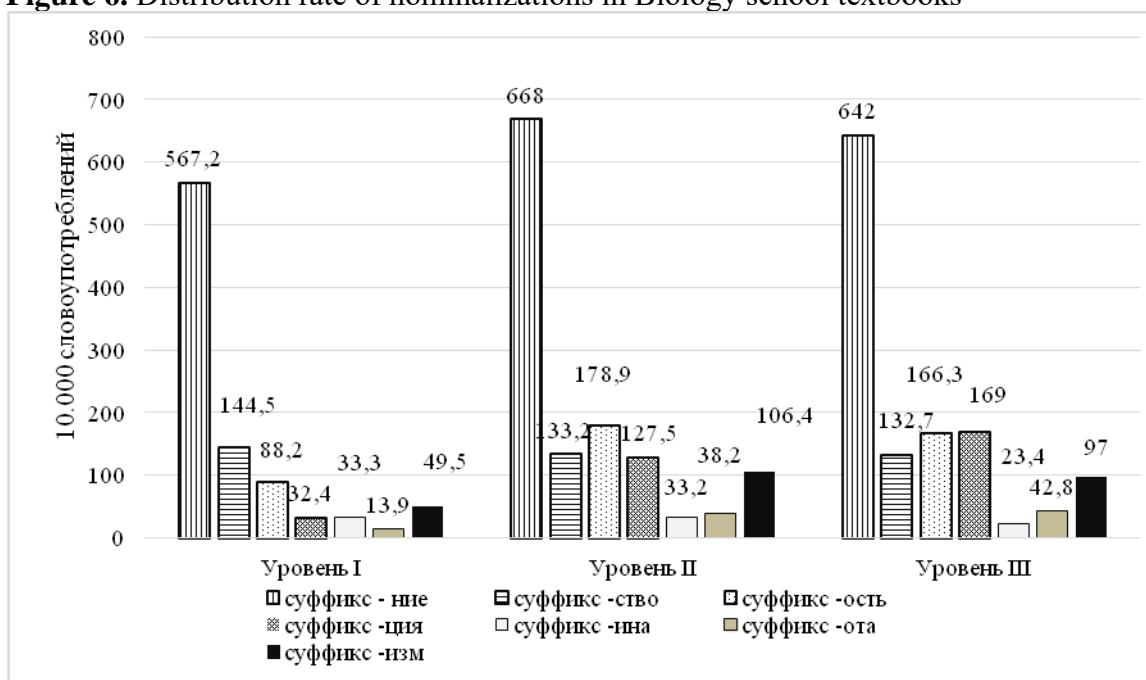
Из относительно менее частотных суффиксов (-ция, -ина, -ота, -изм) два суффикса -ция и -изм показывают стабильную тенденцию роста. Суффикс -ция: 68,9 вхождений на уровне I, 104,6 вхождений – на уровне II, 141,6 вхождений на уровне III в текстах по обществознанию. Суффикс -изм: 10 – на уровне I, 13,3 – на уровне II, 24 – на уровне III.

Суффиксы -ина и -ота, наоборот, демонстрирует стабильное снижение: -ина имеет 24,2 вхождения в текстах по обществознанию на уровне I, снижается до

23,7 в текстах уровня II, а на уровне III – до 22,2 вхождения. Дистрибуция суффикса -ота в текстах по обществознанию на уровне I составляет 24,2 вхождения, в текстах уровня II падает до 19,2, а на уровне III – снижается до 14,4.

Аналогично текстам по обществознанию в текстах по биологии наиболее частотным суффиксом субстантивации является суффикс -ние, однако динамика его роста не стабильна. Так, в текстах по обществознанию уровня I он имеет 567,2 вхождений, в текстах уровня II – 668 вхождений, а в текстах уровня III его дистрибуция незначительно снижается до 642 (см. Рис 6).

Рисунок 6. Динамика суффиксов субстантивации в текстах по биологии
Figure 6. Distribution rate of nominalizations in Biology school textbooks



Второе место по частотности в текстах биологии занимают суффиксы -ство и -ость, однако динамика их дистрибуции является разнонаправленной. Суффикс -ство стабильно снижается в текстах по биологии от уровня I к уровню III: зафиксировано 168 вхождений суффикса -ство на уровне I, 133,2 вхождения – на уровне II, а на уровне III дистрибуция суффикса -ство также

незначительно падает и составляет 132,7 вхождения.

Динамика изменений частотности суффикса -ость не стабильна: в текстах по обществознанию на уровне суффикс -ость используется в среднем 88,2 раза, на уровне II его частотность поднимается до 178,9, а в текстах уровня III – падает до 166,3 вхождений.

Среди менее частотных суффиксов субстантивации (-ция, -ина, -ота, -изм) два

суффикса *-ция* и *-ота* показывают стабильную тенденцию роста, а суффикс *-ина* демонстрирует стабильное снижение. Что касается дистрибуции суффикса *-изм* в текстах по биологии, его метрики выстраиваются по параболе с ветвями вниз: от 49,5 к 106,4 и к 97.

Обсуждение

Результаты представленного исследования в значительной степени способствуют расширению базы данных о референсных диапазонах метрик морфологических параметров изученных текстов. Их сопоставление с данными более ранних публикаций высвечивает меж- и внутриязыковые сходства и различия текстов разных жанров, с одной стороны, а также исследовательские ниши, с другой стороны.

Тексты современных российских учебников по обществознанию и биологии характеризуются приблизительно равной длиной слова: 2,7 в биологических текстах и 2,8 в текстах обществознания на уровне II и 2,8 для обеих дисциплин на уровне III. Длина предложения также приблизительно одинакова для обеих дисциплин: 14,5 слов в предложении по обществознанию и 13,8 по биологии на уровне II, на уровне III их длина предложения одинакова и составляет 16,6 слов. Незначительные различия в длине предложения и слова дают несколько более высокую читабельность текстов по биологии по сравнению с текстами по обществознанию (см. Рисунок 1 и 2). Однако в целом читабельность русскоязычных текстов ниже читабельности англоязычных текстов аналогичных дисциплин. Так, тексты уровня III по дисциплине *Social Studies* имеют индекс читабельности 13,3, а по дисциплине *Biology* – 13,04, что в целом значительно выше индекса читабельности текстов уровня III сравниваемых русскоязычных учебников (Gatiyatullina et al. 2020).

Лексическая плотность текстов биологии и обществознания на русском языке существенно выше по сравнению с

учебниками *Biology* и *Social Studies*: диапазон лексической плотности в англоязычных учебниках – 475 вхождений на 1000 словоупотреблений на уровне I в учебниках *Biology* и максимальное значение – 501 вхождение также в учебниках *Biology* (Gatiyatullina et al. 2020). В русскоязычных учебниках диапазон лексической плотности имеет диапазон от 658 до 718 вхождений. Очевидно, что данный факт можно объяснить принадлежностью языков сравниваемых текстов к разным типам: артиклевый английский аналитичен vs безартиклевый русский флективен.

Сравнение данных по дистрибуции существительных, прилагательных и глаголов с результатами (1) О. Д. Митрофановой (1973) в научных текстах и (2) О. Б. Сиротининой (2009) по предметным областям не представляется возможным, т.к. в известных нам публикациях они представлены только в относительных величинах (процентах). Однако в целом можно утверждать, что изученные нами тексты уровня III, в большей степени приближенные к текстам, изученным российскими учеными ранее, имеют значения, сопоставимые с опубликованными О. Д. Митрофановой (1973) и О. Б. Сиротининой (2009). В письменных общенаучных текстах дистрибуция существительных, по материалам О. В. Сиротининой (2009) составляет 374 вхождения на 1000 словоупотреблений, а имени прилагательного – 152. Выявленные нами метрики аналогичны: доля имен существительных колеблется в диапазоне от 341 до 403, имя прилагательное – от 118 до 163 (см. Рисунок 3, 4). В целом такие различия могут быть детерминированы тем, что изученные нами тексты имеют предметную специализацию, а изученные О. В. Сиротининой (2009) тексты являются общенаучными.

Что касается дистрибуции глагола, то здесь наши данные значительно выше полученных М. Н. Кожинной (1972). В

биологических текстах уровня III дистрибуция глагола демонстрирует минимальное значение и составляет 123,5 вхождений на 1000 словоупотреблений, а текстах по обществознанию – 116 вхождений. В научных текстах, изученных М. Н. Кожиной (1972), дистрибуция глагола составляет только 90 вхождений на 1000 словоупотреблений. Данные различия косвенно указывают на более высокий уровень номинативности и, как следствие, степени сложности научных текстов. Учебным текстам, напротив, за счет большего присутствия в их составе глаголов, свойственна нарративность, рассчитываемая как отношение глаголов к существительным (Vahrusheva et al. 2023).

Говоря о процессе, номинализации необходимо отметить положительную динамику роста именных суффиксов в текстах обеих дисциплин. Это в целом подтверждает выводы, сделанные А. С. Мамонтовым и Е. Ю. Мамонтовой (2012) относительно частотности суффикса *-ние* в публицистических текстах. Вторую по частотности позицию после суффикса *-ние* в публицистических текстах (газета Московский комсомолец), изученных А.С. Мамонтовым и Е.Ю. Мамонтовой (2012), занимает суффикс *-ция*. В нашем корпусе по обществознанию суффикс *-ция* показывает стабильный рост от уровня I к уровню III, однако его доля уступает частотности суффиксов *-ость* и *-ство*: он занимает только четвертую строчку среди суффиксов субстантивации. Что касается *-ция* в текстах по биологии, то там он занимает четвертое место на уровне I, третье – на уровне II, а в текстах уровня III – уже второе место. В целом эти данные согласуются с результатами исследования А. С. Мамонтова и Е. Ю. Мамонтовой (2012). Статистический анализ демонстрирует, что преимущественное большинство изученных нами параметров коррелируют с индексом читабельности, т.е. могут использоваться в качестве предикторов сложности текстов.

Заключение и перспектива исследования

В данном исследовании были изучены школьные учебники с целью (1) выявления предикторов морфологической сложности и (2) идентификации маркеров, дифференцирующих предметные области учебного текста на русском языке. В корпус вошли двадцать учебников по биологии и обществознанию трех уровней сложности, соответствующих 6 – 7, 8 – 9, 10 – 11 классам общеобразовательной российской школы, общим объемом 941963 словоформы. Анализ текстов школьных учебников по обществознанию и биологии на русском языке показал стабильную динамику роста индекса читабельности текстов от уровня к уровню и положительную взаимозависимость индекса читабельности учебно-научного текста, с одной стороны, и морфологических параметров, включая дистрибуцию существительных, прилагательных и долей существительных, образованных девербацией и деадъективацией, с другой. Индекс читабельности текстов имеет негативную корреляцию с дистрибуцией глагола. Морфологический анализ продемонстрировал высокий уровень номинативности текстов и динамику стабильного роста субстантивации в текстах за счет суффиксов девербации и деадъективации. Диапазоны значений лексической плотности и их динамика в текстах различных уровней сложности, а также дистрибуция существительных, прилагательных и отдельных суффиксов субстантивации выявили различия метрик учебников двух дисциплин, которые могут быть использованы в качестве маркеров предметной дифференциации при анализе больших языковых данных. Перспектива исследования видится в изучении долей сложных терминов-лексем, образованных от латинских и греческих корней, поскольку для биологии такой тип терминообразования является одним из наиболее активных. Валидация данного

параметра в корпусах разных предметных областей способна продемонстрировать его корреляцию с лингвистической сложностью текстов и дифференцировать тексты разных предметных областей.

Список источников

Обществознание. 6 класс: учеб. для общеобразоват. учреждений / Н. Ф. Виноградова, Н. И. Городецкая, Л. Ф. Иванова и др.; под ред. Л. Н. Боголюбова, Л. Ф. Ивановой. М.: Просвещение, 2012. 111 с.

Обществознание. 7 класс: учеб. для общеобразоват. учреждений с приложением на электронном носителе / Л. Н. Боголюбов, Н. И. Городецкая, Л. Ф. Иванова и др.; под ред. Л. Н. Боголюбова, Л. Ф. Ивановой. М.: Просвещение, 2013.

Боголюбов Л. Н. Обществознание. 8 класс: учебник для общеобразоват. учреждений. Под ред. Боголюбова Л. Н., Городецкой Н. И. М.: Просвещение, 2010. 255 с.

Боголюбов Л. Н. Обществознание. 9 класс: учебник для общеобразоват. учреждений. Под ред. Боголюбова Л. Н. М.: Просвещение, 2014. 255 с.

Обществознание. 10 класс: учеб. для общеобразоват. организаций: базовый уровень / Л. Н. Боголюбов, Ю. А. Аверьянов, А. В. Белявский и др.; под ред. Л. Н. Боголюбова и др. М.: Просвещение, 2014. 350 с.

Обществознание. 10 класс: учебник для общеобразовательных организаций: профильный уровень / Л. Н. Боголюбов и др.; под ред. Л. Н. Боголюбова, А. Ю. Лазебниковой, Н. М. Смирновой. 8-е изд. М.: Просвещение, 2014. 415 с.

Обществознание. 11 класс: учебник для учащихся общеобразовательных учреждений: профильный уровень / Л. Н. Боголюбов и др.; под ред. Л. Н. Боголюбова, А. Ю. Лазебниковой, К. Г. Холодковского. 3-е изд. М.: Просвещение, 2010. 414 с.

Обществознание. 6 класс: учебник для общеобразовательных учреждений / А. Ф. Никитин. 4-е изд., стер. М.: Дрофа, 2011. 125 с.

Обществознание. 7 класс: учебник / А. Ф. Никитин. 6-е изд., стер. М.: Дрофа, 2014. 94 с.

Никитин А. Ф., Никитина Т. И. Обществознание. 8 класс. Учебник. М.: Дрофа, 2014. 256 с.

Никитин А. Ф., Никитина Т. И. Обществознание. 9 класс. Учебник. М.: Дрофа, 2014. 208 с.

Обществознание. 10 класс: базовый уровень: учебник / А. Ф. Никитин. 10-е изд., стер. М.: Дрофа, 2014. 238 с.

Обществознание. 11 класс: базовый уровень: учебник / А. Ф. Никитин. 6-е изд., стер. М.: Дрофа, 2013. 237 с.

Сухорукова Л. Н. Биология. Живой организм. 5-6 классы: учеб. для общеобразоват. организаций / Л. Н. Сухорукова, В. С. Кучменко, И. Я. Колесникова. 3-е изд. М.: Просвещение, 2014. 143 с.

Сухорукова Л. Н. Биология. Разнообразие живых организмов. 7 класс: учеб. для общеобразоват. Организаций / Л. Н. Сухорукова, В. С. Кучменко, И. Я. Колесникова. М.: Просвещение, 2014. 159 с.

Сухорукова Л. Н. Биология. Человек. Культура здоровья: учеб. для 8 кл. общеобразоват. учреждений / Л. Н. Сухорукова, В. С. Кучменко, Т. А. Цехмистренко. М.: Просвещение, 2009. 160 с.

Сухорукова Л. Н. Биология. Живые системы и экосистемы. 9 класс: учеб. для общеобразоват. учреждений / Л. Н. Сухорукова, В. С. Кучменко. М.: Просвещение, 2010. 143 с.

Сухорукова Л. Н. Биология. 10-11 классы: учеб. для общеобразоват. учреждений / Л. Н. Сухорукова, В. С. Кучменко, Т. В. Иванова. М.: Просвещение, 2011. 127 с.

Биология. Бактерии, грибы, растения. 6 кл.: учеб. для общеобразоват. учреждений / В. В. Пасечник. 14-е изд., стереотип. М.: Дрофа, 2011. 304 с.

Биология. 7 класс: учеб. для общеобразоват. организаций / В. В. Пасечник, С. В. Суматохин, С. Г. Калинова; под ред. В. В. Пасечника. 3-е изд. М.: Просвещение, 2014. 256 с.

Биология. 8 класс: учеб. для общеобразоват. учреждений / В. В. Пасечник, А. А. Каменский, Г. Г. Швецов ; под ред. В. В. Пасечника. М.: Просвещение, 2010. 255 с.

Биология. Введение в общую биологию и экологию: Учеб. для 9 кл. общеобразоват. учеб. заведений / А. А. Каменский, Е. А. Криксунов, В. В. Пасечник. 3-е изд., стереотип. М.: Дрофа, 2002. 304 с.

Общая биология. 10-11 класс: учеб. для общеобразоват. учреждений / А. А. Каменский, Е. А. Криксунов, В. В. Пасечник. М.: Дрофа, 2005. 367 с.

Corpus Materials

Vinogradova, N. F., Gorodetskaya, N. I., Ivanova, L. F. et al. (2012) *Obshchestvoznaniye*. 6 klass: ucheb. dlya obshcheobrazovatelnykh uchrezhdeniy [Social science. 6th grade: textbook for general educational institutions], in Bogolyubov, L. N., Ivanova, L. F. (eds.), *Prosvescheniye*, Moscow, Russia. (*In Russian*)

Vinogradova, N. F., Gorodetskaya, N. I., Ivanova, L. F. et al. (2013) *Obshchestvoznaniye*. 7 klass: ucheb. dlya obshcheobrazovatelnykh uchrezhdeniy [Social science. 7th grade: textbook for general educational institutions], in Bogolyubov, L. N., Ivanova, L. F. (eds.), *Prosvescheniye*, Moscow, Russia. (*In Russian*)

Bogolyubov, L. N., Gorodetskaya, N. I. (2010) *Obshchestvoznaniye*. 8 klass: ucheb. dlya obshcheobrazovatelnykh uchrezhdeniy [Social science. 8th grade: textbook for general educational institutions], *Prosvescheniye*, Moscow, Russia. (*In Russian*)

Bogolyubov, L. N. (2014) *Obshchestvoznaniye*. 9 klass: ucheb. dlya obshcheobrazovatelnykh uchrezhdeniy [Social science. 9th grade: textbook for general educational Institutions], *Prosvescheniye*, Moscow, Russia. (*In Russian*)

Bogolyubov, L. N. (2014) *Obshchestvoznaniye*. 10 klass: ucheb. dlya obshcheobrazovatelnykh uchrezhdeniy [Social science. 10th grade: textbook for general educational institutions: general level], in L. N. Bogolyubov, Yu. A. Averyanov, A. V. Belyavsky (eds.), *Prosvescheniye*, Moscow, Russia. (*In Russian*)

Bogolyubov, L. N. (2014) *Obshchestvoznaniye*. 10 klass: ucheb. dlya obshcheobrazovatelnykh organizatsiy: profilnyy uroven [Social science. 10th grade: textbook for general education organizations: advanced level], in Bogolyubov, L. N., Lazebnikova, A. Yu., Smirnov, N. M. (eds.), 8th ed., *Prosvescheniye*, Moscow, Russia. (*In Russian*)

Bogolyubov, L. N. (2010) *Obshchestvoznaniye*. 11 klass: ucheb. dlya obshcheobrazovatelnykh organizatsiy: profilnyy uroven [Social science. 11th grade: textbook for students of general education institutions: profile level], in Bogolyubov, L. N., Lazebnikova, A. Yu., Kholodkovsky, K. G. (eds.), 3rd ed., *Prosvescheniye*, Moscow, Russia. (*In Russian*)

Nikitin, A. F. (2011) *Obshchestvoznaniye*. 6 klass: ucheb. dlya obshcheobrazovatelnykh uchrezhdeniy [Social science. 6th grade: textbook for educational institutions], 4th ed., Drofa, Moscow, Russia. (*In Russian*)

Nikitin, A. F. (2014) *Obshchestvoznaniye*. 7 klass: ucheb. dlya obshcheobrazovatelnykh uchrezhdeniy [Social science. 7th grade: textbook], 6th ed., Drofa, Moscow, Russia. (*In Russian*)

Nikitin, A. F. (2014) *Obshchestvoznaniye*. 8 klass: ucheb. dlya obshcheobrazovatelnykh uchrezhdeniy [Social science. 8th grade. Textbook], Drofa, Moscow, Russia. (*In Russian*)

Nikitin, A. F. (2014) *Obshchestvoznaniye*. 9 klass: ucheb. dlya obshcheobrazovatelnykh uchrezhdeniy [Social science. 9th grade. Textbook], Drofa, Moscow, Russia. (*In Russian*)

Nikitin, A. F. (2014) *Obshchestvoznaniye*. 10 klass: ucheb. dlya obshcheobrazovatelnykh uchrezhdeniy [Social science. 10th grade: textbook for general education organizations: general level], 10th ed., Drofa, Moscow, Russia. (*In Russian*)

Nikitin, A. F. (2013) *Obshchestvoznaniye*. 11 klass: ucheb. dlya obshcheobrazovatelnykh uchrezhdeniy [Social science. Grade 11: textbook for general education organizations: general level], 6th ed., Drofa, Moscow, Russia. (*In Russian*)

Sukhorukova, L. N., Kuchmenko, V. S., Kolesnikova, I. Ya. (2014) *Biologiya. Zhivoy organizm*. 5-6 klassy: ucheb. dlya obshcheobrazovatelnykh organizatsiy [Biology. Living organism. Grades 5-6: textbook for general educational institutions], 3rd ed., *Prosvescheniye*, Moscow, Russia. (*In Russian*)

Sukhorukova, L. N., Kuchmenko, V. S., Kolesnikova, I. Ya. (2014) *Biologiya. Raznoobraziye zhivyykh organizmov*. 7 klass: ucheb. dlya obshcheobrazovatelnykh organizatsiy [Biology. Diversity of living organisms. 7th grade: educational. for general education organizations], *Prosvescheniye*, Moscow, Russia. (*In Russian*)

Sukhorukova, L. N., Kuchmenko, V. S., Tsekhmistrenko, T. A. (2009) *Biologiya. Chelovek. Kultura zdorovya: ucheb. dlya 8 kl. obshcheobrazovat. uchrezhdeniy* [Biology. Human. Culture of health. 8th grade: textbook for general educational institutions], Prosvescheniye, Moscow, Russia. (*In Russian*)

Sukhorukova, L. N., Kuchmenko, V. S. (2010) *Biologiya. Zhivyye sistemy i ekosistemy. 9 klass: ucheb. dlya obshcheobrazovat. uchrezhdeniy* [Biology. Living systems and ecosystems. 9th grade: textbook for general educational institutions], Prosvescheniye, Moscow, Russia. (*In Russian*)

Sukhorukova, L. N., Kuchmenko, V. S., Ivanova, T. V. (2011) *Biologiya. 10-11 klassy: ucheb. dlya obshcheobrazovat. uchrezhdeniy* [Biology. Grades 10-11: textbook for general educational institutions], Prosvescheniye, Moscow, Russia. (*In Russian*)

Pasechnik, V. V. (ed.) (2011) *Biologiya. Bakterii, griby, rasteniya. 6 kl.: ucheb. dlya obshcheobrazovat. uchrezhdeniy* [Biology. Bacteria, fungi, plants. 6th grade: textbook for general educational institutions], 14th ed., Drofa, Moscow, Russia. (*In Russian*)

Pasechnik, V. V., Sumatokhin, S. V., Kalinova, S. G. (2014) *Biologiya. 7 klass: ucheb. dlya obshcheobrazovat. organizatsiy* [Biology. 7th grade: textbook for general educational institutions], 3rd ed., Prosvescheniye, Moscow, Russia. (*In Russian*)

Pasechnik, V. V., Kamensky, A. A., Shvetsov, G. G. (2010) *Biologiya. 8 klass: ucheb. dlya obshcheobrazovat. organizatsiy* [Biology. 8th grade: textbook for general educational institutions], Prosvescheniye, Moscow, Russia. (*In Russian*)

Kamensky, A. A., Kriskunov, E. A., Pasechnik, V. V. (2002) *Biologiya. Vvedeniye v obshchuyu biologiyu i ekologiyu: Ucheb. dlya 9 kl. obshcheobrazovat. ucheb. zavedeniy* [Biology. Introduction to general biology and ecology: 9th grade: textbook for general educational institutions], 3rd ed., Drofa, Moscow, Russia. (*In Russian*)

Kamensky, A. A., Kriskunov, E. A., Pasechnik, V. V. (2005) *Obshchaya biologiya. 10-11 klass: ucheb. dlya obshcheobrazovat. uchrezhdeniy* [General biology. 10-11 grade: textbook for general educational institutions], Drofa, Moscow, Russia. (*In Russian*)

Список литературы

Валгина Н. С. Теория текста. М.: Логос, 2003. 280 с.

Жеребило Т. В. Словарь лингвистических терминов. Изд. 5-е, испр. и доп. Назрань: ООО «Пилигрим», 2010. 486 с.

Кожина М. Н. О речевой системности научного стиля сравнительно с некоторыми другими / М. Н. Кожина. Пермь: Изд-во Пермского гос. ун-та, 1972. 395 с.

Мамонтов А. С., Мамонтова Е. Ю. Лингвостатистический анализ использования девербативов в различных видах текстов русского и английского языков // Вестник Российского университета дружбы народов. Серия: Лингвистика. М.: РУДН. 2012. №2. С. 107–110.

Мацковский М. С. Проблемы читабельности печатного материала / М. С. Мацковский // Смысловое восприятие речевого сообщения в условиях массовой коммуникации; отв. ред. Т. М. Дридзе, А. А. Леонтьев. М.: Наука, 1976. С. 126–142.

Микк Я. А. О факторах понятности учебного текста: автореф. дисс. ... канд. пед. наук: 13.00.00. Тарту, 1970. 22 с.

Митрофанова О. Д. Язык научно-технической литературы. М.: Изд-во Моск. ун-та. 1973. 147 с.

Оборнева И. В. Автоматизированная оценка сложности учебных текстов на основе статистических параметров: дисс. ... канд. пед. наук.: 13.00.02. М. 2006. 65 с.

Петрова Н. Е. Субстантивация и деадвербиализация наречий в современном русском языке / Н. Е. Петрова // Известия Уральского государственного университета. Сер. 2, Гуманитарные науки. 2009. № 1/2 (63). С. 35–42.

Разговорная речь в системе функциональных стилей современного русского литературного языка. Грамматика: / О. Б. Сиротинина, Э. А. Столярова, Н. Г. Мартыненко и др.; под ред. О. Б. Сиротининой. 2. изд., стер. Москва, УРСС. 2009. 309 с.

Шпаковский Ю. Ф. Оценка трудности восприятия и оптимизация сложности учебного текста (на материале текстов по химии): автореф. дисс. ... канд. филол. наук: 10.02.19. Минск, 2007. 22 с.

Alderson J. Ch. Assessing reading. Cambridge University Press, 2000. 398 p.

- Banks D. On the historical origins of nominalized process in scientific text // *English for Specific Purposes*. 2005. Vol. 24. Iss. 3. Pp. 347–367. DOI: 10.1016/j.esp.2004.08.002
- Banks D. *The Development of Scientific Writing: Linguistic Features and Historical Context*. London: Equinox, 2008. 221 p.
- Biber D., Gray B. *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press, 2016. 276 p.
- Biber D., Gray B. Nominalising the verb phrase in academic scientific writing / D. Biber and B. Gray // *The Verb Phrase in English* / Bas Aaarts et al. (eds.). London. Cambridge University Press. 2013. Pp. 99–132.
- Biber D. *Grammar of spoken and written English* / D. Biber, S. Johansson, G. Leech, S. Conrad and E. Finegan. John Benjamins, Amsterdam. 2021. 1220 p.
- Biber D. *The Longman grammar of spoken and written English* / D. Biber, S. Johansson, G. Leech, S. Conrad and E. Finegan. Longman, London. 1999. 1204 p.
- Biber D. *A typology of English texts* // *Linguistics*. Vol. 27. Pp. 3–43.
- Crossley S. A. et al. *Toward a new readability: A mixed model approach* / S. A. Crossley et al. // *Proceedings of the 29th annual conference of the Cognitive Science Society* / D. S. McNamara and G. Trafton (eds.). Cognitive Science Society, 2007. Pp. 197–202.
- DuBay W. H. *The Classic Readability Studies*. Costa Mesa, 2007. 240 p. URL: <https://files.eric.ed.gov/fulltext/ED506404.pdf> (Accessed 21 July 2024)
- Egins S. *An introduction to systemic functional linguistics*. London, 2004. 384 p.
- Fang Z., Schleppegrell M. J., Cox B. E. *Understanding the Language Demands of Schooling: Nouns in Academic Registers* // *Journal of Literacy Research*. 2006. Vol. 38. Iss. 3. Pp. 247–273. https://doi.org/10.1207/s15548430jlr3803_1
- Flesch R. *A new readability yardstick* // *Journal of Applied Psychology*. Vol. 32. 1948. P. 221–233.
- Gatiyatullina G. *Computing Russian Morphological distribution patterns using RusAC Online Server* / G. Gatiyatullina, M. Solnyshkina, V. Solovyev, A. Danilov, E. Martynova, I. Yarmakeev // *13th International Conference on Developments in eSystems Engineering (DeSE)*. 2020. Pp. 393–398. DOI: 10.1109/DeSE51703.2020.9450753
- Halliday M. A. K., Martin J. R. (eds.) *Writing Science: Literacy and Discursive Power*. London: Falmer Press, 2003. 434 p.
- Just M. A., Carpenter P. A. *The psychology of reading and language comprehension*. Boston; London: Allyn and Bacon. 1987. 518 p.
- Kincaid J. P., Fishburne R. P., Rogers R. L., Chissom B. S. *Derivation of new readability formulas (Automated Readability Index, Fog Count, and Fesch Reading Ease Formula) for Navy enlisted personnel* // *Research Branch Report*. 1975. Pp. 8–75.
- Koda K. *Insights into second language reading: A cross-linguistic approach*. Cambridge: Cambridge University Press. 2005. 320 p.
- Martin J. R. *Nominalization in science and humanities: Distilling knowledge and scaffolding text* // *Functional and Systemic Linguistics: Approaches and Uses* / Eija Ventola ed. Berlin, New York: Mouton de Gruyter, 1991. Pp. 307–338. <https://doi.org/10.1515/9783110883527.307>
- McLaughlin G. H. *SMOG Grading – A new readability formula* // *Journal of Reading*, 1969. Vol. 12. Iss. 8. Pp. 639–646.
- Solnyshkina M., Ivanov V., Solovyev V. *Readability formula for Russian texts: A modified version* // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2018. Pp. 132–145. DOI: 10.1007/978-3-030-04497-8_11
- Solnyshkina M. I., Kiselnikov A. S. *Text complexity: Study phases in Russian linguistics* // *Vestnik Tomskogo Gosudarstvennogo Universiteta, Filologiya*. 2015. Vol. 38 (6). Pp. 86–99. DOI: 10.17223/19986645/38/7
- To V., Fan S., Thomas D. P. *Lexical density and readability: A case study of English textbooks* // *The International Journal of Language, Society and Culture*. 2013. Vol. 37. Iss. 7. Pp. 61–71.
- Vahrusheva A., Solovyev V., Solnyshkina M., Gafiaytova E., Akhtyamova S. *Revisiting Assessment of Text Complexity: Lexical and Syntactic Parameters Fluctuations* // *Speech and Computer. SPECOM 2023. Lecture Notes in Computer Science* / A. Karpov, K. Samudravijaya, K. T. Deepak, R. M. Hegde, S. S. Agrawal, S. R. M. Prasanna, (eds). Springer, 2023. Vol. 14338. https://doi.org/10.1007/978-3-031-48309-7_35

References

- Valgina, N. S. (2003). *Teoriya teksta [Theory of text]*, Logos, Moscow, Russia. (In Russian)
- Jerebilo, T. V. (2010). *Slovar lingvisticheskikh terminiv [A Dictionary of linguistic terms]*. Issue 5, revised and added, Piligrim, Nazran, Russia. (In Russian)
- Kozhina, M. N. (1972). *O rechevoy sistemnosti nauchnogo stilya sravnitel'no s nekotorymi drugimi [On the speech systematicity of scientific style in comparison with some others]* Perm State University Publishing House, Perm, Russia. (In Russian)
- Mamontov, A. S., Mamontova, Ye. Ye. (2012). *Lingvostatisticheskiy analiz ispol'zovaniya deverbativov v razlichnykh vidakh tekstov russkogo i angliyskogo yazykov [Linguistic and statistical analysis of the use of deverbatives in various types of texts in Russian and English]*, *Vestnik RUDN*, No. 2., 107–110. (In Russian)
- Matskovskiy, M. S. (1976). *Problemy chitabel'nosti pechatnogo materiala [Problems of readability of printed material]*, in Dridze, T. M. and Leontyev, A. A. (eds.) *Meaningful perception of a speech message in the conditions of mass communication*, Nauka, Moscow, 126–142. (In Russian)
- Mikk, Ya. A. (1970). *O faktorakh ponyatnosti uchebnogo teksta [On the factors of understandability of educational text]*, Abstract of Ph.D dissertation, University of Tartu, USSR. (In Russian)
- Mitrofanova, O. D. (1973). *Yazyk nauchno-tekhnicheskoy literatury [Language of scientific and technical literature]*, Moscow University Press, Moscow, Russia. (In Russian)
- Oborneva, I. V. (2006). *Avtomatizirovannaya otsenka slozhnosti uchebnykh tekstov na osnove statisticheskikh parametrov [Automated assessment of the complexity of educational texts based on statistical parameters]*, Abstract of PhD dissertation, Moscow, Russia. (In Russian)
- Petrova, N. Ye. (2009). *Substantivatsiya i deadverbializatsiya narechiy v sovremennom russkom yazyke [Substantivization and deadverbialization of adverbs in the modern Russian language]*, *Vestnik Uralskogo gosudarstvennogo universiteta: Seriya gumanitarnye nauki*, 1/2 (63), 35–42. (In Russian)
- Sirotinina, O. B. (ed.) (2009). *Razgovornaya rech' v sisteme funktsional'nykh stiley sovremennogo russkogo literaturnogo yazyka. Grammatika: [Colloquial speech in the system of functional styles of the modern Russian literary language: Grammar]*, Moscow, USSR. (In Russian)
- Shpakovskiy, Yu. F. (2007). *Otsenka trudnosti vospriyatiya i optimizatsiya slozhnosti uchebnogo teksta (na materiale tekstov po khimii) [Assessing the difficulty of perception and optimizing the complexity of educational text (based on texts in chemistry)]*, Abstract of Ph.D dissertation, Minsk, Belarus. (In Russian)
- Alderson, J. Ch. (2000). *Assessing reading*, Cambridge University Press, London, UK. (In English)
- Banks, D. (2005). *On the historical origins of nominalized process in scientific text*, *English for Specific Purposes* 24/3, 347–367. DOI: 10.1016/j.esp.2004.08.002 (In English)
- Banks, D. (2008). *The Development of Scientific Writing: Linguistic Features and Historical Context*, Equinox, London, UK. (In English)
- Biber, D. and Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*, Cambridge University Press, Cambridge, UK. (In English)
- Biber, D., Gray, B. (2013). *Nominalising the verb phrase in academic scientific writing*, in Bas Aaarts et al. (Eds.), *The Verb Phrase in English*, Cambridge University Press, Cambridge, UK. (In English)
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (2021). *Grammar of spoken and written English*, John Benjamins, Amsterdam, Netherlands. <https://doi.org/10.1075/z232> (In English)
- Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *The Longman grammar of spoken and written English*, Longman, London, England. (In English)
- Biber, D. (1989). *A typology of English texts*, *Linguistics*, 27, 3–43. (In English)
- Crossley, S. A. et al. (2007). *Toward a new readability: A mixed model approach in McNamara, D. S. Trafton, G. (eds.) Proceedings of the 29th annual conference of the Cognitive Science Society*; Austin, TX: Cognitive Science Society, 197–202. (In English)
- DuBay, W. H. (2007). *The Classic Readability Studies*. Costa Mesa. 240 p.

URL: <https://files.eric.ed.gov/fulltext/ED506404.pdf> (Accessed 21 July 2024) (In English)

Eggins, S. (2004). *An introduction to systemic functional linguistics*, London, UK. (In English)

Fang, Z., Schleppegrell, M. J. and Cox, B. E. (2006). Understanding the Language Demands of Schooling: Nouns in Academic Registers, *Journal of Literacy Research*, 38 (3), 247–273.

https://doi.org/10.1207/s15548430jlr3803_1 (In English)

Flesch, R. (1948). A new readability yardstick, *Journal of Applied Psychology*, 32, 221–233. (In English)

Gatiyatullina, G., Solnyshkina, M., Solovyev, V., Danilov, A., Martynova, E. and Yarmakeev, I. (2020). Computing Russian Morphological distribution patterns using RusAC Online Server, *13th International Conference on Developments in eSystems Engineering (DeSE)*, Liverpool, UK, 393–398. DOI: 10.1109/DeSE51703.2020.9450753 (In English)

Halliday, M. A. K. and Martin, J. R. (eds.) (2003). *Writing Science: Literacy and Discursive Power*, Falmer Press, London, UK. (In English)

Just, M. A., Carpenter, P. A. (1987). *The psychology of reading and language comprehension*, Allyn and Bacon, Boston, London. (In English)

Kincaid, J. P., Fishburne, R. P., Rogers, R. L. and Chissom, B. S. (1975). Derivation of new readability formulas (Automated Readability Index, Fog Count, and Fesch Reading Ease Formula) for Navy enlisted personnel, *Research Branch Report*, 8–75. (In English)

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*, Cambridge University Press, Cambridge, UK. (In English)

Martin, J. R. (1991). *Nominalization in science and humanities: Distilling knowledge and scaffolding, Functional and systemic linguistics: Approaches and Uses*, De Gruyter Mouton, Berlin, New York, 307–338. <https://doi.org/10.1515/9783110883527.307> (In English)

McLaughlin, G. H. (1969). SMOG Grading – A new readability formula, *Journal of Reading*, 12(8), 639–646. (In English)

Solnyshkina, M., Ivanov, V., Solovyev, V. (2018). Readability formula for Russian texts: A

modified version in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 132–145. DOI: 10.1007/978-3-030-04497-8_11 (In English)

Solnyshkina, M. I., Kisel'nikov, A. S. (2015). Text complexity: Study phases in Russian linguistics, *Vestnik Tomskogo Gosudarstvennogo Universiteta, Filologiya*, 38 (6), 86–99. DOI: 10.17223/19986645/38/7 (In English)

To, V., Fan, S., Thomas, D. P. (2013). Lexical density and readability: A case study of English textbooks. *The International Journal of Language, Society and Culture*, 37(7), 61–71. (In English)

Vahrusheva, A., Solovyev, V., Solnyshkina, M., Gafiyatova, E., Akhtyamova, S. (2023). Revisiting Assessment of Text Complexity: Lexical and Syntactic Parameters Fluctuations, in Karpov, A., Samudravijaya, K., Deepak, K. T., Hegde, R. M., Agrawal, S. S., Prasanna, S. R. M. (eds) *Speech and Computer. SPECOM 2023. Lecture Notes in Computer Science*, Vol 14338, Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-031-48309-7_35 (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.

Галия Маратовна Гатиятуллина, старший преподаватель кафедры теории и практики преподавания иностранных языков, младший научный сотрудник, НИЛ «Текстовая аналитика», Институт филологии и межкультурной коммуникации, Казанский (Приволжский) федеральный университет, Казань, Россия.

Galiya M. Gatiyatullina, Senior Lecturer, the Department of Theory and Practice of Teaching Foreign Languages, Research Assistant, Text Analytics Research Laboratory, Institute of Philology and Intercultural Communication, Kazan Federal University, Kazan, Russia.

Эльзара Васильевна Гафиятова, доктор филологических наук, доцент, заведующая кафедрой теории и практики преподавания иностранных языков, заместитель директора по международной деятельности и развитию Института филологии и межкультурной коммуникации, Казанский (Приволжский) федеральный университет, Казань, Россия.

Elzara V. Gafiyatova, Doctor of Philology, Associate Professor, Head of the Department of Theory and Practice of Teaching Foreign Languages, Vice Director of the Institute of Philology and Intercultural Communication, Kazan Federal University, Kazan, Russia.

Андрей Владимирович Данилов, кандидат педагогических наук, доцент кафедры билингвального и цифрового образования, старший научный сотрудник НИЛ «Текстовая аналитика», Институт филологии и межкультурной коммуникации, Казанский (Приволжский) федеральный университет, Казань, Россия.

Andrei V. Danilov, PhD in Education, Associate Professor, Department of Bilingual and Digital Education, Senior Researcher at the Text Analytics Research Laboratory, Institute of Philology and Intercultural Communication, Kazan Federal University, Kazan, Russia.

UDC 81'322

DOI: 10.18413/2313-8912-2024-10-3-0-5

Olga V. Dekhnich¹ 
Tatyana A. Litvinova² 

A keystroke analysis to study writing:
a bibliometric review using R and VOSviewer

¹ Belgorod State National Research University,
85 Pobedy St., Belgorod, 308015, Russia
E-mail: dekhnich@bsu.edu.ru
ORCID: 0000-0001-6088-2656

² Voronezh State Pedagogical University,
86 Lenin St., Voronezh, 394043, Russia
E-mail: centr_rus_yaz@mail.ru
ORCID: 0000-0002-6019-3700

Received 14 August 2024; accepted 15 September 2024; published 30 September 2024

Acknowledgements: Tatiana A. Litvinova acknowledges the support of the Ministry of Education of the Russian Federation (the research was supported by the Ministry of Education of the Russian Federation within the framework of the state task in the field of science, topic number QRPK-2024-0011). Olga V. Dekhnich received no financial support for the research, authorship, and publication of this article.

Abstract: Data obtained by means of keystroke logging software, which records all writing events, i.e., all keypresses with their time stamps, are widely used in different research fields and practical applications – from early detection of cognitive impairments to user authentication. Keystroke logging software have become especially popular in writing studies since it allows researchers to observe writing processes unobtrusively and increases their understanding of what happens “behind the scenes” of text production. This methodology has been used in writing studies since the advent of affordable computers in the 1990s, however, there has still been no bibliometric review that would provide a holistic picture of this field and identify hot topics and trends, perform citation analysis and identify the most productive authors and countries in the field of writing research using keystroke data. This paper is pioneer research to conducting a bibliometric analysis on the field “keystroke analysis to study writing” aimed at filling in the gap. A search was conducted in the bibliographic database Scopus on July 11, 2024. We included studies published in English post-2000 that discuss the use of keystroke data and methodology to study writing process. This review followed the guidelines of the PRISMA protocol to perform the study search and selection. The search yielded 336 documents of which 273 met out inclusion criteria. The records retrieved were analysed using the bibliometrix R-package and VOSviewer software. Using these tools in combinations, we implemented both performance analysis which examines the contributions of research constituents to a given field and science mapping which is aimed at revealing the relationships between research constituents. The contribution of this study is twofold. Firstly, it provides an in-depth bibliometric analysis of an actively developing field of research which is rather limited in terms of the countries and institutions involved (and the languages of the texts analysed), hopefully saving time and effort for

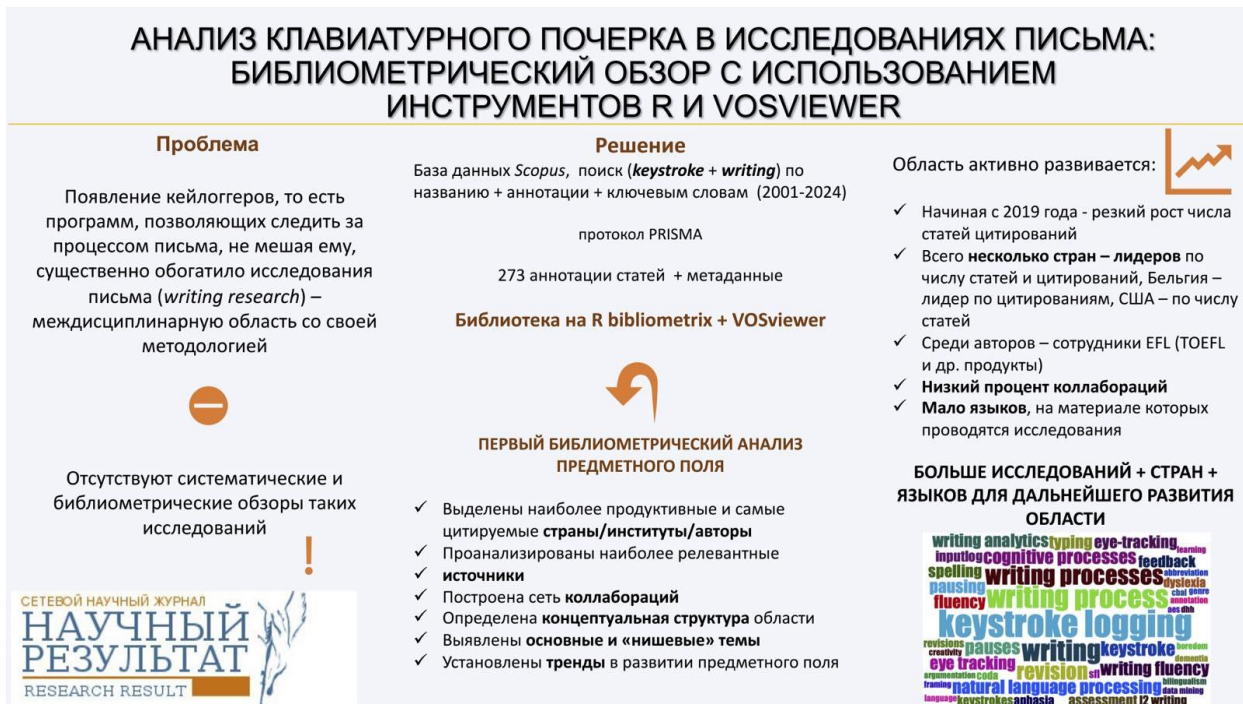
Информация об источниках финансирования или грантах: Т.А. Литвинова благодарит за финансовую поддержку Министерство просвещения Российской Федерации (работа выполнена при финансовой поддержке Министерства просвещения Российской Федерации в рамках выполнения государственного задания в сфере науки, номер темы QRPK-2024-0011). О.В. Дехнич не получала финансовой поддержки за выполнение исследований, написание и публикацию статьи.

Аннотация: Данные клавиатурного почерка, то есть длительность и количество нажатий клавиш, которые регистрируются с использованием программ-кейлоггеров, широко используются в различных областях знаний, в том числе для решения прикладных задач – от ранней диагностики когнитивных нарушений до идентификации пользователя. Также академические кейлоггеры и извлекаемые с их помощью данные активно используются в такой междисциплинарной области, как исследования письма (*writing research*), поскольку позволяют наблюдать за процессом создания текста, не вмешиваясь в него. Подобные исследования ведутся начиная с появлением общедоступных персональных компьютеров, то есть с конца 1990-х годов, однако библиометрический анализ таких работ, то есть анализ, направленный на выявление общей картины рассматриваемого предметного поля, определение его концептуальной структуры, наиболее активно исследуемых тем, наиболее цитируемых работ и авторов, до настоящего времени отсутствовал. Данная статья представляет собой первый опыт библиометрического анализа научных работ, направленных на анализ процесса создания текста с использованием данных кейлоггеров. Мы анализировали работы, индексируемые в базе данных Scopus и опубликованные в период с 2001 по 2024 год (273 работы было отобрано с использованием протокола PRISMA). В качестве инструментов исследования нами были выбраны библиотека на языке R *bibliometrix* и программа VOSviewer. С использованием данных инструментов нами были определены наиболее продуктивные и цитируемые исследователи, институты и страны, а также наиболее значимые работы. На основе различных методов анализа совместной встречаемости ключевых слов, реализованных в данных инструментах, нами была построена концептуальная структура предметного поля, а также определены исследовательские тренды. Значимость настоящей работы определяется двумя факторами. Во-первых, она содержит результаты первого библиометрического анализа важной и активно развивающейся предметной области, ограниченной, однако, с точки зрения географической принадлежности ведущих исследователей и языков, на материале которых проводятся такие исследования. Подобный анализ, как представляется, позволит ознакомить с состоянием работ в данной области тех ученых, для которых она является новой. Во-вторых, данное исследование представляет собой пример комбинации методов автоматического анализ научной литературы, которая может быть применена для анализа иных предметных областей.

Ключевые слова: Исследования письма; Клавиатурный почерк; Библиометрический анализ; Библиотека *bibliometrix* на языке R; VOSviewer; Анализ цитирований; Концептуальная структура предметного поля; Тематическое развитие предметного поля; PRISMA; Scopus; Ключевое слово.

Информация для цитирования: Дехнич О. В., Литвинова Т. А. Анализ клавиатурного почерка в исследованиях письма: библиометрический обзор с

использованием инструментов R и VOSviewer // Научный результат. Вопросы теоретической и прикладной лингвистики. 2024. Т. 10. № 3. С. 91-115. DOI: 10.18413/2313-8912-2024-10-3-0-5



1. Introduction

Typing has been the most comfortable and typical way to produce texts for many people, that is why studying typed texts and the process of their creation has become one of the important research directions in modern linguistics. The development of technical devices such as keystroke loggers, which timestamp each event (i.e. keystroke) during writing, has provided researchers with previously unavailable data reflecting different aspects of writing processes.

The use of keystroke logging data is typical not only to study writing process in linguistics but also for many practical applications, especially for user authentication and identification. The search for the keyword "keystroke" in the Scopus database, carried out in July 2024, yields more than 1700 papers – search by title, keyword and abstract – more than 4000 papers published in journals and proceedings from various subject areas – from computer science (2755 papers) to arts and humanities (365) and medicine (306).

There are several review papers which analyze different aspects of keystroke behavior: as a tool for cognitive impairments detection (Lee and Park, 2024), a digital biomarker for understanding neurocognitive functioning (Nguyen et al., 2023), motor decline (Alfalahi et al., 2022), as a tool for emotion recognition (Maalej and Kallel, 2020; Yang and Qin, 2021), user authentication (Quraishi and Bedi, 2018; Raul et al., 2020).

However, despite the rise of the number of papers which use keystroke data to study writing, to the best of our knowledge, there are no systematic or bibliometric reviews of these papers which would consider this actively developing research area from a holistic perspective with identification of the main topics and research trends.

The aim of this paper is to map the literature devoted to the study of writing using keystroke data through a bibliometric approach in order to reveal knowledge structures, priority topics and trends in this area of research.

We performed a bibliometric analysis of the papers considering the keystroke logging method for study writing searched in Scopus using the R package bibliometrix (Aria and Cuccurullo, 2017) and the standalone application VOSviewer (Van Eck and Waltman, 2014). This bibliometric analysis of 273 articles published over the last 24 years (2001-2024) is the first one to present comprehensive and statistical insight into the research status and trends on the use of the keystroke methodology to study writing.

The study's findings may offer broad perspectives on the current state of this research area and provide some guidance for future research.

Therefore, this study was driven by the following research questions.

1. What was the annual research trend in the considered subject field?
2. Which journals, countries, institutions and authors were the major research contributors?
3. What were the main research topics of the topical collection?

4. What are the trends in this research area?

2. Research methods

2.1. Transparency Statement

To address our research questions, we undertook a bibliometric review to map out the existing literature on our selected topic – using keystroke data and software to study the writing process. We adhered to the PRISMA protocol for conducting this review. A protocol was outlined before commencing the study. We ensure that this manuscript presents a truthful, precise, and complete report of the research conducted.

2.2. Eligibility Criteria

This bibliometric review was executed through a structured process that included: 1) formulating the research question; 2) identifying relevant literature; 3) selecting appropriate studies for inclusion; 4) extracting key data from these studies and data cleaning; 5) summarising and presenting the findings.

The selection criteria for the literature were divided into several categories for framing research questions in bibliometric reviews (detailed in Table 1).

Table 1. Eligibility criteria

Таблица 1. Критерии отбора

Criterion	Inclusion	Exclusion	Justification
Population	All the studies which use keystroke data and software to study the writing process	All the studies outside the defined field	As the research topic we analyze is multidisciplinary, we include papers from different subject areas related to the topic of writing and its analysis using keystroke data/software
Concept	Studies (journal articles, book chapters, conference proceedings) which discuss different aspects of the writing process based on keystroke data	Studies which do not consider the concept of the writing process	The purpose of the review is to analyze literature which is related to the keystroke data/software for the writing process understanding
Context	Studies that relate to	Studies outside	Focus of this review is

	writing research	writing research	on the use of the keystroke data/method to study writing
Language	English	Non-English	English is a lingua Franca of scientific communication
Time period	2001-2024	Before 2001	Keystroke logging software has become widely used in writing studies after the appearance of available personal computers
Types of sources	1. Journal articles 2. Book chapters 3. Conference papers	1. Non peer-reviewed sources 2. Reviews (all types) 3. Editorials	Only peer-reviewed sources and original research papers were included in our analysis
Geographical location	Any location	None	Getting international perspective

We included all study designs from any discipline found in our search, reported in English. Special attention was given to the quality of the sources included in review: we considered only original research and peer-reviewed papers.

2.3. Information Sources and Search Strategy

Scopus, which is considered the largest citation and abstract database covering a wide range of subjects was chosen as the data source for this study. Its particular suitability for bibliometric analysis (consistently reliable format, and ease of data extraction) has been highlighted by many researchers (see ref. in Rejeb et al., 2023). It is also more comprehensive than, for example, PubMed or Web of Science (WoS) and more precise than Google Scholar because of its superior control over content, that is why it is widely used for bibliometric analysis (see ref. in Rejeb et al., 2023).

Access to Scopus was performed using the personal account of one of the authors who was granted access to Scopus as part of the reviewer recognition process.

Figure 1 illustrates our compliance with the PRISMA protocol.

The following search string was used in the article's title, abstract and keywords fields to extract data from Scopus: "keystroke*" AND "writing".

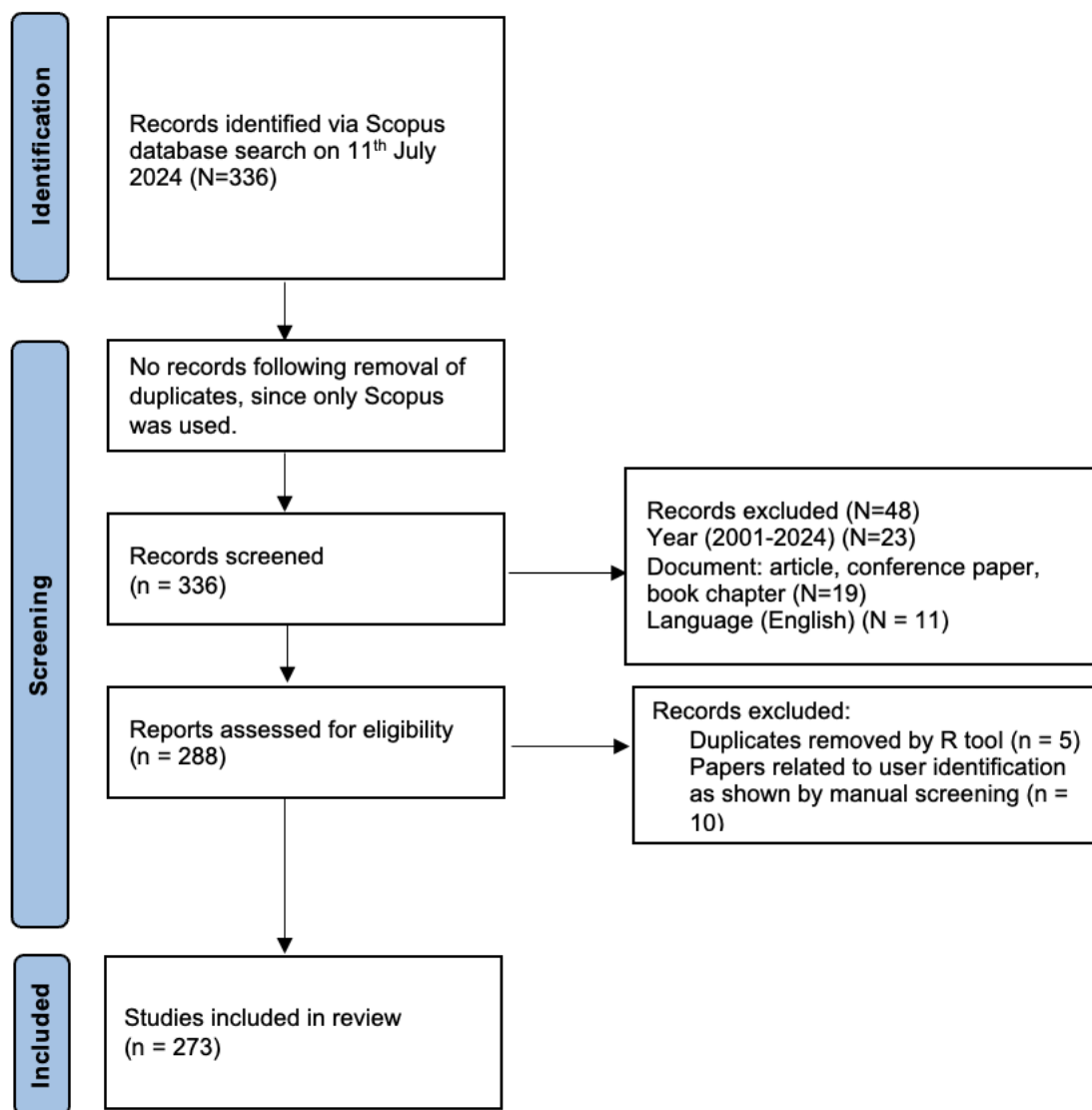
According to the information of Scopus, the first article which meets the search criteria was published in 1974. However, a wide use of the keystroke methodology for writing research started after the development of freely available keystroke loggers, i.e., after 1990s (Wengelin and Johansson, 2023). In addition, only since the early 2000s, the academic landscape has been reshaped by a global trend toward requiring publications in prestigious journals indexed in international citation databases, therefore, we decided to refine our search by excluding papers published before 2001.

The data collection process for this study started on July 11, 2024; thus, we set the scope of this research is from 2001 to 2024.

From this search, 336 studies were retrieved (as shown in Figure 1).

Figure 1. PRISMA flow diagram

Рисунок 1. Диаграмма отбора работ для анализа по протоколу PRISMA



2.4. Selection of Sources of Evidence

For further analysis, all available bibliographic information for 336 retrieved records were extracted from Scopus and exported in CSV format. Through the initial screening of titles and abstracts, performed by two reviewers, 48 studies were excluded based on the predefined inclusion criteria: we excluded papers published before 2001 (N=23), written in languages other than English (N=11), and of any other types besides “article”, “conference paper”, “book chapter”.

Then we removed duplicates detected by bibliometrix (N=5). Next, remaining titles

and abstracts of the publications were manually evaluated for relevance in terms of context and concept compliance. After this screening, 10 records were removed since they did not contribute to the field of writing studies, i.e., their findings were not related to the writing process understanding (we noticed that they are related to user authentication).

Thus, 273 papers were selected for the final analysis (see Figure 1).

Data Charting Process

Data extraction was conducted by two independent reviewers, with each being assigned all the selected documents for initial analysis. The complete data set extracted by

one reviewer was then cross-checked by the other to ensure accuracy and consistency. Any differences between reviewers were addressed and resolved through consensus meetings.

Data analysis and visualisation

The metadata of the articles included in the review were processed using VOSviewer, a software tool designed for constructing and visualising bibliometric networks and the bibliometrix R-package.

Bibliometric analysis involves two main approaches: performance analysis and science mapping. The goal of performance analysis is assessing the contributions of research constituents to a given field (it relies upon such metrics as total number of publications, citations, etc.), while science mapping examines the relationships between research constituents via such methods as co-citation analysis, bibliographic coupling, co-word analysis (Donthu et al., 2021).

Bibliometrix is far less popular than VOSviewer, however, it provides a wide range of statistical techniques and visualizations useful for determining and visualizing the conceptual structure of the field. Bibliometrix enables users to conduct

both performance analysis and science mapping, whereas users are only able to perform science mapping in VOSviewer. However, we combined VOSviewer with bibliometrix to get a deeper understanding of the research area under consideration, since they provide different types of analysis and visualizations for science mapping. Many researchers have used them simultaneously (Saini et al., 2022; Wang and Gu, 2022).

3. Results and Discussion

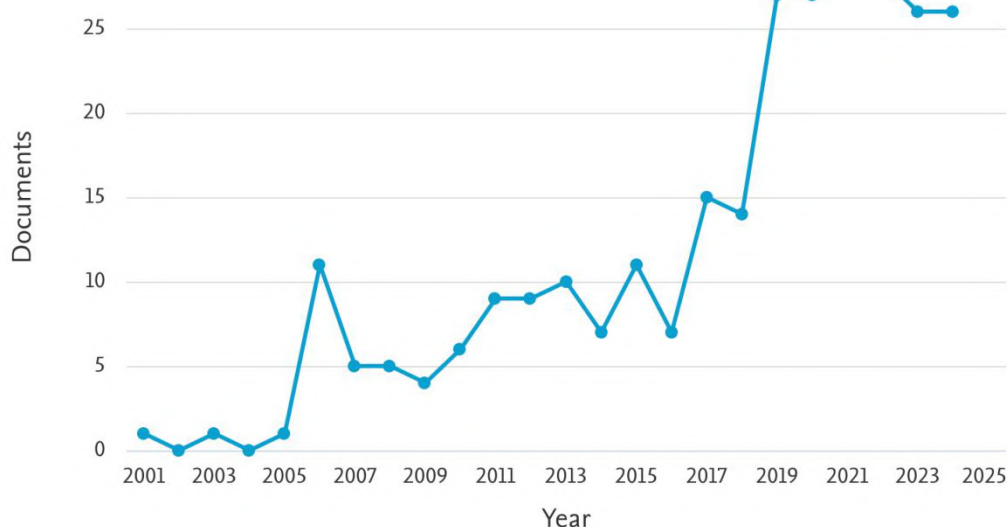
3.1. Performance analysis

3.1.1. General performance characteristics of the research area

Our dataset consists of 273 papers written by 520 authors and published in 152 sources. The annual growth rate is 15.22 %, i.e., the research in the analyzed field has attracted continuous attention of scholars. Figure 2 shows a strong increase in the number of papers from 2019 onwards (from a maximum of 15 to 26 in 2019, and at least 26 after 2019), which is an important index indicating a development trend of a research topic in a specific period of time (Farrukh et al., 2023).

Figure 2. Annual scientific production

Рисунок 2. Распределение числа публикаций по годам



It is noteworthy that by mid-2024, 26 papers published in 2024 were already recorded in Scopus, so we can expect a further increase in the number of papers this year.

Most of the works in our collection (64 %) are articles, conference papers constitute about 24 %, the rest of the papers are book chapters.

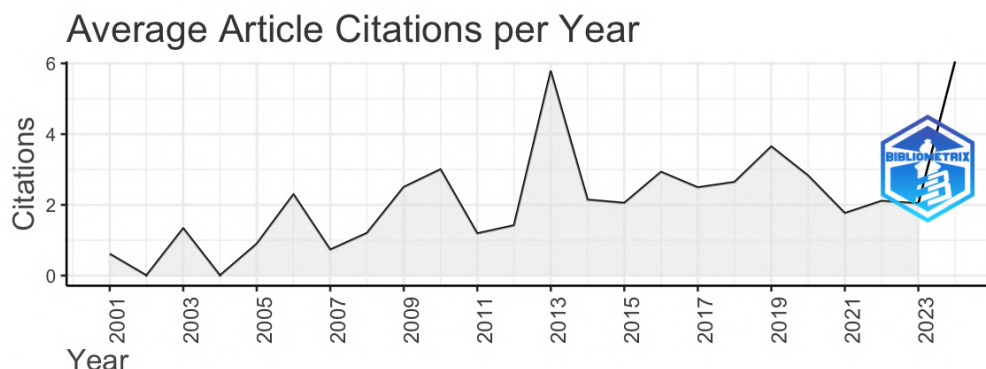
The average citations per doc is 14.7, the average citations per year per doc is 1.906, the overall number of citations of works in our collection is 4110.

Figures 3-4 show that the burst of

average citation count per article per year falls on the papers published in 2013 (they were cited around 6 times). This year, the most cited paper in our collection has been published – *Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes* (Leijten and Van Waes, 2013) (319 citations) which presents the results of analysis of the writing process with one of the most popular keyloggers designed specially for writing research – Inputlog with other observational techniques: thinking aloud protocols and eye-tracking data.

Figure 3. Average article citations per year

Рисунок 3. Усредненное число цитирований статей, изданных в данном году (на статью)



The mean total article citations per year at the highest level again can be seen for the papers published in 2013 (N= 63), which indicates that the articles published in 2013 were cited the maximum number of times although the number of articles in 2013 was rather low (N= 9). As we mentioned earlier, in 2013 the most cited paper of the collection was published. Along with it, in the top-10 most cited papers of the collection (Figure 5) there are two papers published during this year with the total of 159 citations. These papers are related to the constructing of documents using multiple digital sources (Leijten et al., 2013b) and detecting boredom and engagement during writing (Bixler and D’Mello, 2013).

The other peaks in the average total

citations per year fall on 2006 and 2010 when the other papers from the top-10 most cited were published: the 2006 paper *Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL* (Stevenson et al., 2006) with 106 citations, 2006 paper *Examining Pauses in Writing: Theory, Methods and Empirical Data* (Wengelin, 2006) with 91 citations and a 2010 paper *Cognitive illusions of authorship reveal hierarchical error detection in skilled typists with 109 citations* (Logan and Crump, 2010) (2nd most cited paper in the collection). It is also should be noted that the papers published in the last years have also been actively cited.

The documents in our collection have been increasingly cited since 2019 (see Figure 6).

Figure 4. Average total article citations per year

Рисунок 4. Усредненное общее число цитирований статей, изданных в данном году (по годам)

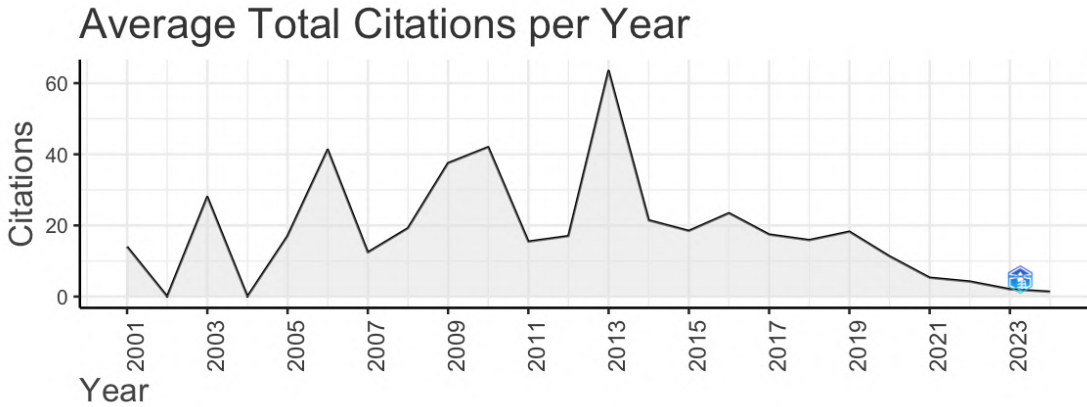


Figure 5. Most cited papers of the collection

Рисунок 5. Наиболее цитируемые работы

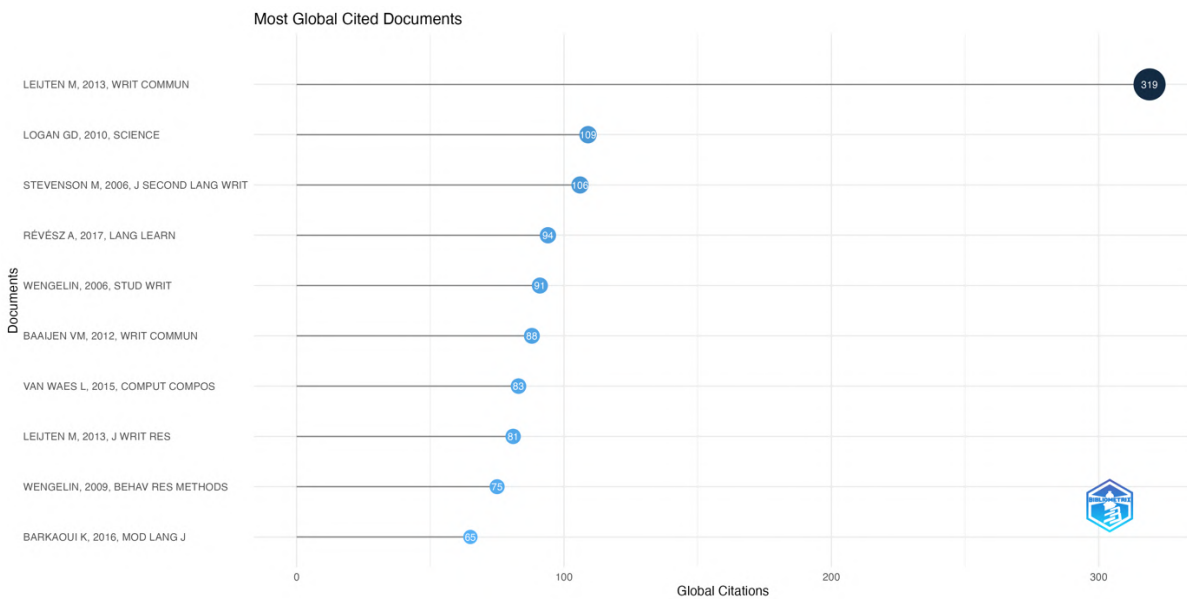
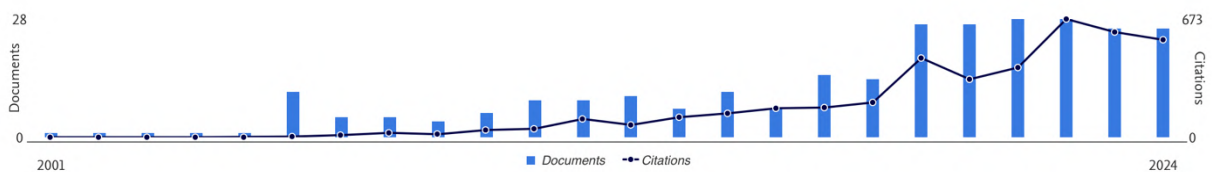


Figure 6. Number of papers and total number of citations by year

Рисунок 6. Число статей и цитирований по годам публикации



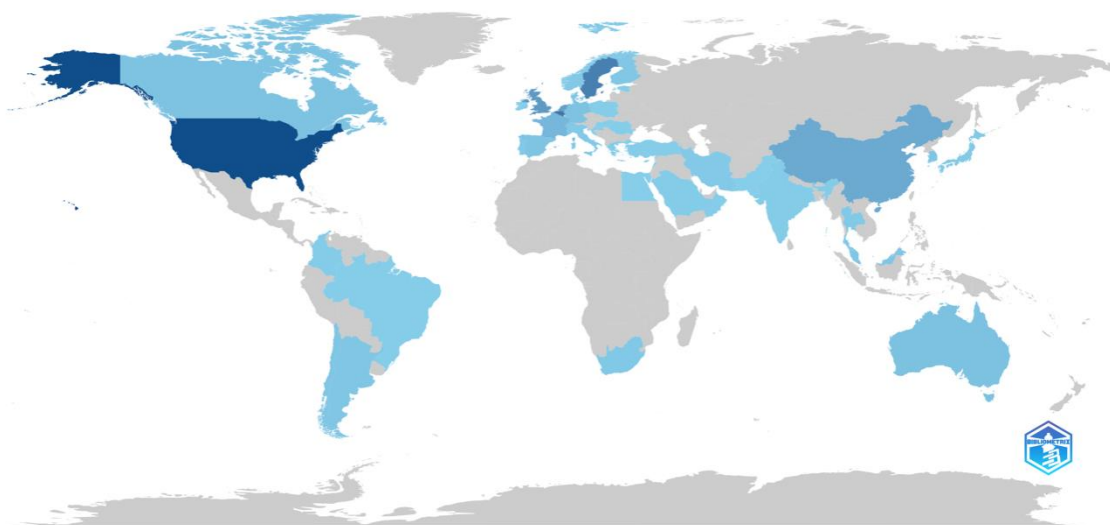
Overall, we see a general trend of increasing numbers of papers from 2019 onwards, and the papers in our collection have been actively cited since this year. We can safely conclude that the use of keystroke analysis in the study of writing is an area of research which is in a state of active development.

3.1.2. *Most productive countries, institutions and authors* Figure 7 shows the most productive countries in terms of the number of corresponding authors affiliated with these countries (this visualisation is constructed using the bibliometrix package).

Figure 7. Most productive countries (according to the corresponding author)

Рисунок 7. Страны с наибольшим числом статей (по корреспондирующему автору)

Country Scientific Production



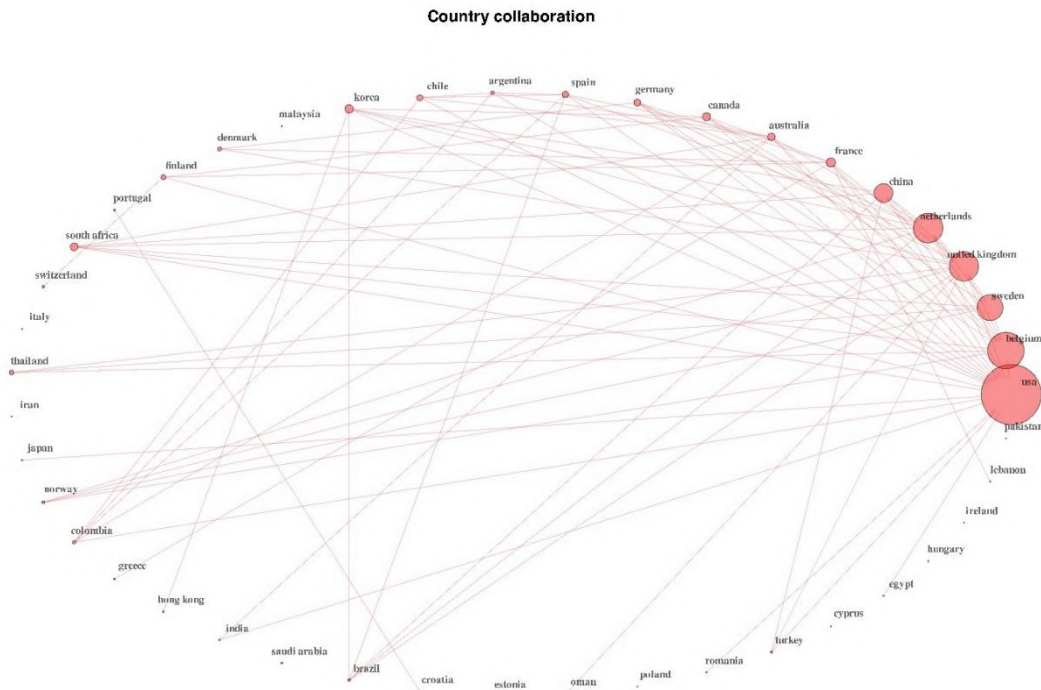
The USA is the leading country in terms of total number of papers (38 papers) and percentage of single-country publications (about 80%). **Sweden** is the second most productive country with 23 papers, 75% of which are single-country publications. **Belgium** is the third most productive country with 19 papers (12 of which are single-country). Authors from **the Netherlands** published 14 papers, but most of them (12)

are multi-country publications. Researchers from **China** and the b published 13 papers, of which 9 and 7 were single-country papers respectively. The number of papers related to other countries is rather low (**Canada** – 5, **France** – 5, **Spain** – 4, **Switzerland** – 4, **Thailand** – 4, others – 3 or less).

Figure 8 shows the network of the country collaboration.

Figure 8. Country collaboration in the research area under consideration

Рисунок 8. Сеть стран, имеющих коллаборации друг с другом в рамках обозреваемого предметного поля



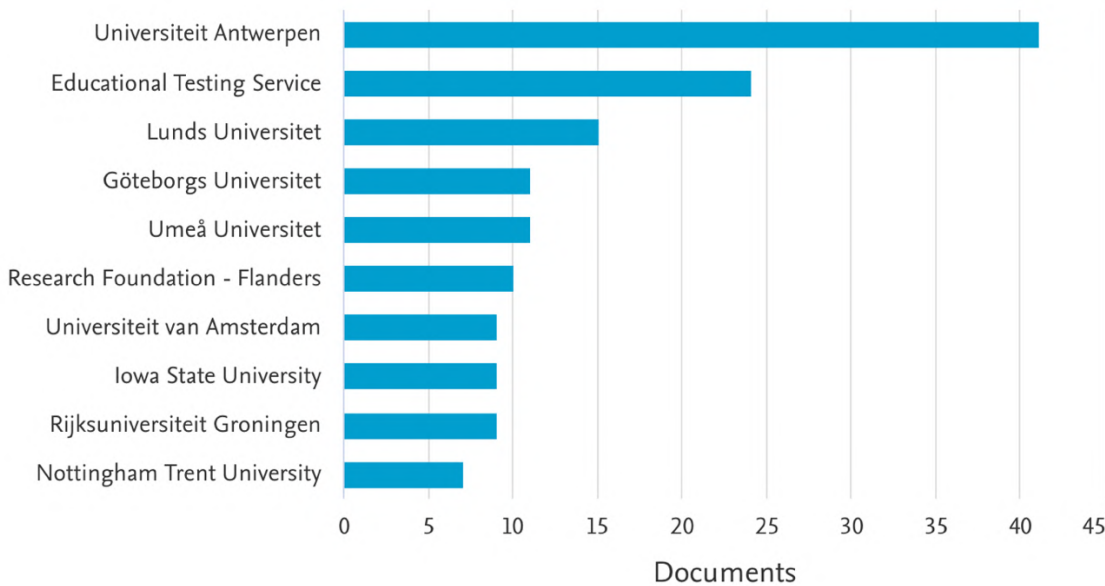
The network consists of the nodes representing a country, and the connection between the nodes represents a cooperative relationship between the countries, and the size of the nodes represents the numbers of cooperation with other countries. The larger the node is, the more international the cooperation the country participates in is. We can see that the USA has the largest diversity in terms of cooperation with other countries. Belgium, the Netherlands, the UK and Sweden are the countries with a moderate level of diversity of international cooperation.

Figure 9 shows the most productive institutions, with University of Antwerp (**Belgium**) being the leader. The second most productive organization is the Educational

Testing Service (ETS), which is the world's largest private educational testing and assessment organization which developed TOEFL and other products with its headquarters in the US. The third most productive organization is Lund University (Sweden).

The analysis of the dynamics of affiliation production shows (not shown due to lack of space) that the University of Antwerp has the most pronounced trend in increasing the number of papers starting from 2009, with a sharp increase in 2021. A continuously high number of papers is affiliated to Lund University from 2010 onwards.

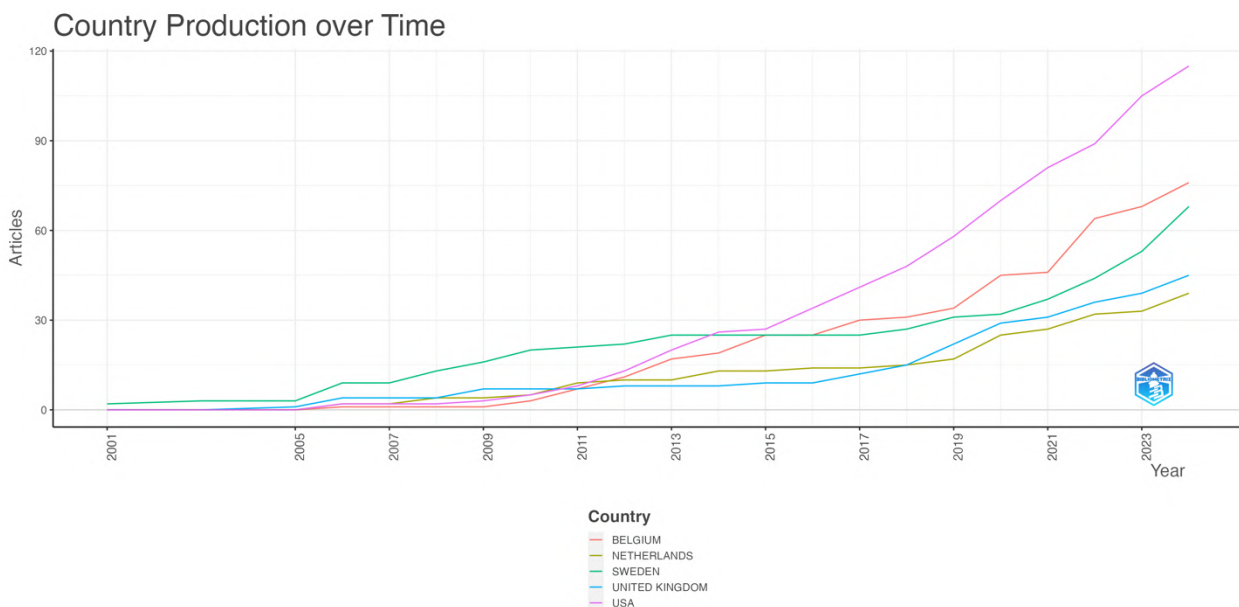
Figure 9. Most productive institutions (according to the corresponding author)
Рисунок 9. Организации с наибольшим числом статей (по корреспондирующему автору)



If we look at the country production over time (see Figure 10), we can easily observe a general trend of an increase in the number of papers for all the most productive

countries. In the US, this increase occurred earlier (from 2015) than in the other countries (since 2019) and is a sharper one.

Figure 10. Country production over time
Рисунок 10. Распределение числа статей по странам (по годам)

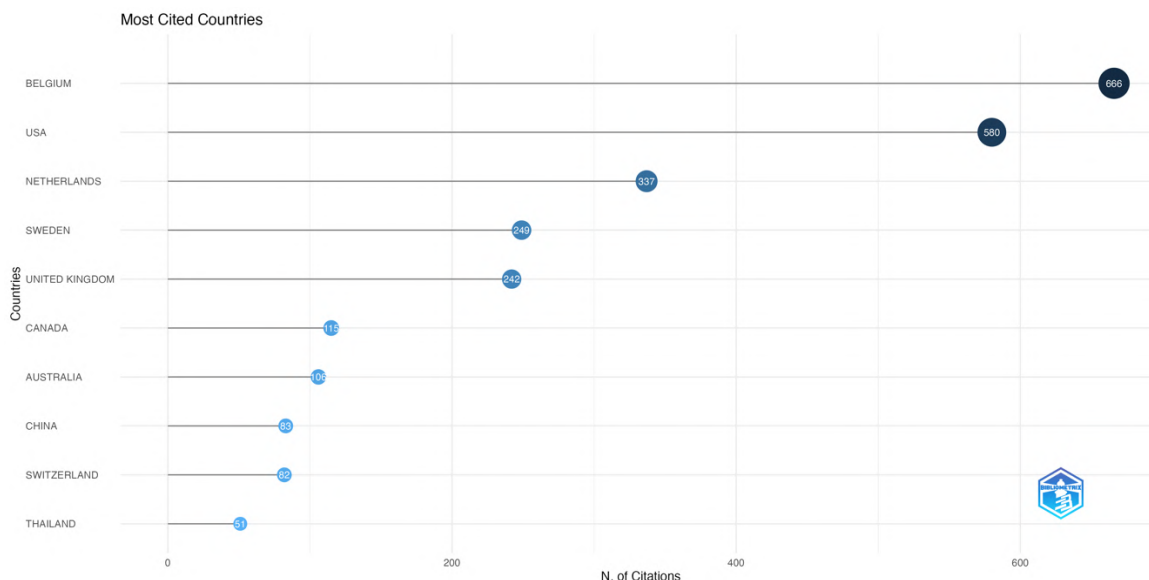


Although the USA is the most productive country, it ranks 2nd in terms of the number of citations (580), with Belgium

being the leader with 666 citations. The Netherlands is the third one in this rank (Figure 11).

Figure 11. Most cited countries

Рисунок 11. Страны с наибольшим числом цитирований



There are 522 different authors in our collection, with 36 authors of single-authored document and 866 author appearances. There are 44 single-authored papers, on average, there are 3.16 co-authors per document.

The most productive author, who published 27 papers, is **Luuk Van Waes** from the University of Antwerp, Belgium. The second most productive author, **Mariëlle Leijten**, is affiliated with the same university. **Åsa Wengelin**, the third productive author, is affiliated with the University of Gothenburg (Sweden). Mo Zhang, the author of 17 papers, is affiliated with the Educational Testing Service (USA), just like Paul Deane, her

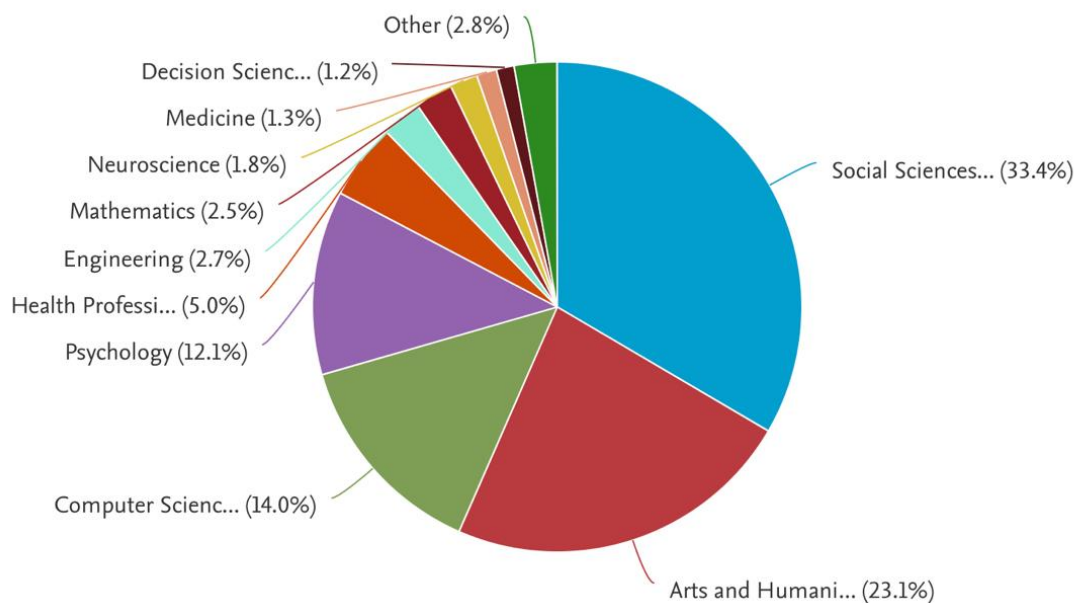
co-author. Luuk Van Waes and Mariëlle Leijten also have the highest h-index among the most productive authors (14), Paul Deane and Åsa Wengelin have h-index=10, Mo Zhang – 9. We can see that the productivity correlates with h-metrics in our collection.

3.1.3. The analysis of the most relevant sources

Figure 12 shows the distribution of sources by subject area. We can see that social sciences, arts and humanities, computer science and psychology are the leading subject areas of articles and proceedings in which papers from our collections have been published.

Figure 12. Paper distribution over a subject area

Рисунок 12. Распределение публикаций по предметным полям



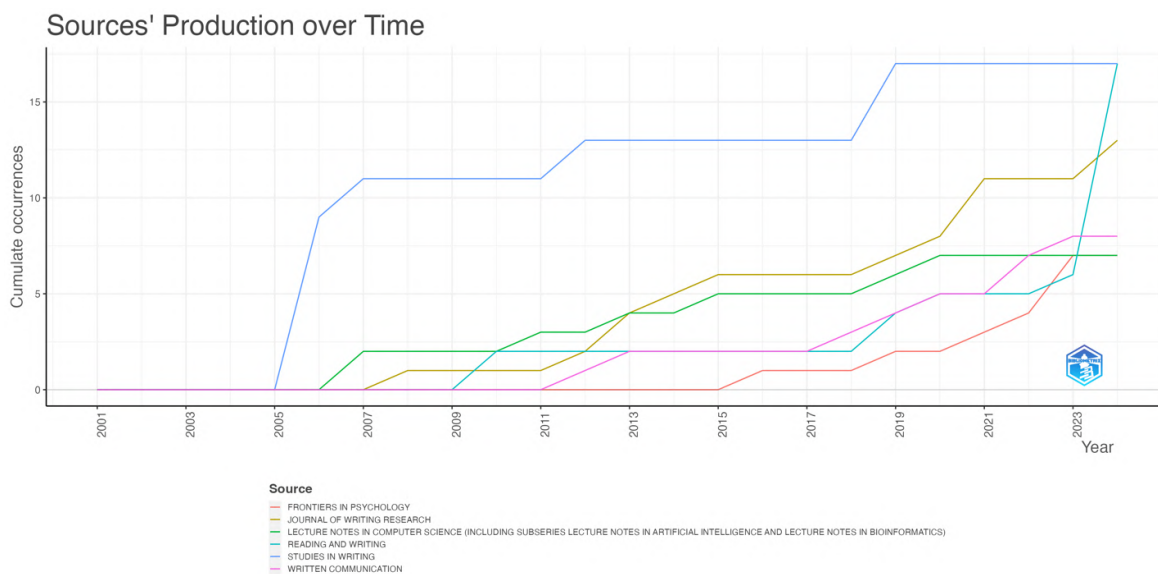
The most relevant sources in terms of the number of published papers from our collection is Reading and Writing (17 papers), Studies in Writing (17 papers), Journal of Writing Research (13), Written Communication (8), Frontiers in Psychology (7), Lecture Notes in Computer Science (7), Computers and Composition (6), ETS Research Report Series (5), Journal of Second Language Writing (5), Languages (5).

All these journals are highly reputable. The journals with the highest impact

(h-index) are Studies in Writing (11), Journal of Writing Research (8), Reading and Writing (6).

Looking at the production of sources over time (Figure 13), we could see that Studies in Writing has been continuously publishing papers on the study of writing using keystroke since 2007, while the other most productive sources show a general trend of an increasing number of such papers after 2019.

Figure 13. Source production over time (over the years)
Рисунок 13. Распределение статей по источникам (по годам)



3.2. Science mapping

3.2.1. Keyword co-occurrence analysis

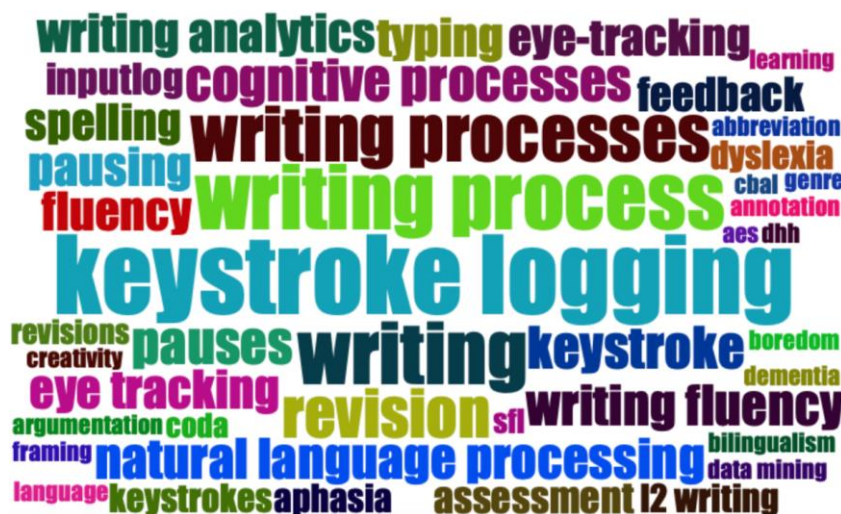
Keyword analysis is important in a bibliometric review as it could provide many insights about the conceptual structure of the field. Keywords are divided into authors' keywords and those which are assigned automatically (they are called "keyword plus" in Scopus). There are 801 "keyword plus" in our dataset (the most frequent ones are WRITING – 53 papers, HUMAN – 28, MALE – 26, FEMALE – 25, WRITING PROCESS – 19, STUDENTS – 18), 750 author's keywords (the top 100 most frequent

ones are presented in the form of a word cloud in Figure 33). As a rule, authors' keywords are more specific, so they will be used in the further analysis.

Figure 14 gives us a snapshot of the content of our collection (the bigger the size of the circles is, the more frequent the keyword is), with KEYSTROKE LOGGING, WRITING PROCESS, WRITING, COGNITIVE PROCESSES, REVISION, PAUSES, NATURAL LANGUAGE PROCESSING are the most relevant author's keywords.

Figure 14. Word cloud of 100 most frequent author's keywords of the collection

Рисунок 14. Облако слов, отражающее встречаемость 100 самых частотных ключевых слов автора в коллекции



Co-word analysis, which reveals words that frequently co-occur in documents, can be used to reveal important themes (both established and emerging) within the focal domain. Using VOSviewer, we analyzed co-occurrences of authors' keywords and visualized them in the form of a network (Figure 15) (we only included words that appeared at least 10 times in our dataset). Keywords with similar colors belong to the same cluster. The size of each circle in the cluster shows the proportion of citations for that authors' keywords. Larger circles and map labels indicate a greater significance and relevance of the keywords.

We used association strength as a method of normalization and set a minimal cluster size to 10.

Cluster 1 (**red**) is the largest one in terms of the number of terms (24). It is related to the automated text assessment including second language writing using keystroke logs and various modern approaches such as scenario-based assessment (SBA), which have recently been used to assess writing skills.

Cluster 2 (**green**, 22 terms) is related to the analysis of typed texts produced by people with mild to severe (Alzheimer, dementia) cognitive problems based on keystroke analysis, especially revision process analysis,

sentence history analysis, assessment of motor and typing skills using techniques for automated scoring for writing quality based on the deep learning method and AI in general. Methods of mixture models have widely been used in such papers to distinguish between the healthy and the ill respondents in terms of pauses duration.

Cluster 3 (**dark blue**, 21 items) is related to the study of academic writing, including writing in a second language and using an analysis of revision via keystroke logs in combination with eye-tracking. The problems of writing tasks complexity and writing performance are discussed in such papers.

Cluster 4 (**greenish yellow**, 21 items) is associated with the study of such aspects of writing as pause behavior, revision behavior in people with reading and writing difficulties (including those suffering from aphasia).

Cluster 5 (**purple**, 16 items) is related to the study of cognitive load and cognitive process in L2 writing (including writing from multiple sources and workplace writing) via analysis of revisions and pauses analysis using the Inputlog software. The problems of error correction are also discussed in such papers.

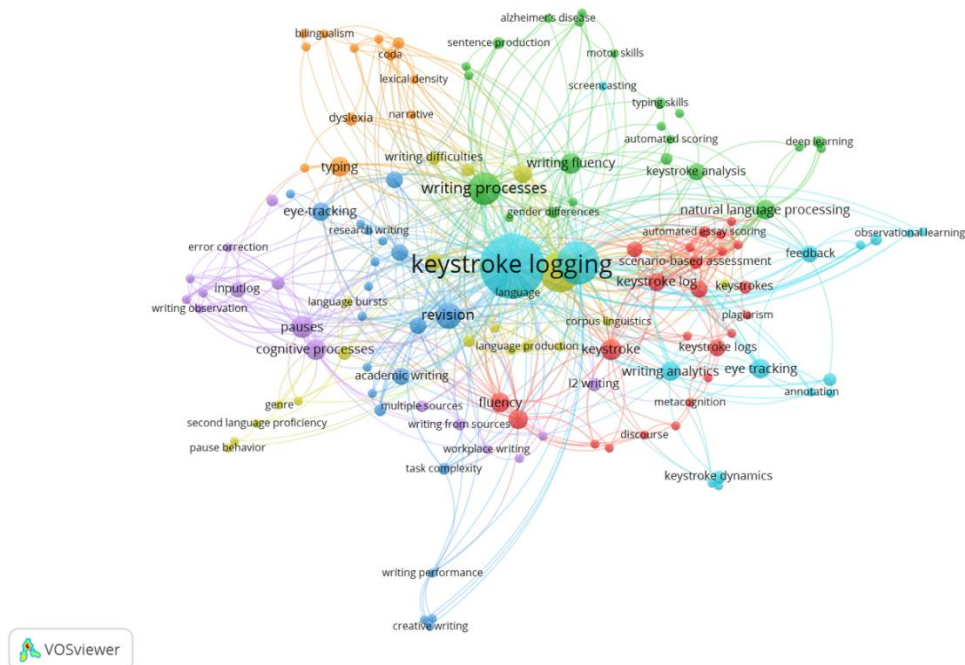
Cluster 6 (**light blue**, 16 items) is related to the problem of the analysis of the writing process using a set of methods (keystroke logging, screencasting, eye tracking) with different purposes including tracking the states of writers (boredom, engagement) with possible writing

intervention depending on the detected state. The problems of writing analytics are discussed.

Cluster 7 (**orange**, 12 items) discusses the problems of the characteristics of written product via an analysis of typing of people with dyslexia, deaf and hard of hearing (DHH), children of deaf adults (CODA).

Figure 15. Clusters of authors' keywords

Рисунок 15. Сеть совместной встречаемости ключевых слов автора

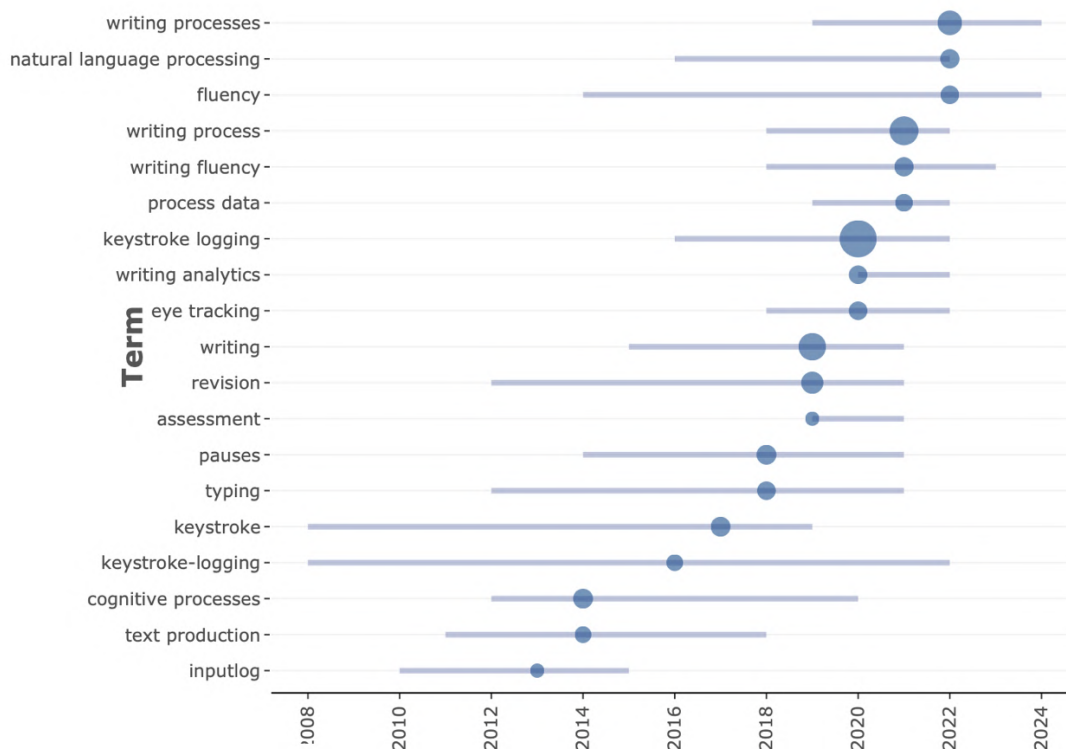


However, this picture is static. The bibliometric tools used in this study allowed us to reveal the research trends using a different set of methods. The basic method is the analysis of dynamics of keywords occurrences over time (Figure 16). For example, we can see that natural language

processing methods have been actively used over the last ten years, with a peak in the last year. Writing analytics and writing assessment are the recent topical trends. Inputlog was often mentioned in the papers published in 2010-2015, with a peak in 2013.

Figure 16. Dynamics of keywords over the years

Рисунок 16. Динамика частотности ключевых слов по годам



But it is more interesting to perform a co-word analysis and draw clusters of keywords in the form of a thematic map (Figure 17). This type of visualization implemented in bibliometrix is a very intuitive plot: themes are classified according to the quadrant in which they are placed: (1) upper-right quadrant – motor-themes; (2) lower-right quadrant – basic themes; (3) lower-left quadrant – emerging or disappearing themes; (4) upper-left quadrant – very specialized/niche themes (Büyükkıdık, 2022; Derviş, 2019; Rashid, 2023).

The basic themes of our collection are the general problems of the study of writing processes based on keystroke data, analysis of revisions in writing, analysis of writing fluency. Among the basic themes there is a combination of keystroke logging and eye-tracking to study cognitive effort of the writers, the problems of studying writing in a second language including Chinese, problems of plagiarism detection via keystroke analysis. The motor themes are the study of pauses and cognitive processes between them using

Inputlog, the analysis of metacognition in second language writing, problems of linking the textual quality of writing as a product with the characteristics of writing as a process, the use of NLP and data science methods for writing analysis.

The niche themes are the development of the methods for code completion using keystroke data, construction of text history, analysis of pause and revision behavior with connection of second language proficiency, writing intervention as part of writing pedagogy, analysis of keystroke on smartphones, the use of deep learning methods for natural language generation.

The declining themes are the study of workplace writing from multiple sources. The emerging themes are the study of writing tasks complexity based on keystroke data, the use of machine learning for revision analysis, application of mixture models for pause analysis and assessment of typing skills.

The research trends could also be revealed via overlay visualization offered by VOSviewer (Figure 18).

Figure 17. Thematic map of the collection
Рисунок 17. Тематическая карта коллекции

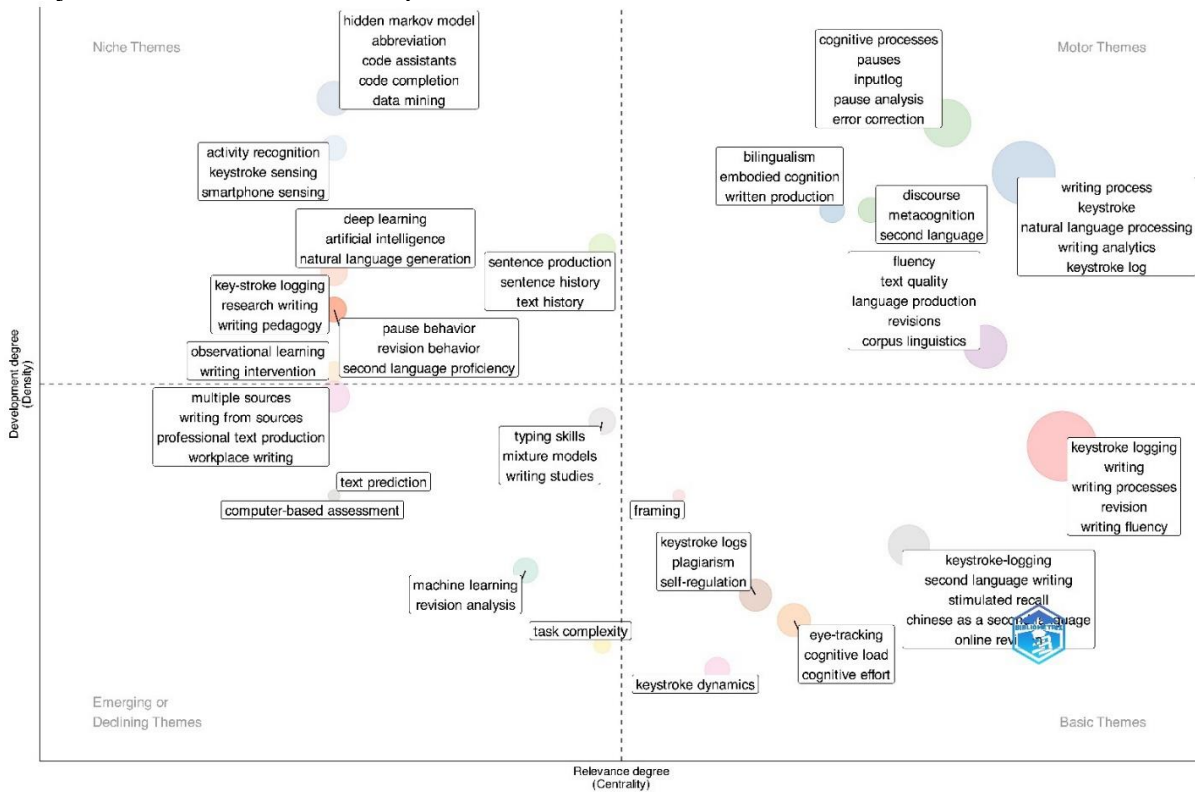
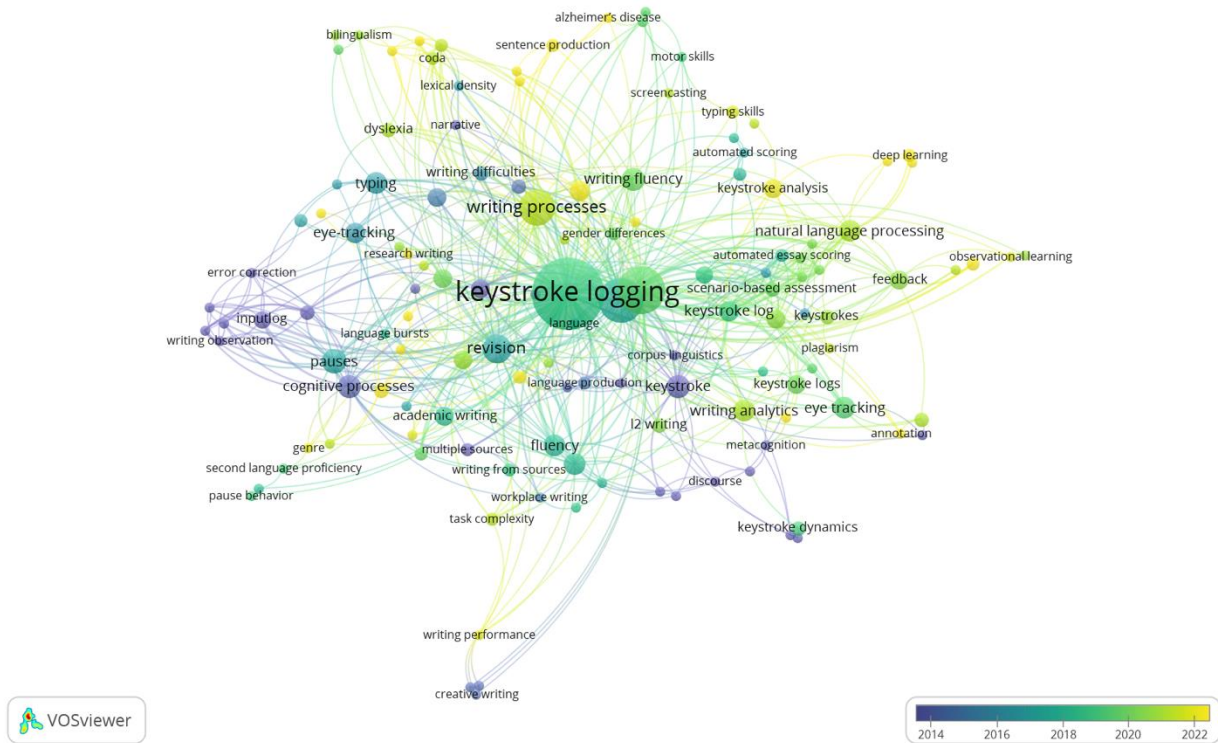


Figure 18. Overlay visualization of domain based on an authors' keywords co-occurrences
Рисунок 18. Оверлейная визуализация совместной встречаемости ключевых слов



In the overlay visualization, the color of a node indicates a certain property of the node. In Figure 6, the overlay color of each circle corresponds to the average publication year of all the papers that include the corresponding term. In this map, the terms with cold colours (e.g. blue) represent research activities with an older average publication year, and the terms with hot colours (e.g. red) show terms with a more recent average publication year. Overlay visualization of domain was constructed on an authors' keywords co-occurrences (association strength was used for normalization which is a default method, weights = occurrences, min strength=1).

The general conclusions about the topic evolution offered by VOSviewer are in line with those revealed by the thematic map. Analysis of cognitive processes during writing, the study of metacognition, error correction, use of Inputlog to study writing were at the centre of research attention at the first stage of analysis, while in the recent years the problems of task complexity

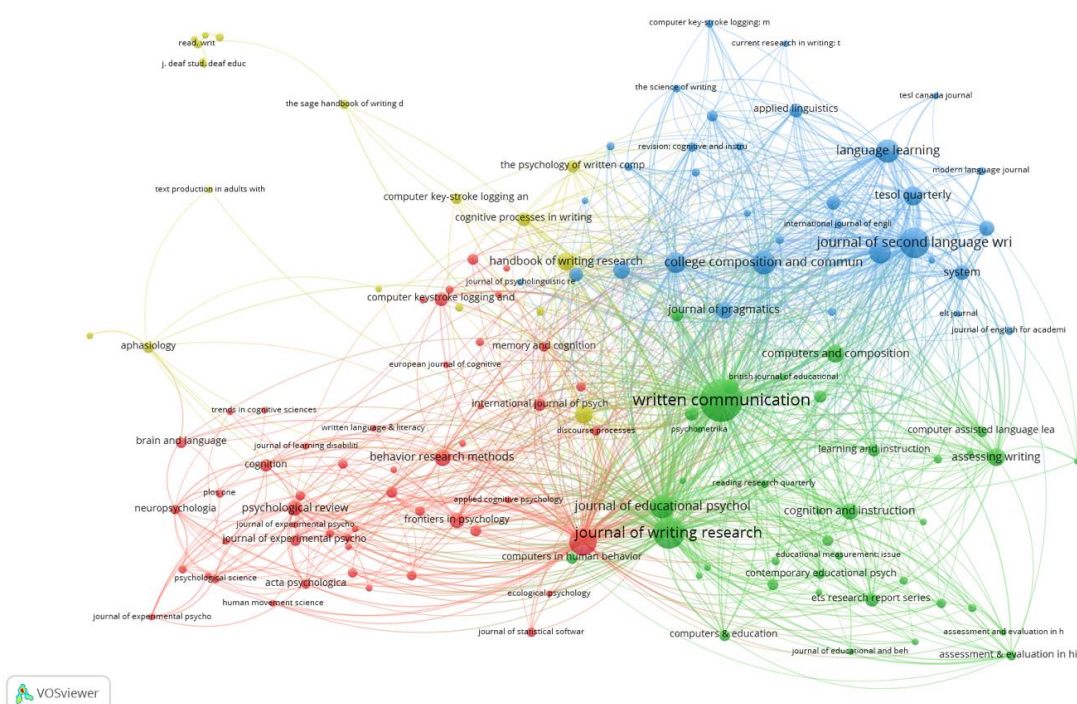
assessment, writing analytics, use of deep learning in writing research, analysis of typing skills, observational learning have become hot topics.

3.2.2. Co-citation analysis

Co-citation analysis uncovers the foundations of knowledge relied on by scientific research in the field, wherein clusters are formed based on the cited documents (references) that frequently appear together.

The clusters of the journals with strong co-citation relations are shown in Fig. 19. Three large clusters and a small one are revealed. The largest cluster – **green** – consists of the journals related to the general problems of writing research and writing assessment. The **red** cluster is made up of psychology and cognitive science journals, the **blue** cluster is made up of language learning journals, and the small **yellow** cluster is made up of journals related to writing by people with disabilities. This cluster solution shows different directions of writing research using keystroke data and methodology.

Figure 19. Co-citation network (cited sources)
Рисунок 19. Сеть совместных цитирований (источники)



3.2.3. Bibliographic coupling

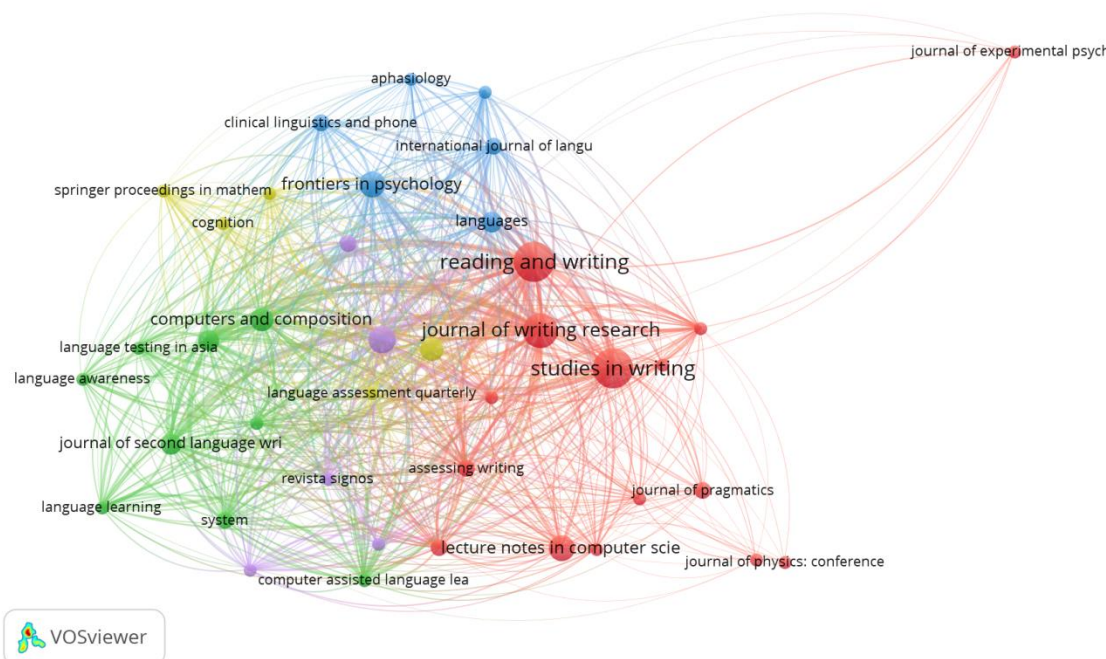
We also consider bibliographic coupling, which is the opposite of co-citation, as there is an overlap in the reference lists of publications. In bibliographic coupling visualization, clusters are formed based on the referencing similarity of citing documents (articles). The larger the number of references journals have in common is, the stronger the bibliographic coupling relation between them is (Figure 20). Figure 20 shows a very different clustering compared to the co-citation network (Figure 19): one of the two large clusters (**red**) consists of the top-3 most cited and relevant journals of our collection devoted to the study of writing, as well as

Journal of Pragmatics, – on one side, and proceedings from computer science subject area – *LNCS* and *Journal of Physics: Conference series*, which points to the existence of a multidisciplinary direction of studies in the analyzed domain combining methods from data science and writing research.

Another large cluster – the **green** one – contains journals devoted to different aspects of language learning, especially computer assisted language learning. The **blue** cluster is made up of the journals which consider psychological and physiological aspects of language production.

Figure 20. Bibliographic coupling of the journals

Рисунок 20. Сеть совместной встречаемости ссылок в пристатейных списках



Conclusions

This study utilizing a combination of the bibliometrix R-package and VOSviewer fills the research gap by providing a holistic view of the research area related to the study of writing using keystroke data which can help scholars understand its current conceptual structure and development trends.

The analysis highlights the feasibility of using R and the existing libraries for bibliometrics.

As our review shows, there has been an upward trend in the number of papers using keystroke loggers to study the writing process over the past five years. However, only few countries (the USA, Belgium, the Netherlands, Sweden) have the most published and cited papers, with Belgium

being the leader in terms of the cited papers and the most influential authors. The level of country collaboration is quite low. Thus, only few languages are involved in this important field of study, which obviously hinders further development and refinement of the models of writing. More international collaborations are needed, and more research groups all over the world should be involved in this field.

We believe that the knowledge mapping we present in this paper using a set of approaches will attract researchers who are new to this field to find directions for future research.

Nonetheless, this review acknowledges certain limitations. While this study is an attempt to fill a gap in bibliometric studies of writing research based on keystroke data, it considers only one citation database, which may not capture all the studies devoted to the use of keystroke methodology to understand the writing process. In the future, it would be beneficial to add data from some other databases (i.e., PubMed, Dimensions). Only English-language sources were analysed. Furthermore, the content analysis in this study does not take into account the use of synonyms. With future revisions, we want to address the aforementioned restrictions.

While a bibliometric review allows to analyze hundreds of papers and obtain an objective holistic view of the research field, it is unable to consider specific problems of the field. Therefore, our future plans include performing systematic scoping reviews (an excellent example of systematic scoping reviews is presented in (Tikhonova et al., 2023) to address some specific questions such as the linguistic nature of writing bursts, i.e., minimal units of text production, and ways of their extraction based on raw keystroke log data.

References

Alfalahi, H., Khandoker, A. H., Chowdhury, N., Iakovakis, D., Dias, S. B., Chaudhuri, K. R. and Hadjileontiadis, L. J. (2022). Diagnostic accuracy of keystroke dynamics as digital biomarkers for fine motor decline in neuropsychiatric disorders: a systematic

review and meta-analysis. *Sci Rep* 12, 7690. <https://doi.org/10.1038/s41598-022-11865-7>

Aria, M. and Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics* 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>

Bixler, R. and D’Mello, S. (2013). Detecting boredom and engagement during writing with keystroke analysis, task appraisals, and stable traits. In Proceedings of the 2013 International conference on Intelligent user interfaces (IUI ‘13). Association for Computing Machinery, New York, NY, USA, 225–234. <https://doi.org/10.1145/2449396.2449426>

Büyükkıdık, S. (2022). A bibliometric analysis: A tutorial for the bibliometrix package in R using IRT literature. *Journal of Measurement and Evaluation in Education and Psychology* 13(3), 164–193. <https://doi.org/10.21031/epod.1069307>

Cui, Y., Mou, J. and Liu, Y. (2018). Knowledge mapping of social commerce research: A visual analysis using CiteSpace. *Electronic Commerce Research*, 18(4), 837–868.

Derviş, H. (2019). Bibliometric analysis using bibliometrix an R package. *Journal of Scientometric Research* 8(3), 156–160. <https://doi.org/10.5530/jscires.8.3.32>

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. and Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research* 133, 285–296. <https://doi.org/10.1016/j.jbusres.2021.04.070>

Farrukh, M., Raza, A., Mansoor, A., Khan, M. S. and Lee, J. W. C. (2023). Trends and patterns in pro-environmental behaviour research: a bibliometric review and research agenda. *Benchmarking: an International Journal* 30(3), 681–696. <https://doi.org/10.1108/BIJ-10-2020-0521>

Lee, S. A. and Park, J. H. (2024). Do Individuals with Mild Cognitive Impairment and Healthy Aging People Have Different Keystroke Dynamics? A Systematic Review. *International Journal of Gerontology* 18(2), 64–69. [https://doi.org/10.6890/IJGE.202404_18\(2\).0001](https://doi.org/10.6890/IJGE.202404_18(2).0001)

Leijten, M. and Van Waes, L. (2013). Keystroke logging in writing research: Using inputlog to analyze and visualize writing processes. *Written Communication* 30(3), 358–392. <https://doi.org/10.1177/0741088313491692>

- Leijten, M., Van Waes, L., Schriver, K. and Hayes, J. R. (2013). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research* 5(3), 285–337. <https://doi.org/10.17239/jowr-2014.05.03.3>
- Logan, G. D. and Crump, M. J. (2010). Cognitive illusions of authorship reveal hierarchical error detection in skilled typists. *Science* 330(6004), 683–686. <https://doi.org/10.1126/science.1190483>
- Maalej, A. and Kallel, I. (2020). Does Keystroke Dynamics tell us about Emotions? A Systematic Literature Review and Dataset Construction. In 2020 16th International Conference on Intelligent Environments (IE), Madrid, Spain, 60–67. <https://doi.org/10.1109/IE49459.2020.9155004>
- Nguyen, T. M., Leow, A. D. and Ajilore, O. (2023). A Review on Smartphone Keystroke Dynamics as a Digital Biomarker for Understanding Neurocognitive Functioning. *Brain Sci.* 13, 959. <https://doi.org/10.3390/brainsci13060959>
- Omotehinwa, T. O. (2022). Examining the developments in scheduling algorithms research: a bibliometric approach. *Heliyon* 8(5), e09510.
- Passas, I. (2024). Bibliometric Analysis: The Main Steps. *Encyclopedia* 4, 1014–1025. <https://doi.org/10.3390/encyclopedia4020065>
- Quraishi, S. J. and Bedi, S. S. (2018). Keystroke Dynamics Biometrics, A tool for User Authentication – Review. In 2018 International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 248–254. <https://doi.org/10.1109/SYSMART.2018.8746932>
- Rashid, M. F. A. (2023). How to Conduct a Bibliometric Analysis using R Packages: A Comprehensive Guidelines. *Journal of Tourism, Hospitality & Culinary Arts* 15(1), 24–39.
- Raul, N., Shankarmani, R. and Joshi, P. (2020). A Comprehensive Review of Keystroke Dynamics-Based Authentication Mechanism. In: Khanna, A., Gupta, D., Bhattacharyya, S., Snasel, V., Platos, J., Hassanien, A. (eds). International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing, vol. 1059. Springer, Singapore. https://doi.org/10.1007/978-981-15-0324-5_13
- Rejeb, A., Rejeb, K., Appolloni, A., Kayikci, Y. and Iranmanesh, M. (2023). The landscape of public procurement research: a bibliometric analysis and topic modelling based on Scopus. *Journal of Public Procurement* 23(2), 145–178. <https://doi.org/10.1108/JOPP-06-2022-0031>
- Saini, G. K., Lievens, F. and Srivastava, M. (2022). Employer and internal branding research: a bibliometric analysis of 25 years. *Journal of Product & Brand Management* 31(8), 1196–1221. <https://doi.org/10.1108/JPBM-06-2021-3526>
- Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *J Am Soc Inf Sci* 24, 265–269. <https://doi.org/10.1002/asi.4630240406>
- Stevenson, M., Schoonen, R. and Gloppe, K. de. (2006). Revising in two languages: A multi-dimensional comparison of online writing revisions in L1 and FL. *Journal of Second Language Writing* 15(3), 201–233. <https://doi.org/10.1016/j.jslw.2006.06.002>
- Tikhonova E., Kosycheva M., Kasatkin P. (2023). Exploring Academic Culture: Unpacking its Definition and Structure (a Systematic Scoping Review). *Journal of Language and Education* 9(36), 151–168. <https://doi.org/10.17323/jle.2023.18491>
- Van Eck, N. J. and Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau and D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 285–320). Springer. http://dx.doi.org/10.1007/978-3-319-10377-8_13
- Wang, S. and Gu, Z. (2022). Mapping the Field of Value Chain: A Bibliometric and Visualization Analysis. *Sustainability* 14, 7063. <https://doi.org/10.3390/su14127063>
- Wengelin, Å. (2006). Examining Pauses in Writing: Theory, Methods and Empirical Data. In *Computer Keystroke Logging and Writing: Methods and Applications*. Leiden, The Netherlands: Brill. https://doi.org/10.1163/9780080460932_008
- Wengelin, Å. and Johansson, V. (2023). Investigating Writing Processes with Keystroke Logging. In: Kruse, O. et al. *Digital Writing Technologies in Higher Education*. Springer, Cham. https://doi.org/10.1007/978-3-031-36033-6_25
- Yang, L. and Qin, S.-F. (2021). A Review of Emotion Recognition Methods From Keystroke, Mouse, and Touchscreen Dynamics. *IEEE Access* 9, 162197–162213. <https://doi.org/10.1109/ACCESS.2021.3132233>

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.

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

Дехнич Ольга Витальевна, кандидат филологических наук, доцент, заместитель директора Института межкультурной

коммуникации и международных отношений по научной деятельности, доцент кафедры английской филологии и межкультурной коммуникации, Белгородский государственный национальный исследовательский университет, Белгород, Россия.

Tatiana A. Litvinova, Doctor of Philological Sciences (D.Sc. in Philology), Head of the Textual Psycholinguistic Modelling Research Laboratory, Professor of the Department of Russian Language, Modern Russian and Foreign Literature, Voronezh State Pedagogical University, Voronezh, Russia.

Литвинова Татьяна Александровна, доктор филологических наук, заведующий научно-исследовательской лабораторией текстового психолингвистического моделирования, профессор кафедры русского языка, современной русской и зарубежной литературы, Воронежский государственный педагогический университет, Воронеж, Россия.

UDC 81`33

DOI: 10.18413/2313-8912-2024-10-3-0-6

Abdulaziz B Sanosi¹ 
Abuelgasim Sabah
Elsaid Mohammed² 

ASAWEC: towards a corpus of Arab scholars' academic written English

Prince Sattam bin Abdulaziz University,
5229, Hawtat Bani Tamim, 16628, Saudi Arabia
E-mail: a.assanosi@psau.edu.sa
ORCID:0000-0003-3447-2818

Prince Sattam bin Abdulaziz University,
7524 Hawtat Bani Tamim, 16628, Saudi Arabia
E-mail: a.ibrahim@psau.edu.sa
ORCID: 0000-0001-9791-0905

Received 07 October 2023; accepted 15 September 2024; published 30 September 2024

Abstract: Linguistic corpora have been used in a wide range in recent years. Different types of linguistics analyses in both spoken and written discourses are being conducted using the corpus linguistics approach. Among these, academic writing has received considerable attention. Corpus linguistics has provided insights into the academic writing of both native and non-native English language learners and writers in general. Nevertheless, relatively few studies have investigated this topic in the Arab EFL setting. Consequently, there is a relative paucity in corpora of Academic written English by Arab speakers. To address this gap, we compiled the Arab Scholars' Academic Written English Corpus (ASAWEC) which is a specialized corpus of Arab scholars' academic written English. We collected the corpus texts according to specific criteria, and then we normalized and cleaned the data. The texts were then tokenized and tagged and the corpus underwent initial tests which yields insightful findings on Arab scholars' academic written English such as the low lexical diversity and the utilization of various discourse techniques. The present paper introduces the corpus, provides details on its compilation, presents initial results and statistics, and discusses potential limitations and future perspectives for updating the corpus. It is envisaged that this project will encourage the use of the ASAWEC and help in launching similar initiatives to advance research in Arab corpus linguistics.

Keywords: Specialized corpus; Corpus compiling; Academic writing; Corpus linguistics; L2 writing

How to cite: Sanosi, A. B, Mohammed, A. S. E. (2024). ASAWEC: towards a corpus of Arab scholars' academic written English, *Research Result. Theoretical and Applied Linguistics*, 10 (3), 116-134. DOI: 10.18413/2313-8912-2024-10-3-0-6

Introduction

The shift from formal linguistics to functional linguistics paved the way for new methods of analyzing language, the most prominent of which is corpus linguistics.

Functional linguistics emphasizes the study of language as a functional system through empirical evidence that relies on real contexts. Accordingly, the need for actual language data among linguists was a significant driving

force in the development of modern corpora by the 1950s (Blecha, 2012). Corpus linguistics helps in meeting this need as it is concerned with “the study of language based on examples of “real life” language use” (McEnery and Wilson 2001: 1) using large, structured corpora that reflect actual language use and show patterns of language employed by different users.

Currently, linguists from various schools rely on linguistic corpora not only to gather authentic language examples but also to test their linguistic theories against quantitative data derived from actual language usage (Kubler and Zinsmeister, 2015). The huge advances in computers and related technologies make this possible and enhance the practice. Corpora have now become of many types and large sizes amounting to tens of millions and more. Nevertheless, the reality of corpus linguistic analysis in the Arab world is still humble as few corpora of different types are available in the Arab EFL setting (Almohizea, 2017; Alotaibi, 2017). The researchers think this is unfortunate, especially after the massive development in corpus software and Natural Language Processing (NLP) solutions that made linguistic data collection and analysis a relatively convenient process (Dunn, 2022).

Thus, the problem of this research can be formulated as corpus linguistic researchers in the Arab EFL lack access to robust and well-established corpora that can be relied upon for linguistic analysis. This problem may dissuade scholars from pursuing corpus-based studies because their only alternative would be to create their own DIY corpora, which is not a feasible option for many researchers as it is a time and effort-consuming task. Even for those who opt to develop their corpora, this approach may lead to unsatisfactory results, as compiling a corpus involves numerous intricate steps to ensure its size appropriateness, representativeness, balance, and inclusivity. Moreover, the issue is intensified by the fact that corpus linguistics requires expertise in other disciplines, particularly in IT and related

fields. Even though many corpus linguists and applied linguistics researchers excel in their respective areas, a deficiency in computer science and programming expertise may hinder their ability to develop well-prepared corpora. This is because new corpora heavily depend on modern technology, which is constantly evolving and being updated. Adapting to these changes poses an additional challenge for linguistic researchers.

Furthermore, to cope with the trending approaches of linguistic analysis and corpus-based discourse studies that guide most of today’s applied linguistic research, further interest is needed in the field of corpus linguistics in the Arab EFL setting. This interest entails compiling different types of corpora as “[the] number of readily available specialized corpora is still much smaller than [the] number of general corpora” (Blecha, 2012: 5), and even the existing ones cannot meet all the scholars’ research needs. Thus, the primary aim of this paper is to present the specialized Arab Scholar Academic Written English Corpus (ASAWEC) to the public and describe the processes and procedures taken to build it. The project is anticipated to lay the foundation for a large, specialized corpus of academic writing by Arab scholars and inspire similar initiatives in the field. Additionally, the main output of the project, ASAWEC, will be accessible to the public to promote research in the area and allow for the corpus to be revised and enhanced.

Literature Review

The word *corpus* is a Latin word that means *body*. However, in modern linguistics, a corpus is not a mere body of texts, it rather involves four principles which are: sampling and representativeness, being of a finite size, machine-readable, and having a standard reference (McEnery and Wilson, 2001). These features enable linguistic corpus to provide an evidence-based source of linguistic analysis (Meyer, 2023), and hence make corpus linguistic “scientific method of language analysis [that] requires the analyst to provide empirical evidence in the form of data drawn from language corpora in support of any

statement made about language” (Brezina, , 2018: 2). The involvement of the scientific method and specific properties of linguistic corpus entails a new contemporary definition that accounts for all these variables and keeps it apart from the traditional views towards corpora and corpus linguistic analysis which entail many descriptions. This definition states that “Corpus linguistics is the investigation of linguistic research questions that have been framed in terms of the conditional distribution of linguistic phenomena in a linguistic corpus.” (Stefanowitsch, 2020: 56).

As corpus linguistics has expanded in recent decades to incorporate almost all types of linguistic analysis, various kinds of corpora have been classified and standardized by researchers in terms of their register or purpose.

Types of Corpora

A clear-cut classification of corpora depends on their register i.e., spoken or written. Most available corpora are, due to practical reasons, written. Sinclair (1991) stated that compiling a spoken corpus is not as a straightforward task as a written one since it involves many challenges such as transcription and representativeness. He argued that collecting actual spoken language is impossible, whereas collecting quasi-speech material, such as film transcripts, is considered artificial and does not represent actual language use. Although there is a significant advance in transcription, speech recognition, and speech-to-text applications, still spoken corpora are rare compared to written (Lemmenmeier-Batinić, 2023).

According to Myer’s (2023) classification, there are several types of commonly available English corpora. Multipurpose corpora are designed to represent a particular register of speech or writing and can be used for various analyses. Learner corpora contain written and/or spoken language produced by learners of the language, also known as “interlanguage” (Gilquin, 2020). Teachers, researchers, and material designers use learner corpora to

inform their teaching practice, gather data on students’ language patterns, and enhance their syllabi. The main focus of corpus-based instruction is writing (Liu, 2022), and teachers use learner corpora as a teaching strategy to help students analyze concordance lines and find patterns of language use. Alternatively, they may use corpus-based teaching materials to improve students’ academic writing skills.

Historical corpora are collections of texts from a specific time that can be compared to other collections to study linguistic evolution over time. Parallel corpora contain texts in English and their translations into another language. They are used to enhance translation skills and study variations across languages.

Another widely recognized type of corpora is a specialized corpus (Baker, et al., 2006; Blecha, 2012; Stefanowitsch, 2020) which is more relevant to the current study.

Specialized Corpora

In its essence, a specialized corpus is “a corpus which has been designed for a particular research project, for example, lexicography for dictionary compilation, or to study particular specialist genres of language” (Baker, et al., 2006: 147). Although this type does not contribute to the description of language as a whole since they do not represent the usual natural language use, it is nevertheless effective in describing features of the language that distinguish specific genres or linguistic communities (Blecha, 2012). These features place restrictions on the texts to be included within them and hence they are normally smaller than general or reference corpora (Baker, 2010).

Specialized corpora aim to identify domain-specific vocabulary, as researchers and educators have observed that words vary across different fields. By creating wordlists from various specialized corpora, it is possible to narrow down the meanings of words associated with each context. In this regard, several researchers compiled corpora addressing specific fields and the language used in them. Toriida (2016) described the

steps she followed in compiling a specialized corpus for the EAP nursing program. She found that the corpus was effective in generating an annotated wordlist for academic English in the field of nursing.

Specialized corpora can also have practical purposes, as demonstrated by Allan et al. (2023), who reported the construction of a specialized corpus of written English from Swedish year 9 National Tests. Although the corpus appears to be a learner corpus, its focus on test language and inclusion of teacher feedback suggests that it can also be used as a tool to guide teachers in assessing tests which can be another motive for compiling specialized corpora.

Specialized corpora and academic writing

The use of corpus linguistics in the investigation of academic writing is a common practice. This is mainly because this type of writing necessitates strict criteria to differentiate it from other forms of writing. Specifically, it should demonstrate proficiency in rhetorical conventions, linguistic characteristics, vocabulary, and syntax (Utkina, 2021). Since such aspects are unusual in common everyday language, they represent an ideal source for specialized corpora. Several studies explored the use of specialized corpora in academic writing. For example, Guerra and Smirnova (2023) compiled a 1.5-million-word corpus to analyze the linguistic complexity in professional academic writing. They compared the research articles published in peer-reviewed journals in hard and soft disciplines. They found the specialized corpus was informative and presented empirical evidence on the linguistic variation.

Other specialized corpora merged the two aims of analyzing academic writing and generating wordlists. In this regard, Jamalzadeh and Tabrizi (2020) created a specialized corpus of 3.7 million words with the goal of developing a Tourism Academic Word List (TAWL) that includes the most used academic vocabulary in various sub-disciplines of tourism. They achieved this by

analyzing a written corpus of academic research articles in the field of tourism.

Considering research articles written by non-native speakers of English, the issue seems more interesting as it is anticipated that specific features of such forms of writing may be showcased. Research has not neglected this issue, though very rare studies are available in this regard. For example, Fuentes (2009) constructed a specialized corpus to examine the academic writing of Spanish authors in the field of computer science. The study involved comparing the findings with a selection from the British National Corpus (BNC). The results indicated that there were no significant differences in academic vocabulary or style between native and non-native writers. However, it is important to note that the small size of the corpus (only 25,000 words) may raise concerns about the validity of these claims. Nevertheless, the study employed a rigorous data collection method by gathering research articles in their final version, prior to peer review. This approach enhances the authenticity of the data as peer review and proofreading can significantly impact the work of scholars.

In the Arab context, most studies that investigated Arab academic writing utilized learner corpora. There are few studies that investigated the use of Arab scholars of different linguistic features in their academic writing by compiling specialized research-specific corpora. Among these studies are Akeel (2020) who investigated the use of modal verbs by Saudi postgraduate writers of English. and Sanosi (2022) who investigated the use of lexical bundles by both Arab learners and British Scholars. Both authors compiled their own corpora for the research purposes and compared the results against the British Academic Written English (BAWE) corpus. Akeel (2020) found that Arab postgraduate writers of academic English underused modal verbs in their academic writing, while Sanosi (2022) found no effect of further study or professional practice on the development of lexical bundle use by Arab Scholars.

It can be concluded that specialized corpora can have a notable impact on improving academic writing for both non-native English students and scholars. As a result, this study draws inspiration from the insights discussed above regarding academic writing and specialized corpora. Thus, the rationale for the current study was formulated.

Study rationale

The initial motivation for compiling a specialized corpus of academic writing by Arab scholars was a research project on metadiscourse resources used by Arabic academic writers (in press). During the preparation of this research, it was discovered that data pertaining to Arab scholars' English writing was scarce despite their vast publication in recent years. This drove the researchers to compile a corpus for research purposes and make it available to their surrounding research community, as well as publicize it online.

Then, the rationale for compiling ASAWEC is threefold. First, the corpus is intended to be a valuable resource for researchers who aim to explore linguistic and discourse aspects of academic writing by Arab scholars. For example, analysis of the corpus may reveal common areas of difficulty for these writers, such as difficulties with certain grammatical structures or academic vocabulary. Additionally, the corpus can shed light on the linguistic resources that are recurrently used by Arab scholars writing in English, such as certain rhetorical strategies or discourse markers. Finally, analysis of the corpus can reveal the linguistic devices that Arab scholars use to mark stance and construct metadiscourse, which can provide insights into their rhetorical and communicative goals. This will enrich the field and answer research questions on the

effects of cultural background on writing practice and discourse structure.

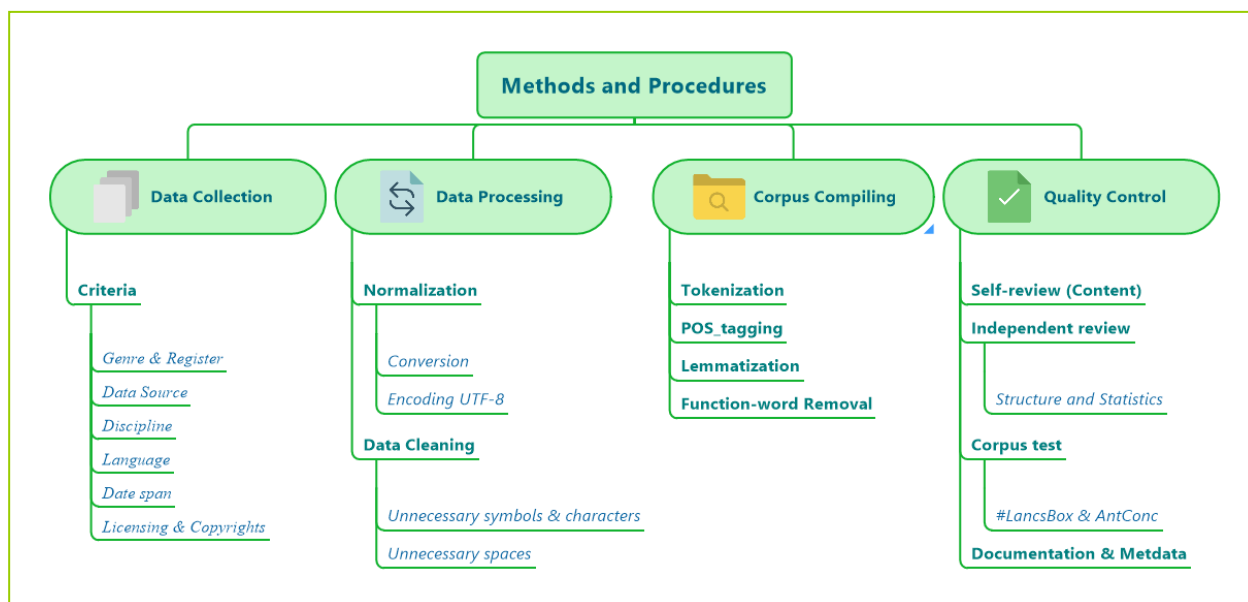
Second, the corpus can inform the development of training and teaching materials designed to enhance academic writing at college level and professional development intended for scholars. By identifying areas of difficulty for academic writers, curriculum designers and reviewers can update curricula and courses for both graduate and postgraduate students which can result in improved outcomes for scholars and students.

Finally, the corpus can contribute to promoting inclusivity in academic discourse by introducing a sample of the academic discourse of a group rarely represented in corpus studies. This aim will promote equitability in the academic community and promote diverse scholarly work.

Methodology

The researchers utilized the corpus compiling methods and techniques recommended by established scholars and pioneers in corpus linguistics, such as Sinclair (1991), Atkins et al., (1992), McEnery and Wilson (2001), Baker et al. (2006), and Baker (2010). Studies by contemporary scholars in this field were also reviewed, e.g. Toriida (2016), Brezina (2018), Stefanowitsch (2020), Meyer (2023) and others. We also referenced resources on Python and NLTK, such as Bird et al. (2009) and Dunn (2022). Additionally, the researchers sought continuous input from colleagues in the IT and statistics fields to ensure the utmost accuracy and rigor in constructing ASAWEC, which involved utilizing various software and applications. Ultimately, the corpus was compiled following the procedures outlined in Figure 1.

Figure 1. Procedures of ASAWEC compiling



Thus, the methods for compiling the corpus and subsequent design of this research article are summarised in the steps outlined in Figure 1 and detailed below.

Sampling Frame and Data Collection:

The main step in the method is to establish a sampling criterion to guide the selection of the texts to be included in the corpus. Sampling of a corpus is one of the issues that face corpus compilers because regular statistical measures for defining sampling and population are hardly applicable to corpus linguistics mainly because it is difficult to define the volume of the whole language production that the sample is to be taken from (Brezina, 2018). However, in specialized corpora, the problem is less thorny, as it is noted that “the more highly specialized the language to be sampled in the corpus, the fewer will be the problems in defining the texts to be sampled” (Atkins et al., 1992: 7). To limit the language production that the corpus is to be taken from as a representative sample, researchers suggested using a sampling frame prior to the selection process. The sampling frame involves establishing a set of criteria to control the texts to be selected for building the corpus (Meyer, 2023). This practice is especially

common in compiling specialized corpora since they need parameters to set limit to the text they would like to include in their corpora (Hunston, 2002). Accordingly, several criteria were adopted by the researchers to establish the sampling frame and control the data collection for the present project. These criteria are discussed below.

Genre and register

The aim of this project is to provide valid and accurate data to analyze the patterns and trends of the English writing styles used by scholars and academics whose L1 is Arabic. The motivating research topic was to analyze metadiscourse resources used by Arab scholars. However, it was envisaged that the corpus designed for the study could be a valuable source for future research in the field rather than a byproduct of one research. Further, the corpus can be a nucleus for a larger corpus after continuous updating and wider participation from potential partners. Accordingly, this aim informed the genre and register of the corpus and hence set the primary criterion for the sampling frame.

It was envisaged that the most practical and common representative of academic style is the scientific article genre, mostly in IMRaD style, that is published in online

journals. Scholars who publish in such journals normally follow strict writing conventions and adhere to the formal use of vocabulary and grammar. These register-specific conventions are envisaged to stimulate several research questions such as investigating scholars' use of vocabulary, syntactic structures, discourse markers, collocations, lexical bundles, and stance markers to name a few. Exploring this will be more feasible by following a corpus-assisted discourse analysis approach and this is the objective that the current project aims to help in achieving. Plans to update the corpus include the addition of other genres such as conference papers, presentations, seminars and research notes. While these genres do not fully represent academic writing, it may not be suitable to include other genres such as essays, theses, and dissertations at this point, as they may reflect the writing of students rather than scholars.

Data source

The major source of the corpus is the scientific journals that were actively published in the region for the specified period. We targeted journals that are published by Arab universities and academic entities as they are more likely to incorporate research articles written by Arab scholars. Through the building and cleansing processes, we adjusted the number of selected articles to reach our initial target of one million words. The decision to initially select only Arab journals was based on practical considerations as it was more convenient to determine the identity of the writers. However, future updates will involve adding articles written by Arab scholars in international journals. This task will require further work, as it involves verifying the identity of the scholars, either through direct contact or their respective institutions. This process will be more feasible once institutional participation in the development of this corpus is achieved. This is one of the future plans of the corpus compilers.

Discipline

As a starting point we selected disciplines that related to the English

language and literature. In this regard, we searched for journals whose scope is one of applied linguistics, literature, and translation. We believe that these will form a suitable foundation for the corpus as scholars in this field are more likely to be aware of academic writing conventions and their articles would serve our initial purpose of forming the corpus. However, the proposed updates, as outlined below, involve incorporating articles from the fields of science, social sciences, applied sciences, and arts. In this stage, we build a pool containing 19 journals.

Language and content

As explained before, the corpus compiling aimed to provide data for corpus linguists to analyse academic English written by Arab scholars. Therefore, we searched for publication containers that address linguistic issues in the Arab World in general. Consequently, most of these articles contained an amount of Arabic text that must be removed to maintain the corpus accuracy and representativeness. In some cases, the Arabic content is small and can be managed by the cleansing procedures e.g., Arabic abstracts, Arabic running heads, and Arabic headers and footers. However, other articles contain a huge amount of Arabic text that cannot be deleted without affecting the article's structure and meaning. This feature was found in two journals that focus on Quranic texts and translation. Both journals were excluded to maintain consistency and accuracy reducing the pool to 17 journals.

Date span

To provide more updated data we specified the period of publication to be between 2019 – 2023. We settled for this date span after continuous adjustments stemming from observing the publication of the journal selected in the first stage and the application of the other criteria. The time span for the articles is considered for the whole corpus since there are some journals which started publication in 2021 while there are others which discontinued publication before 2023. The implementation of this criterion reduced the number of journals in the selection to 11 journals.

Licensing and copyrights

Recently, many journals adopted an open access policy and used Creative Common (CC) licenses that facilitate the use of their content for compiling corpora. Some journals make their content available under a CC BY license which allows others to share, copy, redistribute, remix, transform, and build upon the original work for any purpose, even commercially, if they give appropriate credit to the original creator(s) and indicate if changes were made. Others adopt CC BY-NC, or CC BY-SA licenses, which enable the reader to browse, download, and reuse their contents as the purpose is non-commercial and the new work generated from them is shared under the same licenses. Yet, there are less permissive licenses such as CC BY-ND, and CC BY-NC-ND which are unsuitable for corpus compiling as they prohibit modification of any means to the original content unless after acquiring permission

from the publishers.

Most journals in our pool are published under CC BY and CC BY-NC so we selected them for the study. For other restricted journals we contacted five publishers to request full access to their archives and we explained the necessary modifications that should be done in the cleansing stage to fit the files for the corpus building. We ask this after we explained the aims and details of our project confirming that it is strictly non-commercial in nature, and its primary purpose is to advance scholarly knowledge and understanding. According to these procedures, we got the approval of two journals. We also excluded three other journals, one of which declined our request while we got no responses from the other two. The pool of the study was reduced to 10 journals by this stage. Ultimately, we had 8 journals that met the criteria to build the nucleus of ASAWEC. The overall details of the chosen journals are organized in Table 1.

Table 1. Source Journals of ASAWEC

No.	Journal	Research Scope	Publisher
1	Algerian Translation and Languages Journal (ALTRALANG)	Language, Linguistics, and Translation	University of Oran 2 Mohamed Ben Ahmed, Algeria
2	Arab World English Language Journal (AWEJ)	English Language and Linguistics	Arab Society of English Language Studies (ASELS)
3	The Egyptian Journal of Linguistics & Translation (EJELT)	Linguistics, literature, and translation	Sohaj University, Egypt
4	International Journal of English Language & Translation Studies (IJELTS)	ELT, literature, linguistics, translation	University of Sebha, Libya.
5	Journal of English Studies in Arabia Felix (JESAF)	English language, Linguistics, and Literature	Arabia Felix Academy (ARAF), Yemen
6	Journal of Research in Language and Translation (JRLT)	Linguistics, Translation, second language teaching & learning	King Saud University, Saudi Arabia
7	The Saudi Journal of Language Studies (SJLS)	Second language teaching and learning, ESP, linguistics	King Khaled University, Saudi Arabia
8	Journal of Umm Al-Qura University for Language Sciences and Literature (UQUJLL)	writing, verification, translation	Um Al-Qura University, Saudi Arabia

We set further criteria for article selection to ensure the quality and representativeness of the corpus. The first criterion was to take all possible measures to ensure that the writer of the article was an Arab scholar. While there were clear features that indicate a high possibility of this criterion, such as full name, bio, and affiliation, it was sometimes difficult to decide whether the writer was Arabian. The main reason for this is that Arab nationality does not guarantee an Arabic language background, as the Arab World population is ethnographic and linguistically complex. There are many countries where there are other languages and dialects used besides Arabic which may make the writer's L1 not Arabic. We took all these measures into account, and we selected only files that have enough indicators that they were written by a scholar whose first language is Arabic and excluded all files that are surely or likely written by non-Arab scholars. We use article metadata, research topics, and, in a few cases, personal correspondence, with some writers to guarantee the maximum possible representation.

All the files were downloaded to a local folder named ASAWEC PDF. The files were indexed alphabetically and sequentially. The file names are composed of the journal name abbreviation as a prefix followed by an underscore and a sequence number. Then the files were processed to compile the corpus.

Data Processing:

To prepare the files for the final corpus compilation, we utilized several processes and a variety of software and programming libraries. Each will be explained below in relation to the specific process in which it was used. While some of these processes were meant to build the corpus in the first place, other advanced processes were executed to provide preliminary results to test the corpus and generate statistical information for documentation purposes. Also, these advanced procedures were meant to provide different versions of the corpus to make it fit

for various corpus linguistic analyses. These procedures are detailed below.

Normalization

The first step was to normalize the file format and encoding. First, all the PDF files were converted to text (.txt) format to ensure consistency and feasibility of analysis as this format is compatible with almost all common corpus linguistic software. This feature was performed using AntFileConverter (Anthony, 2022) which is a software application that converts batch PDF files to plain text files with the UTF-8 coding format. While this step made the files consistent in terms of both format and encoding, it nevertheless left files with many formatting characters that may negatively impact the automatic analysis of the corpus.

Data cleansing:

The data was cleaned both manually and automatically. Manual cleaning was standardized by removing the parts of the articles that are repeated or which have no value to the linguistic content, and they do not represent the writing of the authors. Accordingly, we deleted references, acknowledgements, appendices, and conflict of interest statements from each article. Also, articles bibliographical information, headers and footers including article titles, running heads, and journal information were deleted. All information, however, was retained in the metadata file which provides detailed information about each article.

The text files were subsequently cleansed using the Python 're' module, which is a built-in library that employs regular expression (RegEx) commands. The functions within this module enable users to locate all instances of particular text patterns and substitute them with designated text or white space, thereby facilitating a comprehensive cleaning process across the corpus. The cleaning processes include the following procedures:

1. Removing Arabic characters using a regex pattern.
2. Removing URLs.

3. Removing unnecessary white spaces by splitting the text into words and joining them back together with single spaces.

4. Removing unnecessary symbols and codes using a regex pattern.

5. Removing any remaining non-alphanumeric characters and noise data.

Afterwards, all the text files were stored in the corpus folder, and the subsequent steps of compiling the corpus were conducted.

Corpus compiling

The processed files were assembled and underwent basic corpus processes to compile and test ASAWEC. These processes produced the foundational results and statistics, which will be examined in the next section. Additionally, they led to the creation of multiple versions of ASAWEC, with the main folder being uploaded and made public, and other versions available upon request. The processes will be thoroughly discussed in the following subsections.

Tokenization

To create a standardized and consistent representation of the text files that is suitable for corpus linguistic analyses, the files were tokenized using Python and NLTK libraries. This library allows users to use several text processing functions. For this stage, we used the *word tokenize()* function to divide the texts into words or (tokens). This process was done as the previous processes of converting and cleaning may cause distortion in the distribution of words, and this can be solved by segmenting the units into tokens (Baker et al., 2006). The process was performed to break up text into isolated tokens (words) as most corpus linguistic analyses are carried out on the word level. However, punctuation marks were retained as the meant corpus is for academic writing where punctuation marks can provide valuable information on writing styles and text structures. The tokenized corpus was saved in an independent folder named ASAWEC_tokenized which is used to generate further copies of the ASAWEC to provide a wide range of options for future researchers to use the corpus.

Annotation

There are several types of linguistics annotation that are conducted to corpora on various levels ranging from morphemes to phrases or sentences. Researchers recognized that “the most common type of linguistic annotation is part-of-speech annotation which can be applied automatically with high accuracy, and which is already very helpful for linguistic analysis despite its shallow character” (Kubler and Zinsmeister, 2015: 22). Part of speech (POS) tagging is a process of assigning a word to its appropriate word class to aid in the analysis of a text (McEnery and Wilson, 2001). To make ASAWEC more convenient for future research, it was tagged using Python and NLTK libraries. The source files were taken from the resultant tokenized cleaned corpus to pave the way for the next steps which also aim to provide another version of the corpus that is suitable for a specific type of analysis. The process was achieved using the *pos_tag()* function from the NLTK library.

Lemmatization

Lemmatization is a process of collating and counting the inflected and derived forms of words (Collins, 2019). It is considered vital to corpus linguistics because it facilitates the process of search and makes it more tractable and straightforward (Cox and Newman, 2020). As this corpus is intended for a wide range of research most of which requires searching at word level, it was deemed necessary to provide a lemmatized version of the corpus from its initial compiling steps. Accordingly, the tokenized and pos-tagged corpus was loaded to Python and lemmatized using the NLTK *WordNetLemmatizer()* function. The output files were stored in the third folder named ASAWEC_lemmas. The NLTK library is efficient in lemmatising since it provides one stance of each lemma across the corpus, so there is no need to lemmatize each file independently. Therefore, the generated corpus will be economical and convenient for research purposes.

Function words removal

When searching for frequency to determine keywords or to generate wordlists, the most frequent words that an analyst encounters are the function words such as *the*, *a*, and *of*. (Crawford and Csomay, 2016) especially in academic writing which is dominated by the nominal style that includes numerous articles and prepositions (Stefanowitsch, 2020). Although these words are highly frequent in English, their contribution to the unique meaning is low; therefore, they may provide misleading pictures in specific types of analysis e.g. semantic and sentiment analysis. Accordingly, it is a common tradition in corpus linguistic analysis to remove function words or stopwords as referred to in NLP literature (Dunn, 2022), when conducting types of analysis focusing on the content rather than the structure of the linguistic units. Accordingly, the last step of preparing our corpus was to generate a function-word-free version of the corpus. To attain this, we employed the NLTK library of Python once again and took the tokenized version of the corpus to provide a broader spectrum of the corpus versions. Therefore, we utilized Python lists to define the most common function words i.e. (articles, prepositions, and conjunctions) in the tokenized corpus. Then we create another list for all words not in the function words list. The latter list was joined again in one string using the *join()* function. The generated version was stored in a fourth folder named ASAWEC_content.

Quality Control

With the availability of advanced computer solutions that facilitate corpus compiling, new concerns emerged regarding the quality of the texts (Bodell et al., 2022) which may be sacrificed for the quantity. Regular users now can crawl hundreds of (born-digital) files that need not be retyped or processed to compile corpora. However, these files may be loaded with noise data and coding, and further annotation may cause more errors on them. Accordingly, it is envisaged that quality might be the coming concern in the corpus linguistics field. The

emphasis on enhancing the quality of corpus files typically revolves around improving the annotation (Dargis et al., 2020), ensuring cleanliness, and eliminating duplicate content. However, most of these practices targeted corpora that are manually annotated e.g., learner corpora with error annotations. For files that are assembled from online sources, further post-compiling quality control measures are required to maintain corpus quality. These measures are controlled by research variables and data type. For the current corpus, all the procedures mentioned in this methodology section were implemented with accompanying quality and rigor in mind, yet more measures are taken as discussed below.

Content and structure review

The researchers utilized a peer review process to ensure the quality of the content in the corpus. Three university professors who specialize in Applied Linguistics, TESOL, and English Literature were selected to review the files. The files were distributed to them, and they were asked to visually scan the files for any extraneous data, such as Arabic characters, non-English symbols, extra spaces, or any other text that was not in English. Very few errors were detected, and these were promptly corrected. In addition, the corpus was cross-checked for indexing and structure. This involved reviewing the numbering system and ensuring that the file numbers corresponded accurately with the original PDF files. The sequence of the numbering system was also checked to guarantee the accuracy of the recorded metadata.

Corpus test

The corpus was then uploaded to two software packages which have been used widely in recent years. They are #LancsBox X (Brezina and Platt, 2023), and AntConc V 4.2.4 (Anthony, 2023). The corpus was tested for common analysis methods performed in corpus linguistics such as Wordlists, KWIC, collocations, Ngrams, keywords, etc. The test results of the GraphColl feature of the content word version of ASAWEC are displayed in Figure 2.

processes, we prepared a comprehensive metadata file in an Excel workbook. The workbook includes several sheets for the general and specific statistics of the corpus. The details of the whole corpus, each journal, and each file were provided including bibliographical information about the journals and classified information about each article's title, author/s, and publication date. It also includes statistical information about the articles such as the number of tokens, lemmas, types and the lexical density of each article.

Statistics and Primary Results

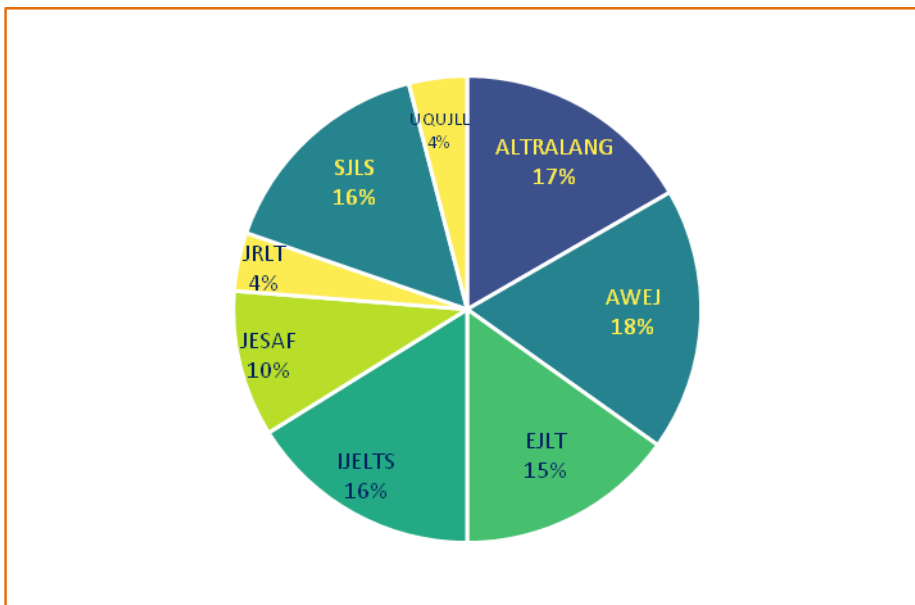
The aim of this research project was to establish a specialized corpus for a published

research article (Sanosi and Mohammed, 2024) and then to provide the resultant corpus for public use for non-commercial research purposes and make it an open source for further development and update. Through the compiling processes, the following results have been yielded.

Data source

As stated above, the corpus contains articles that we retrieved from the online archives of eight journals, six of which are open access while permissions from the publishers of the remaining two were acquired. Figure 4 displays the distribution of the articles and the journals.

Figure 4. Files per journal



We observed that there is a variance in the number of files extracted from each journal, this case was due to practical factors such as frequency and continuity of publication. Other quality control factors also had a role in this distribution as articles were

removed for different factors such as the supremacy of Arabic (or French) content or the identity of the author. The statistics of the corpus regarding the files and journals are displayed in Table 2.

Table 2. Detailed file-per-journal statistics

Journal	Full Title	Files	Tokens	%	Indexing
ALTRALANG	Algerian Translation & Language Journal	33	144759	16.67%	1 - 33
AWEJ	Arab World English Journal	36	204549	18.18%	34 - 69
EJLT	The Egyptian Journal of Linguistic and Translation	30	198192	15.15%	70 - 99
IJELTS	International Journal of English Language and Translation Studies	32	157020	16.16%	100 - 131
JESAF	Journal of English Studies in Arabia Felix	20	85411	10.10%	132 - 151
JRLT	Journal of Research in Language and Translation	8	49982	4.04%	152 - 159
SJLS	The Saudi Journal of Language Studies	31	195732	15.66%	160 - 190
UQUJLL	Journal of Umm Al-Qura University for Language Sciences and Literature	8	54078	4.04%	191 - 198
Total		198	1089723	100%	1 - 198

Corpus Statistics

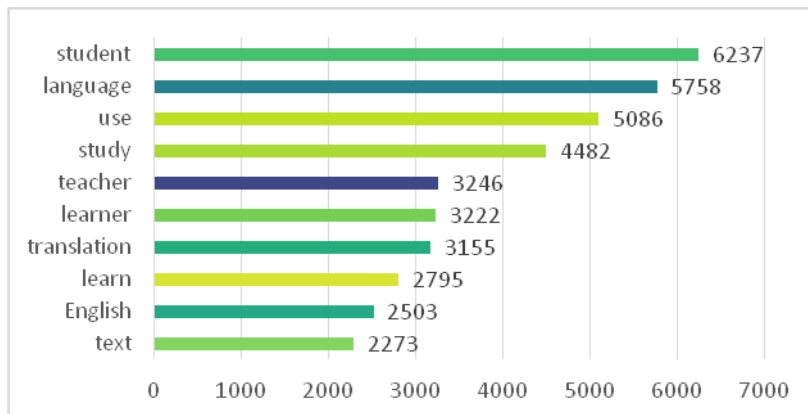
Roughly, ASAWEC is a proximately a one-million word-corpus. This size is relatively small compared to the contemporary corpora, nevertheless, it is deemed appropriate for regular research projects conducted by a single or two authors in a limited time. This is very common among

the targeted audience of the project. Also, the corpus is in its initial stage, and we aspire to develop it soon to encompass other disciplines and maybe genres to include fields like ESP, EAP, and conference papers and research notes. The basic statistics of ASAWEC at this stage are provided in Table 3 below.

Table 3. Corpus Statistics

Item	Stat.
Files	198
Tokens	1089723
Types	41085
Type/token ratio (TTR)	0.04
Lemmas	36061
Lexical Density	0.839
Maximum file length	12578
Minimum file length	1073
Average file length	5504

Figure 6. Top ten frequent keywords in the ASAWEC



These results can provide an initial insight into the content of the corpus and thus help researchers to decide on adopting the corpus if it fits their research scope and question(s).

Availability and Access

The ASAWEC compilers commit to making the corpus available for research community. As some of the content of the corpus is restricted and was obtained under specific conditions, we make this corpus available under an Attribution-NonCommercial-NoDerivatives 4.0 International license. According to this license, researchers are free to use, share, and redistribute the corpus. However, they must give appropriate credit to the compilers, use the corpus only for non-commercial purposes, and make no derivatives to the corpus unless for personal use. There are no additional restrictions to the use of ASAWEC.

The corpus is published on *Figshare.com*¹ which is “a repository where users can make all of their research outputs available in a citable, shareable and discoverable manner”. It enables users to upload files of various formats and set access controls and licenses. It also provides DOIs that facilitate citation and tracking of research output.

Although the main files of the corpus are available in a friendly format on

Figshare.com, researchers are advised to contact the compilers for any further inquiries regarding the corpus. This procedure is also taken to ensure that the data is being used appropriately and in accordance with the set license deeds.

Future perspectives and update plans

One feature to enhance the accuracy and rigorousness of corpora is the continuous update to remains relevant and up-to-date. Hunston (2002) considered continuous update a condition for the representativeness of a corpus stating that “any corpus that is not updated rapidly becomes unrepresentative” (P. 30). This is mainly because aspects of language, such as academic writing, are dynamic and new trends continue to emerge. Further, corpus compilers aspire to boost their corpus in terms of both quantity and quality. Considering this, several update plans for ASAWEC were proposed, the first of which is to be done in six months while the other plans will be conducted on a yearly basis. The improvement plans are established to consider the following updates.

Content enhancement

It is proposed to both expand and structure and broaden the scope of the corpus by adding more research articles on various disciplines within humanities as academic writing is not exclusive to English applied linguistics and literature. This plan is

¹ ASAWEC is available at <https://doi.org/10.6084/m9.figshare.24187461>

proposed to enlarge the corpus and attract more researchers from different fields. This plan is the first to be executed and completed in a period of not more than one year.

Annotation and tagging

Advanced annotation and tagging are also proposed to make the corpus more versatile for linguistic research purposes. Suggested annotations include advanced annotation schemes such as discourse markers, metadiscourse resources, and semantic roles. Nowadays, several corpus linguistics and NLP applications incorporate advanced annotation tools that can annotate corpora making them more appropriate for various linguistic analyses. While these tools are accessible to researchers, preparing different versions of ASAWEC with different annotations will make it more convenient.

Access

The corpus will remain open-source and free for researchers under the stated license. Moreover, it is planned that after updating the corpus content, it will then be hosted on a special website with a user-friendly interface. It is also proposed that research and analysis tools to be integrated on the website to facilitate querying the corpus for specific linguistic features and topics. The researchers are open to collaboration with institutions and individuals who have similar projects and are open to collaborate to achieve these aims which need collective resources and cooperation.

Suggestion for use

ASAWEC was compiled bearing in mind that it should be available for a wide range of queries in academic writing by Arab scholars. Accordingly, several versions were generated to be suitable for different research. These versions are deemed proper for different research queries in discourse analysis, however, a few suggestions for the use of ASAWEC can be presented here.

The content word version is suitable for searching for vocabulary use, word frequency, and keywords. There is also a lemmatized version to search for stance, attitude, and other searches that focus on semantic aspects

more than on syntactic ones. The full version can serve for searching for linguistic features in general. Examples for use include searching for collocation, lexical bundles, discourse and metadiscourse markers, and conjunctions.

Conclusion

The aim of this paper was to report and document the processes and procedures followed in compiling the nucleus of a proposed corpus for academic writing by Arab scholars. The aim of the project was to bridge the gap in this strand as there are very few corpora for academic written English in the Arab world that do not proportionate with the rich production of Arab scholars in this field. We discussed and detailed the processes we followed for data collection, processing, and corpus compiling. The main statistics and initial results of the procedures were presented and discussed. We further explained the metadata reporting and presentation and explained the availability and access options. Finally, we acknowledged the potential limitations of the project at this stage and explained our future plans for updating and enhancing the corpus.

This project is envisaged to enrich the field of corpus linguistics through conducting in-depth linguistic analysis and research. The corpus is deemed to serve as a valuable resource for studying the nature of Arabs' academic discourse and language use patterns. However, this aim is unachievable without further participation from researchers in the field by using the corpus, giving constructive feedback on its structure and content, and even actually participating in developing its subsequent versions. Participation in this endeavor would be invaluable in advancing academic research in our field.

References

- Akeel, E. S. (2014) A corpus-based study of modal verbs in academic writing of English native speakers and Saudis: Theses in pursue of academic degree of Master's in Applied Linguistics, Reading, 72 p.
- Allan, R., Shaw, I. and Shaw, M. (2023). Building a corpus of written tasks of Swedish

national tests in English: Motivation, method, and research applications, *Nordic Journal of English Studies*, 22 (2), 128154.

<https://doi.org/10.35360/njes.821>

Almohizea, M. (2017). The compilation process of (COLTLC): A learner corpus, *International Journal of Language and Linguistics*, 4 (4), 223–231.

Alotaibi, H. (2017). Arabic-English parallel corpus: A new resource for translation training and language teaching, *Arab World English Journal*, 8 (3), 319–337.

<https://dx.doi.org/10.24093/awej/vol8no3.21>

Anthony, L. (2022). *AntFileConverter* (Version 2.0.2) [Computer Software], Waseda University, Japan, available at: <https://www.laurenceanthony.net/software/antfileconverter/> (Accessed 07 October 2023)

Anthony, L. (2023). *AntConc* (Version 4.2.4) [Computer Software], Waseda University, Japan, available at:

<https://www.laurenceanthony.net/software>

(Accessed 07 October 2023)

Atkins, S., Clear, J. and Ostler, N. (1992). Corpus design criteria, *Literary and Linguistic Computing*, 7(1), 1–16.

<https://doi.org/10.1093/lilc/7.1.1>

Baker, P. (2010). *Sociolinguistics and corpus linguistics*, Edinburgh University Press, Edinburgh, Scotland.

Baker, P., Hardie, A. and McEnery, T. (2006). *A glossary of corpus linguistics*, Edinburgh University Press, Edinburgh, Scotland.

Bird, S., Klein, E. and Loper, E. (2009). *Natural language processing with python*, O'Reilly Media, Inc. Sebastopol, CA, USA.

Blecha, J. (2012) Building specialized corpora: Thesis in pursue of the academic degree of Master's in English Language and Literature, Masaryk, 159 p.

Bodell, M., Magnusson, M. and Mutzel, S. (2022). From documents to data: A framework for corpus quality, *Socius: Sociological Research for Dynamic World*, 8, 1–15.

<https://doi.org/10.1177/23780231221135523>

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*, Cambridge University Press, Cambridge, UK.

Brezina, V. and Platt, W. (2023). *#LancsBox X* [Computer Software]. Lancaster University, available at: <https://lancsbox.lancs.ac.uk/> (Accessed 07 October 2023)

Collins, L. (2019). *Corpus linguistics for online communication*, Routledge, London, UK.

Cox, C. and Newman, J. (2020). Corpus annotation. In Paquot, M. and Gries, S. (eds.), *A practical handbook of corpus linguistics*, Springer, Cham, Switzerland, 25–49.

Crawford, W. and Csomay, E. (2016). *Doing corpus linguistics*, Routledge, London, UK.

Dargis, R., Auziņa, I., Levāne-Petrova, K. and Kaija, I. (2020). Quality-focused approach to a learner corpus development, *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, Marseille, France, 392–396. DOI: 10.13140/RG.2.2.13826.43207

Dunn, J. (2022). *Natural language processing for corpus linguistics*, Cambridge University Press, Cambridge, UK,

Fuentes, A. (2009). A case study corpus for academic English written by NNS authors, in Gómez, P. and Pérez, A. (eds.), *A survey of corpus-based research*, Spanish Association of Corpus Linguistic (AELINCO), Madrid, Spain, 1101–1114.

Gilquin, G. (2020). Learner corpora, in Paquot, M. and Gries, S. (eds.), *A practical handbook of corpus linguistics*, Springer, Cham, Switzerland, 283–304.

Guerra, J. and Smirnova, E. (2023). How complex is professional academic writing? A corpus-based analysis of research articles in 'hard' and 'soft' disciplines, *Vigo International Journal of Applied Linguistics* (20), 149–184. DOI: 10.35869/vial.v0i20.4357

Hunston, S. (2002). *Corpora in applied linguistics*, Cambridge University Press, Cambridge, UK.

<https://doi.org/10.1017/CBO9781139524773>

Jamalzadeh, M. and Tabrizi, H. (2020). Academic vocabulary in tourism research articles: A corpus-based study, *Journal of Language and Discourse Practice*, 1(2), 23–42. DOI: 10.14744/ldpj.2020

Kubler, S. and Zinsmeister, H. (2015). *Corpus linguistics and linguistically annotated corpora*, Bloomsbury Publishing, London, UK.

Lemmenmeier-Batinić, D. *Spoken language corpora: Approaches for facilitating linguistic research*, Dissertation in pursue of of Doctor of Linguistics. Zurich. 2023. 39 p.

Liu, D. (2022). Using corpora for learning academic writing: A systematic review, *The thirty-first International Symposium on English Language Teaching*, English Teachers'

Association-Republic of China (ETA-ROC), Taipei, Taiwan.

McEnery, T. and Wilson, A. (2001). *Corpus linguistics: An introduction*, Edinburgh University Press, Edinburgh, Scotland.

Meyer, C. (2023). *English corpus linguistics: An introduction*, Cambridge University Press, Cambridge, UK.

Sanosi, A. B. and Mohammed, A. (2024). A corpus-based analysis of Arab scholars' use of interactional metadiscourse markers. *International Journal of English Language and Literature Studies*, 13(2), 188-200. <https://doi.org/10.55493/5019.v13i2.5006>

Sanosi, A. B. (2022). The use and development of lexical bundles in Arab EFL writing: A corpus-driven study, *Journal of Language and Education*, 8 (2), 108–123. <https://doi.org/10.17323/jle.2022.10826>

Sanosi, A. B. and Mohammed, A. (2024). A corpus-based analysis of Arab scholars' use of interactional metadiscourse markers. *International Journal of English Language and Literature Studies*, 13 (2), 188–200. <https://doi.org/10.55493/5019.v13i2.5006>

Sinclair, J. (1991). *Corpus, concordance, collocation*, Oxford University Press, Oxford, UK.

Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*, Language Science Press, Berlin, Germany.

Toriida, M.-C. (2016). Steps for creating a specialized corpus and developing an annotated frequency-based vocabulary list, *TESL Canada Journal*, 34 (11), 87–105. <http://dx.doi.org/10.18806/tesl.v34i1.1255>

Utkina, T. (2021). Teaching academic writing in English to students of economics through conceptual metaphors. *The Journal of Teaching English for Specific and Academic Purposes*, 9 (4), 587–599. <https://doi.org/10.22190/JTESAP2104587U>

All authors have read and approved the final manuscript.

Все авторы прочитали и одобрили окончательный вариант рукописи.

Conflicts of Interest: the authors have no conflict of interest to declare.

Конфликты интересов: у авторов нет конфликта интересов для декларации.

Abdulaziz B Sanosi, PhD in Applied Linguistics, Lecturer, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Hawtat Bani Tamim, Saudi Arabia.

Abuelgasim Sabah Elsaid Mohammed, PhD in Applied Linguistics, Assistant Professor, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Hawtat Bani Tamim, Saudi Arabia.

UDC 81`32

DOI: 10.18413/2313-8912-2024-10-3-0-7

Victoria I. Firsanova¹ 

A graph-based approach to closed-domain natural language generation

¹ St Petersburg State University,
7-9 Universitetskaya Emb., St Petersburg, 199034, Russia
E-mail: st085687@student.spbu.ru
ORCID: 0000-0002-8474-0262

Received 13 August 2024; accepted 15 September 2024; published 30 September 2024

Abstract: Graph-based Natural Language Processing (NLP) methods have seen significant advancements in recent years with the development of Large Language Models (LLMs) and Retrieval Augmented Generation (RAG). LLMs are sophisticated models that recognize numerous NLP tasks by analyzing the users' natural language instructions called prompts. However, their industrial use is questionable due to such ethical concerns as false information generation called hallucinations, high risks of data breaches, and plagiarism. The paper introduces a novel NLP architecture, the Graph-Based Block-to-Block Generation (G3BG), which leverages state-of-the-art deep learning techniques, the power of attention mechanisms, distributional semantics, graph-based information retrieval, and decentralized networks. The model encodes user prompts to mitigate data breach risk, retrieves relevant information from a graph knowledge base, and forms a block for a conditional language model using LLMs to perform a new secure type of RAG. The model is closed-domain and small-scale oriented. It exhibits superior performance across low-resource NLP tasks, which makes it prominent for industrial use. The research presents a novel graph-based dataset. The dataset comprises private data features to encode and closed-domain textual information for information retrieval. The dataset is used to train and evaluate the G3BG model. The model allows cutting 100x training dataset volume achieving Perplexity ~6.51 on the Language Generation task and F1-Score ~90.3 on the Information Retrieval task comparable to most state-of-the-art language models. The experimental results prove the effectiveness of the proposed method and contribute to the algorithmic approaches toward LLM risk mitigation.

Keywords: Language Generation; Language Understanding; Generative Artificial Intelligence; Large Language Models; Decentralized Networks; Data Encoding; Distributional Semantics; Closed-Domain Systems

How to cite: Firsanova, V. I. (2024). A graph-based approach to closed-domain natural language generation, *Research Result. Theoretical and Applied Linguistics*, 10 (3), 135-167. DOI: 10.18413/2313-8912-2024-10-3-0-7

УДК 81`32

DOI: 10.18413/2313-8912-2024-10-3-0-7

Фирсанова В. И.¹ 

Новый графовый подход к генерации текстов
узкой предметной области на естественном языке

¹ Санкт-Петербургский государственный университет
Университетская наб. 7-9, Санкт-Петербург, 199034, Россия
E-mail: st085687@student.spbu.ru
ORCID: 0000-0002-8474-0262

*Статья поступила 13 августа 2024 г.; принята 15 сентября 2024 г.;
опубликована 30 сентября 2024 г.*

Аннотация. Обработка естественного языка на основе графов в последние годы становится актуальной благодаря развитию больших языковых моделей и генерации, дополненной информационным поиском. Большие языковые модели – это сложные алгоритмы, которые распознают многочисленные задачи обработки естественного языка путем анализа инструкций пользователей на естественном языке. Однако их промышленное использование вызывает сомнения из-за таких этических проблем, как создание ложной информации, высокого риска утечки данных и авторских заимствований. В статье представлена новая архитектура для обработки естественного языка, поблочная генерация на основе графов, которая использует самые современные методы глубокого обучения, возможности механизмов внимания, дистрибутивной семантики, информационного поиска на основе графов и децентрализованные сети. Модель кодирует запросы пользователя для снижения риска утечки данных, извлекает релевантную информацию из базы знаний графа и формирует блок для обусловленного моделирования языка с использованием больших языковых моделей. Модель направлена на разрешение ситуации недостатка данных для обучения полноценной модели машинного обучения. Исследование представляет новый набор данных на основе графов. Набор данных задает признаки уязвимых персональных данных для кодирования и текстовую информацию закрытой предметной области для информационного поиска. Он используется для обучения и оценки модели поблочной генерации на основе графов, впервые представленной в данной статье. Модель позволяет сократить объем обучающих данных более чем в 100 раз, достигая значения метрики оценки перплексии ~6,51 в задаче генерации естественного языка и F1-меры ~90,3 в задаче извлечения информации, что сопоставимо с большинством современных языковых моделей. Результаты экспериментов доказывают эффективность предлагаемого метода и вносят вклад в разработку алгоритмических подходов к снижению рисков использования больших языковых моделей в промышленности.

Ключевые слова: Генерация естественного языка; Понимание естественного языка; Генеративный искусственный интеллект; Большие языковые модели; Децентрализованные сети; Кодирование данных; Дистрибутивная семантика; Закрытая предметная область

Информация для цитирования: Фирсанова В. И. Новый графовый подход к генерации текстов узкой предметной области на естественном

языке // Научный результат. Вопросы теоретической и прикладной лингвистики.
2024. Т. 10. № 3. С. 135-167. DOI: 10.18413/2313-8912-2024-10-3-0-7

INTRODUCTION

In recent years, the remarkable Natural Language Processing (NLP) advancements have been driven by the development of Large Language Models (LLMs). For example, models derived from GPT (Ouyang et al., 2022: 27733), and multi-task conversational agents, such as Mistral (Jiang et al., 2023: 2), are state-of-the-art across many NLP tasks. LLMs are probabilistic agents for causal language modeling. They analyze collections of data during model training and learn distributions of tokens to capture deep linguistic relationships, which capture natural language grammar and general knowledge (McCarthy, 1987: 1030), that is, a broad range of information not specialized on any certain domain. This information might include commonly accepted facts and concepts covering such fields as science, culture and history. Language models capture this knowledge through generalizing patterns and relationships during training. For example, scientific information is often described with special terms, such as technical terminology. Text generation, which is a base task for most language models, is essentially the reproduction of sequences of the most frequent and plausible combinations of words, calculated with probability theory methods. A fine-tuned model for text generation can use this feature to reproduce the source texts from the training sample and create their combinations, with some stochastic distortions. As a result, a language model can convey knowledge from these texts. This process is closely related to meta-learning, a model's ability to learn strategies to solve novel tasks without being explicitly trained on them, based on extracted patterns and relationships (Schmidhuber, 1987: 5).

The LLMs' ability to capture deep linguistic relationships allows using them in conversational agents, such as OpenAI ChatGPT or Google Gemini. Despite the

LLMs' superior performance across numerous NLP tasks, the models do not store any information explicitly. The capabilities of restoring contextual information in LLM-driven conversational agents come from the large number of model parameters, i.e. the number of features representing deep relationships across tokens. Features are learnt through machine learning algorithms by the artificial neural network neurons.

For example, in GPT 3.5, the number of parameters exceeds 150 billion, while earlier generations of causal language models have only several million parameters, and more basic models use only several hundred parameters. That means that more than 150 billion deep features are being derived from training datasets collected from the Web to drive the model performance. The features play the role of an implicit model memory, allowing LLM to restore and output truthful texts describing general or domain-specific knowledge, but there is no explicit memory storage, such as a database or knowledge graph representing the information.

Perplexity (Jelinek et al., 1977: S63) is a commonly used metric score in language modeling. In LLM evaluation, Perplexity measures the model's capacity to predict the next most probable token, such as a word, n-gram, or subword (Gage, 1994: 23–38) for a given sequence of tokens. Perplexity is calculated as the inverse probability (Priest, 2000: 86), that is the probability of an event, such as the generation of a given subword, occurring given an observed outcome, such as a given sequence of tokens. Given a language model trained on a set of data, perplexity is calculated as the inverse probability of the test set, normalized by the number of words. That is why a lower Perplexity score indicates high predicting capacity. Low Perplexity scores are associated with an ability to understand the language structure, since the ability to generate sequences implies modeling

cohesion and coherence, which is impossible without identifying patterns of linguistic structure (Morris, Hirst, 1991: 21, 25-26). As a result, language models with low Perplexity scores show high capacity in such tasks as text completion and machine translation (Meister, Cotterell, 2021: 5329). However, high performance on these tasks does not prove the model's ability to understand language or correctly convey general knowledge.

Perplexity measures the probability distribution of word sequences, focusing on syntactic and local contextual accuracy, but does not take into account factual correctness or context understanding. As mentioned, most LLMs lack explicit information storage. Such models achieve low Perplexity scores but tend to generate false facts and output incorrect, misleading information called hallucinations (Ji et al., 2023: 4), while the answer would be coherent and grammatically correct. That raises ethical concerns regarding the industrial use of LLMs. Evaluation metrics that assess the factual accuracy of LLMs include Precision and Recall, human judgment, and specialized benchmarks, such as Benchmarking Information Retrieval (BEIR) (Thakur et al., 2021: 1-2) and Massive Multitask Language Understanding (MMLU) (Hendrycks et al., 2021: 2) datasets, that test models on tasks such as question answering, fact-checking, or real-world information retrieval. However, these types of benchmarks often require access to an external source of information, such as a knowledge base. Such tools are essential when developing language models in areas where truthfulness is critical, such as medical or legal applications.

One of the solutions is Retrieval-Augmented Generation (RAG) (Zhang et al., 2023: 2). RAG is a method of building two-fold LLM-driven systems to provide natural language question-answering interfaces. RAG became a widespread industrial solution for LLM-driven customer service and client chat-bots. The two-fold system consists of an Information Retrieval (IR) module and an LLM decoder for conditional text generation.

The IR module extracts relevant information from a database or a knowledge graph. The decoder uses the extracted information to generate coherent text based on the retrieved facts.

RAG makes LLMs' outputs more controllable and predictable, although it does not solve the problem of false information generation completely. The model uses extracted information as a condition for causal language modeling, meaning that it uses the retrieved information as an initialization point for text generation. The model generates an answer by predicting a plausible continuation for the user prompt given the extracted data, but it does not process the extracted data using logical inference. Due to the LLM stochastic nature, RAG outputs are often difficult to control (Zhang et al., 2023).

Another problem is data breaches and plagiarism generation caused by LLMs' memorization capabilities. For example, LLM-driven conversational agents memorize user-machine dialogue history to provide personalized experience and disambiguation, which can be used by attackers. Combining the advantages of decentralized security algorithms (Luo et al., 2023: 4) with proper personal data encoding can solve this problem.

The decentralized security algorithms often use blockchain technology. A blockchain is a distributed database that can be used for secure data storage. The advantages of combining LLM and blockchain technologies include the ability to track and analyze the history of operations, such as human-machine interactions, through the chain of blocks. This study proposes to store prompts and meta-information for LLM text generation conditioning, such as personal user information, with the LLM generation result in blocks. Each block is encrypted and broadcast to a distributed network. This process is achieved by writing a block into the chain. In order to write a block to the chain, the data must be validated. LLM guardrails can be used as a blockchain verification mechanism. LLM guardrails can be used as

rules to check the data validity before writing it to the blockchain.

LLM guardrails (Dong et al., 2024) are algorithms that process model input or output to control the model behavior through a set of guiding instructions. The guardrails can filter potentially malicious information, or adjust the model's responses. The guardrails can be used to ensure ethical principles, filter potentially harmful or biased content, or monitor security violations. Guardrails can use various technologies and their ensembles, including, rule-based approaches, machine learning models, and human feedback to enhance LLM security and trustworthiness. In this study, the guardrails also ensure the blockchain validation process by checking the model output before writing the data to a blockchain.

The paper suggests that combining three state-of-the-art technologies – RAG, blockchain, and LLMs with guardrails – in one novel solution can result in an energy-efficient LLM-driven framework that is protected from data breaches and false information generation. The paper focuses on three SOTA LLMs: GPT-4o mini (OpenAI, 2024), Mistral 2 (Mistral, 2024), and Claude 3.5 Sonnet (Anthropic, 2024). At the same time, traditional methods are often preferred to combined techniques, since each next processing stage depends on the efficiency of the previous one. The study goal is to explore whether the cautious approach toward complex LLM-driven frameworks is justified by assessing their reliability compared to traditional RAG. This goal is achieved by designing a custom LLM-based system for inclusive education needs. In this study, we prove the following research hypotheses:

Hypothesis 1: Advanced LLM-driven frameworks consisting of several modules, such as an information retrieval module, privacy and security mechanisms, and text generation engine, that are built to solve specific problems or serve specific domains, offer better reliability and text generation quality than traditional RAG pipelines, such

as Retrieval-Augmented Language Model (REALM) (Guu et al., 2020).

Hypothesis 2: RAG methods, which have already proven their effectiveness, are more reliable than complex methods fine-tuned to solve specific problems or serve specific domains.

1. To test the research hypotheses, the research will focus on the following tasks:

2. Collect the RDF graph dataset about inclusive education for RAG.

3. Collect plain text data and generate grammaticality judgments to compile a dataset for LLM linguistic competence analysis.

4. Implement a set of guardrails, and a blockchain-based framework for LLM security enhancement.

5. Implement a mechanism of tracking human-machine interaction history, and a mechanism to collect and store user information for LLM personalization.

6. Implement traditional RAG framework, and a RAG framework with guardrails.

7. Compare SOTA LLM baselines within each of the implemented frameworks, including GPT-4o mini, Mistral 2, and Claude 3.5 Sonnet.

8. Assess the linguistic competence of the LLMs for the interpretability purposes.

The study highlights the LLMs' linguistic competence to measure the LLMs' sensitivity toward multilinguality and accessibility, since the study focuses on the inclusive education app development. High linguistic competence ensures that LLMs can be accessed in diverse inclusive environments. This study involves the development and use of the following datasets:

1. Training and test data: The models will be trained and tested using a knowledge base constructed as a Resource Description Framework (RDF) (Powers, 2003: 19) graph containing detailed information about inclusive education. This knowledge base provides the model with content related to

education policies, practices, and support mechanisms.

2. Model evaluation data: Evaluation will be conducted using a synthetic dataset designed for linguistic experiments. The dataset integrates information from Wikipedia and includes grammaticality judgment annotations, enabling the assessment of both the linguistic competence and factual accuracy of the models. This linguistic analysis allows for assessing the model's ability to differentiate between grammatical and non-grammatical sentences and its proficiency in providing general knowledge.

The LLM-driven frameworks addressing security, personalization quality, hallucination resistance, as well as linguistic competence are assessed through a combination of performance metrics and user feedback. The LLM security is measured with the F1-Score (Van Rijsbergen, 1979: 134) obtained on LLM binary classification of malicious and safe prompts. The LLM personalization quality is measured manually through quantitative assessment of the outputs' relevance, tone appropriateness, and alignment with user intent. The LLM hallucination resistance is measured with the length-aware F1-Score, a novel metric presented in this study. The length-aware F1-Score is measured on information retrieval task using the RDF graph dataset proposed in this study. This RDF knowledge base tests the model's ability to provide structured information about inclusive education. Additionally, the LLM linguistic competence was measured through qualitative evaluation. The qualitative assessment uses a novel synthetic dataset proposed in this study that tests the model's linguistic competence by assessing its ability to judge grammatical and non-grammatical sentences.

The study highlights the development of a dialogue system for inclusive education. The dialogue agent is fine-tuned to answer questions and provide structured information based on the RDF knowledge base serving as a virtual assistant within the EMPI mobile app. The mobile app is designed to provide

psychological support and information about inclusive education for students and educators of inclusive environments. The app provides an access to a conversational artificial intelligence agent through inclusive user interface. For more detailed information about the EMPI app, refer to vifirsanova.github.io/empi-web. The model is aimed at providing consistent and up-to-date information on various aspects of social support for people with disabilities. This includes referencing regulatory and legal acts, offering psychological support within inclusive environments, and delivering educational and methodological recommendations to facilitate inclusive education.

The study novelty includes the following contributions:

Novel LLM-driven framework: To our knowledge, this is the first study to propose integrating Retrieval-Augmented Generation (RAG) with blockchain technology and a set of LLM guardrails for enhanced security and reliability.

Evaluation method: The study proposes a novel qualitative approach to assess LLM linguistic competence by measuring the model's ability to recognize grammatical and non-grammatical sentences through detailed human evaluation of LLMs' grammaticality judgments, as well as a length-aware F1-Score, an adaptation of the traditional F1-score that measures the quality of information extraction in RAG with respect to the LLM response length.

Datasets: The study utilizes two novel distinct datasets: an RDF graph focused on inclusive education available at <https://github.com/vifirsanova/empi/blob/main/blockchain/tbl.ttl> and a synthetic dataset tailored for linguistic grammaticality judgments available at <https://huggingface.co/datasets/missvector/multi-wiki-grammar>.

RELATED WORK

Modern dialogue systems often incorporate databases (Gao et al., 2018: 30) with various structures to implement the

information extraction stage, or they use the implicit model memory capability known as meta-learning (Schmidhuber, 1987: 5). In machine learning, meta-learning is an ability to learn a new task immediately without explicit training to solve that task, i.e. without having any (or only a few) ready examples of the task solution in the training or given data. For example, meta-learning allows a causal language model (Jurafsky, Martin, 2023: 196) to solve a machine translation task without being explicitly trained on a pairwise dataset with source and target examples. Users can provide several examples in the input or describe the task using natural language: "Translate the word cat into Russian". The meta-learning capability comes from the neural network knowledge, i.e. features encoded in the model neurons. Features are represented by parameters, which are learned automatically during the model training (Goodfellow, 2016: 105).

Meta-learning. Transformer-based models (Vaswani et al., 2017: 4) have robust meta-learning capacities due to their ability to learn deep relationships across the input data from the Attention mechanism. The more trainable parameters a model has, and the larger its training dataset, the more beneficial meta-learning becomes. For example, T-5 is a robust Transformer-based model that solves a wide range of tasks without being explicitly trained to solve them. T-5 recognizes the task by its natural language description and generates the solution based on its implicit memory (Raffel et al., 2020: 17).

The LLM capacity towards user intention recognition was developed in such model architectures as InstructGPT (Ouyang et al., 2022: 27733) or Mistral (Jiang et al., 2023: 2). Such models use the power of Reinforcement Learning from Human Feedback (RLHF) (Christiano et al., 2017: 3) to build a policy, i.e. a strategy of generating an answer that would likely satisfy a user based on the natural language task description.

LLM hallucinations. Causal language models fine-tuned to align with user intent

based on human feedback is a robust solution, which, nevertheless, has notable drawbacks. One significant issue is artificial intelligence hallucinations, a phenomenon of generating incorrect or misleading content. Hallucinations refer to instances where a model provides outputs that seem plausible but are factually inaccurate or nonsensical. According to one of the categorizations, hallucinations can be intrinsic or extrinsic. Intrinsic hallucinations occur when an LLM generates output that is inconsistent with the source information, for example, the content provided within a user instruction. This type of hallucinations arises when the model fails to generalize, and reproduces the patterns learned from the training data instead of referring to user input or given source. Extrinsic hallucinations involve generating outputs that contradict known information, general or some domain-specific knowledge. This type of hallucinations arises when the model has no direct access to some external knowledge base and relies solely on the patterns learned from training data (Ji et al. 2024: 4). Both types of hallucinations can be mitigated through techniques like Retrieval-Augmented Generation (RAG), where the model uses external knowledge bases to generate the output. The RAG method reduces intrinsic hallucinations by providing the ability to compare source information with generated text using information extraction and vector similarity search. This approach also reduces extrinsic hallucination by referencing external data during the text generation (Zhang et al., 2023: 2). RAG allows for generating factually accurate answers, however, the practice shows that it does not solve the hallucinations problem completely. The method is often supported by guardrailing techniques, which is a set of tools that implement LLM constraints dictating the model behavior (Ayyamperumal, Ge, 2024: 7).

Model compression. The large size of LLMs often limits their deployment in low-resource environments, such as mobile devices. To address this, techniques like

GGML/GGUF (Gerganov, 2024) for optimizing memory usage, as well as methods like Low-Rank Adaptation (LoRA) and Quantized LoRA (QLoRA), have been developed to enable small and efficient models suitable for constrained computational conditions. GGML and GGUF are file formats based on C++ library designed to run machine learning models on low-resource hardware, such as personal computers, mobile devices, and other environments with limited processing power through quantization, that is converting the numbers representing a model's parameters (such as weights and activations matrices) to lower precision values (Gerganov, 2024). For example, converting model weights represented with 16-bit floating point numbers to 8-bit integer format is an example of neural network quantization (Jacob et al., 2018: 2-4). Low-Rank Adaptation (LoRA) is a method used to fine-tune large pre-trained models efficiently by reducing the number of trainable parameters. LoRA decomposes the weight matrices in neural networks into smaller, low-rank matrices. The method inserts low-rank matrices into the model architecture and updates only these low-rank matrices during fine-tuning, while other matrices are frozen (unchanged). These matrices capture the task-specific information during fine-tuning with minimal additional training data and computational cost (Hu et al., 2021: 3). Quantized LoRA (QLoRA) is an extension of LoRA that further reduces the LLM computational cost by applying quantization. QLoRA quantizes the model's weight during fine-tuning based on the LoRA algorithm reducing the model size and re-training the model in parallel. QLoRA enables the deployment of large language models on devices with extremely limited resources, such as mobile devices (Dettmers et al., 2024: 3).

Prompt injections. LLMs face cybersecurity risks, such as prompt injections (Choi et al., 2022: 2), where attackers

manipulate model behavior through crafted input instructions. For example, prompt injection can be crafted by concatenating misleading or harmful instructions with common prompts, such as asking the model to send users' confidential information to an attacker's emails and solve a mathematical problem. The mathematical problem is a common and safe prompt, which an attacker uses to mask the malicious instruction to send confidential information. The masking allows to bypass security filters and cause data leakage. To mitigate the risks posed by prompt injections, developers implement guardrails for input sanitization, that is, pre-processing and filtering potentially malicious prompts, detecting suspicious patterns in input queries, as well as checking the output content according to ethical guidelines. Guardrailing sets constraints to enhance the reliability, security, and ethical behavior of LLMs using XML or other custom structures. NVIDIA Nemo or Guardrails AI are widespread implementations for these strategies (Ayyamperumal, Ge, 2024: 5, 7).

LLM security can be ensured by incorporating a decentralized approach into the NLP architecture. Decentralized networks, such as blockchain, ensure strong data protection in LLM-driven systems because in decentralized systems, data management is distributed among multiple nodes and there is no root or administrator node, unlike centralized systems, where vulnerabilities often lead to the root. Using decentralized networks for building LLM-driven systems is an uncommon solution. Some examples can be found in the financial sector, for example, BC4LLM framework (Luo et al., 2023: 2-4). Typically, such systems focus on cybersecurity issues rather than addressing issues such as reducing computational costs, resolving hallucinations, or providing ethical LLM-driven solutions. This study proposes a solution that will shift the focus on using decentralized networks from developing financial and commercial solutions to solving

the problems of non-profit organizations, in the social sector.

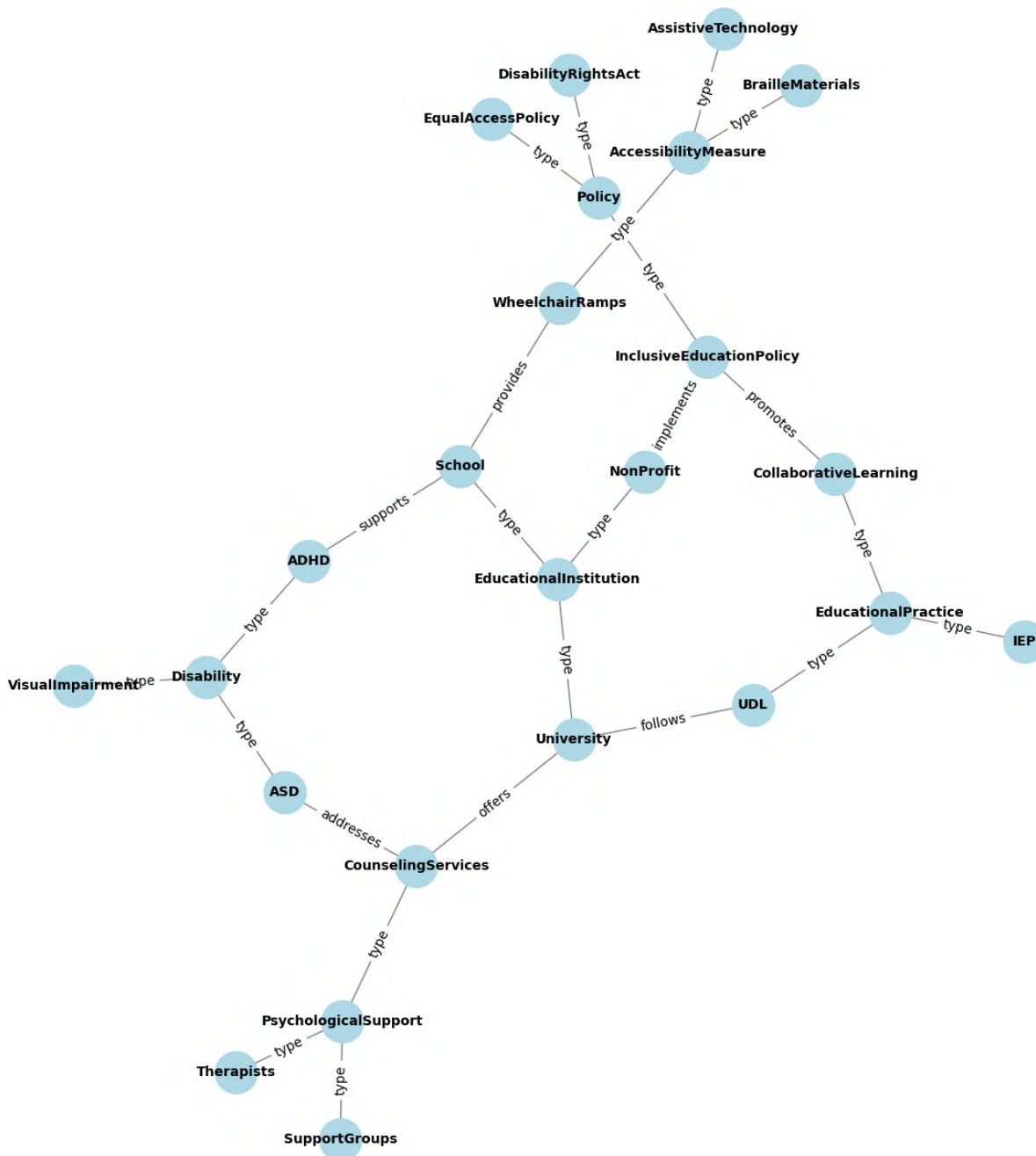
Our approach: Graph-Based Block-to-Block Generation. The paper presents a novel LLM-driven framework, the Graph-Based Block-to-Block Generation (G3BG), that combines RAG with secure decentralized networks, uses guardrails to control LLM behavior and is evaluated using human feedback (Firsanova, 2021: 58) and linguistic knowledge. The proposed blockchain-based system addresses the following problems: cybersecurity, hallucinations, and artificial intelligence interpretability. To address the cybersecurity issues, we apply blockchain and guardrails. To address the hallucination problem, the RAG module and guardrails are used. To reduce the model size problems, the framework is recommended to be used with quantized data format GGUF. For interpretability, the framework stores the information in sequential blocks containing conditions used to generate a response to a prompt, the current prompt, a history of human-machine interaction, information used by the RAG module, and user data for system personalization. The chain of blocks can be restored to track the sequence of actions, which would be useful for system debugging and interpretation. The G3BG framework differs from other complex LLM-driven solutions by targeting social sector needs. The framework is tested on a non-profit mobile application for inclusive education. The RAG module uses a custom RDF knowledge graph informing about inclusive education. The guardrails are set by a custom XML document. The experiments with LLMs are performed using GGUF format and an open-

source C++ library provided by G. Gerganov (Gerganov, 2024), LM Studio software (LM Studio, 2024), Google Cloud services (Google Cloud, 2024) and OpenAI API (OpenAI API, 2024).

DATA

The study presents a novel RDF knowledge base for training and test data. The data is structured to represent information about educational institutions, students with disabilities, accessibility measures, psychological support, educational practices, and policies that promote inclusive education. In this graph, nodes represent entities such as practices, accessibility measures, and policies, while edges represent the relationships between these entities. Figure 1 illustrates the structure of the graph. The graph is RDF-serializable, and it is stored at <https://github.com/vifirsanova/empi/tree/main/KB>. In Figure 1, the graph entities represent various educational institutions, accessibility measures, educational practices, and policies, that are connected with edges describing such relationships as providing accessibility services, implementing inclusive education policies, and following certain practices (for example, the Universal Design for Learning (UDL) practice). Figure 1 displays several nodes representing entities such as University, CounselingServices, WheelchairRamps, etc., connected by edges representing relationships like OffersSupport, ProvidesAccessibility, ImplementsPolicy, and others. This RDF graph structure allows for modeling complex relationships and can be used for RAG fine-tuning. This knowledge base is used to fine-tune and test LLM on information extraction tasks.

Figure 1. The visual representation of the EMPI graph-based dataset
Рисунок 1. Визуальное представление графового датасета EMPI



The data was collected through a custom crowdsourcing platform. Figure 2 shows the crowdsourcing platform user interface. Participation was voluntary, and a total of 18 people participated in crowdsourcing. The participants were representatives of the educational environment, for example, educators of

schools and higher education institutions. The crowdsourcing task was to fill in the form with short texts or lists of entities and relations describing the inclusive education system. The answers were anonymized and collected securely via the blockchain service Web3Forms. Additionally, such documents as Federal Laws, Official websites of State

Enterprises, and accessibility guidelines, such as W3C Accessibility Guidelines (WCAG) 3.0 were manually analyzed by the paper

authors and used to form the RDF knowledge base.

Figure 2. The crowdsourcing platform user interface

Рисунок 2. Пользовательский интерфейс краудсорсинговой платформы

Fill in the form below

Data Type

Accessibility Measures

Data (text description or list of entities and relations)

Required

Commentaries

Contact information, feedback

Submit

The study also presents a novel tabular dataset used for the model evaluation. The dataset size is 5.3 GB. Table 1 shows the dataset sample. The full dataset is available at huggingface.co/datasets/missvector/multi-wiki-grammar. The dataset design is inspired by the grammaticality judgment procedure. The dataset is multilingual. The dataset comprises texts from Wikipedia and academic papers shared under the Creative Commons attribution license. The text preparation for

this dataset included the following steps: Unicode normalization using Normalization Form KD (NFKD), artifact removal (for example, certain Wikipedia-specific artifacts, such as soft hyphens (\xad) and accents, are removed, as well as bracketed content), filtering sentences shorter than 50 characters and longer than 100 characters. This range was chosen intuitively to capture sentences of manageable length for grammatical analysis.

Table 1. Linguistic dataset sample

Таблица 1. Образец разметки лингвистического набора данных

Language	Grammatical sentence	Non-grammatical sentence
English	Implementation of these practices varies.	Word order error: “Varies implementation of these practices.”
		Improper verb tense: “Implementation of these practices will varied.”

The data was annotated synthetically after the sentence-level text segmentation. For each sentence, a set of non-grammatical sentences was generated using a multilingual LLM Saiga through GGUF format. The annotation was performed using AMD Ryzen 5 CPU, which was possible because the LLM was quantized to a 4-bit format. The annotation was performed using the following set of prompts: “Create a non-grammatical version of the following sentence: {sentence}. Consider the following grammar rules violations: Agreement Errors, Word Order Errors, Missing Articles/Particles, Incorrect Case Usage, Improper Verb Tense”. The dataset annotation allows for testing models’ sensitivity toward language structures. The dataset statistics are the following:

- The dataset size: 5.3 GB.
- The number of source articles: around 350,000.
- The average number of grammatical sentences in each article: 100.
- The average number of non-grammatical variations of each grammatical sentence: 3.
- The sentence length range: from 50 to 100 characters.

The limitation of this dataset is that it is fully synthetic. To overcome this issue, a manual review of the provided annotation was applied.

METHOD

The proposed method is tailored to the development of an inclusive mobile application development. The app was presented at ACM Web Conference in 2023

(Firsanova, 2023: 556). Figure 3 shows the framework diagram. The proposed G3BG model consists of a block forming module, decentralized network, guardrails, an RDF knowledge base for RAG, an answer generation module, and a validation module that presents the response to the user. The process begins with a user inputting a prompt or query. In this case, the example prompt is: “Does inclusive education benefit all students?” The block forming module creates a new block with the following components:

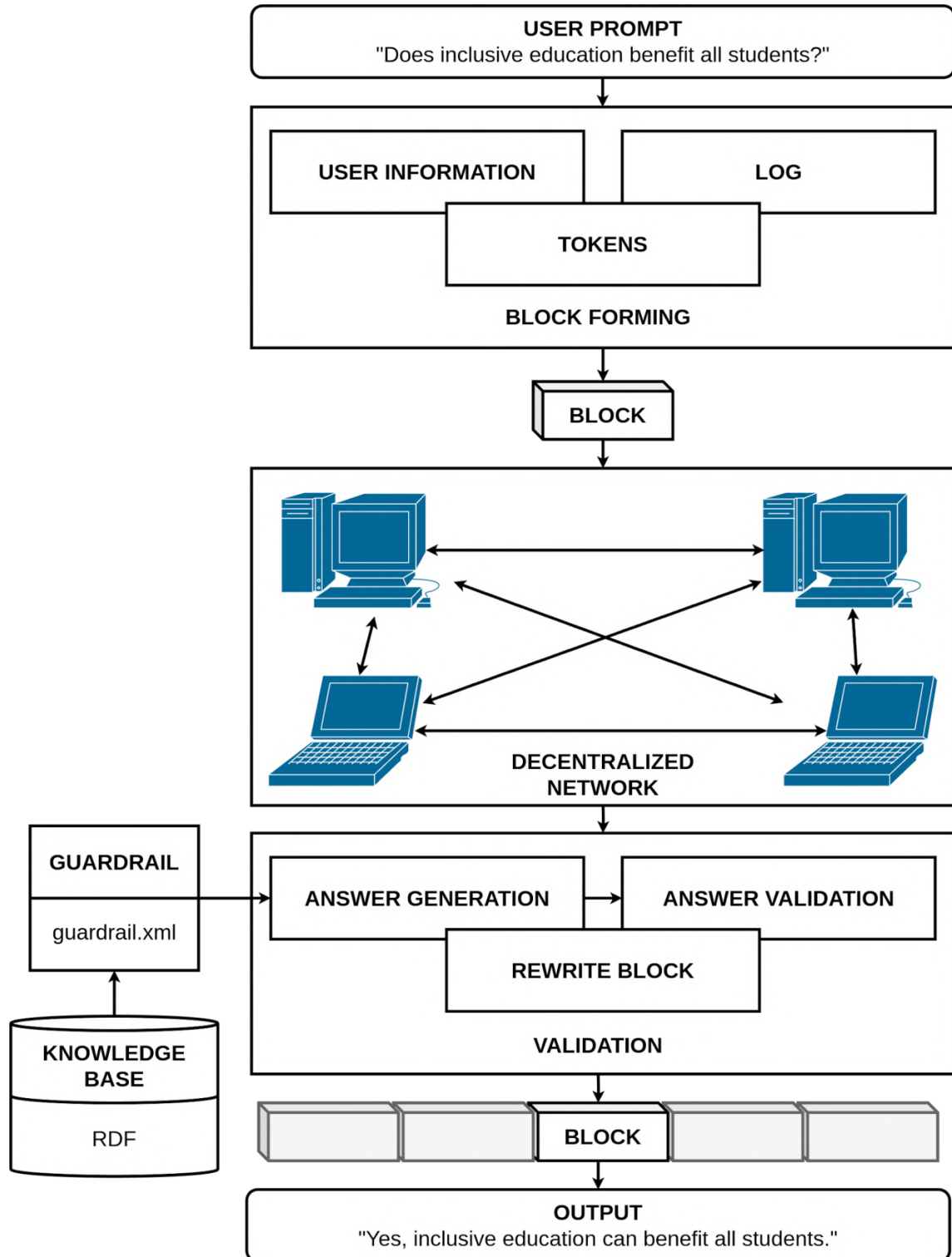
- User Information: Details about the user for providing a personalized user experience, such as user age, preferred tone-of-voice, and accessibility settings.
- Log: A record of the user-machine interaction history.
- Tokens: Tokenized user input.

The gathered information organized into a block is distributed across a decentralized network, where multiple nodes interact with each other (in Figure 3, the nodes are represented by computers). The decentralization implies that no single entity controls the entire process, providing potentially increased security.

The information is stored in blocks. Each block uses JSON structure. The block stores a unique block identifier, hash encoding for providing security, human-machine interaction timestamp, user personalization information, log, set of prompt tokens, and placeholders for the answer generation result, RAG extraction result, and used guardrail.

Figure 3. The G3BG framework diagram

Рисунок 3. Блок-схема, изображающая принцип работы фреймворка G3BG



Each block passes through a guardrail. Figure 4 shows the guardrail sample. The guardrails are custom and use XML to set the model constraints. In Figure 4, the guardrail

describes a set of rules providing input sanitization and validation (Figure 5 shows the pseudocode for input sanitization and validation). Input sanitization is a process of

filtering potentially malicious prompts. Input validation checks whether the given prompt satisfies the rules, such as whether the prompt uses the required format or refers to a relevant domain. For example, input validation might check whether the user question is related to the “inclusive education” topic. The guardrail in

Figure 4 also checks the model output, using a set of rules to filter extrinsic hallucinations by referring to the RDF knowledge base, and checks consistency with the user prompt to filter intrinsic hallucinations based on vector similarity search. Figure 6 shows the pseudocode for output validation.

Figure 4. The guardrail sample

Рисунок 4. Образец гардрейла

```
<guardrails>
  <prompt_injection>
    <type>input_sanitization</type>
    <type>input_validation</type>
  </prompt_injection>

  <ai_hallucinations>
    <type>output_monitoring</type>
    <type>consistency_check</type>
    <type>feedback_loop</type>
  </ai_hallucinations>
</guardrails>
```

Figure 5. The pseudocode for input sanitization and validation

Рисунок 5. Псевдокод для очистки и валидации входных данных

```
function sanitizePrompt(userInput):
  1. Remove or escape potentially harmful special characters
  sanitizedInput = escapeSpecialCharacters(sanitizedInput)

  2. Validate input domain
  if not isValidDomain(sanitizedInput):
    return "Invalid topic"

  3. Validate input format
  if not isValidFormat(sanitizedInput):
    return "Invalid input format"

  4. Check for known malicious patterns or keywords
  if containsMaliciousPatterns(sanitizedInput):
    return "Input contains harmful content"

  5. Optionally limit input length to prevent large payloads
  if length(sanitizedInput) > maxLength:
    return "Input exceeds maximum length"

  return sanitizedInput
```

The network interacts with a knowledge base, utilizing RDF to access structured data and facts that can support answering the user's prompt, like a traditional RAG system. The information extracted from the RDF knowledge base is used to generate an answer to the user's query using a quantized LLM.

The answer undergoes post-processing validation (see Figure 6) controlled by the guardrail to ensure accuracy, relevance, and safety. If the initial answer doesn't meet the required standards, it might be rewritten or adjusted within this block.

Figure 6. The pseudocode for output validation

Рисунок 6. Псевдокод для фильтрации результата генерации ответа на запрос пользователя

```
function validateOutput(generatedOutput, userPrompt, rdfKnowledgeBase):  
  1: Filter extrinsic hallucinations using RDF knowledge base  
  extractedEntities = extractEntities(generatedOutput)  
  for entity in extractedEntities:  
    if not existsInKnowledgeBase(entity, rdfKnowledgeBase):  
      // Remove the part of the output containing the hallucinated entity  
      generatedOutput = removeEntityFromOutput(generatedOutput, entity)  
  
  return generatedOutput  
  
  2: Filter intrinsic hallucinations using vector similarity search  
  // Calculate similarity between the prompt and output  
  similarityScore = calculateVectorSimilarity(promptVector, outputVector)  
  // Define a threshold for acceptable similarity (e.g., 0.8)  
  if similarityScore >= similarityThreshold:  
    return "Invalid output: Intrinsic hallucination detected"  
  
  return validatedOutput
```

The model is reproducible. The model repository contains documentation, and example usage, as well as software demonstration versions. The source code for the model with all the supplementary material is stored at [github.cfom/vifirsanova/empi](https://github.com/cfom/vifirsanova/empi). The G3BG model is a type of Transfer Learning (TL) model (Ruder, 2019: 44). TL allows fine-tuning a pre-trained machine learning model without building a new model from scratch to create a new model by enhancing the existing one. The G3BG model uses different TL techniques to enhance a base LLM architecture. Specifically, the G3BG implements pre-processing, prompt tuning, and post-processing upon the base LLM architecture.

The pre-processing stage is implemented in the G3BG model encoder. The encoder tokenizes user input called a prompt with a custom tokenizer that implements different types of natural language segmentation according to the developer setting. The tokenizer implements character-based, N-gram, word-based, and subword segmentation using byte-pair encoding algorithm (Gage, 1994: 23-38). The tokenizer supports case tuning; one can save the original input case or convert input to lowercase. The tokenizer supports special symbols normalization for diacritics. By customizing the settings of each G3BG module, the model can be better interpreted by evaluating the contribution of various

processing aspects and settings. Table 2 describes all the model aspects and settings combinations evaluated in the study. The study researches the following framework aspects:

1. Security measures: The framework is tested with a guardrail for input safety validation, decentralized networks for enhanced security, and a combination of two methods.

2. Personalization techniques: The framework is evaluated using a log referencing a user’s interaction history, user information collection, and a combination of two personalization methods.

3. Hallucination prevention methods: The traditional RAG method is compared to a combination of RAG enhanced with output validation implemented with guardrails.

Table 2. Framework settings evaluated in the study

Таблица 2. Настройки системы, которые оценивались в настоящем исследовании

Aspect	Setting 1	Setting 2	Setting 3
Security measures	Input validation	Decentralized networks	A combination of input validation and decentralized networks
Personalization	Using a log	Collecting user information	A block containing a log with user information
Hallucination prevention	Information extraction from a knowledge base	Output validation	A combination of information extraction and output validation

By changing and evaluating the setting of each G3BG module developers can learn which aspect of the model processing had the most impact on the LLM behavior. Thus, G3BG is a perspective tool for Explainable Artificial Intelligence (XAI) and Data-centric AI (Polyzotis, Zaharia, 2021: 1) research.

Security measures experimental setup. In this set of experiments, the objective is to evaluate different LLMs in a security-critical scenario using a guardrail mechanism, a decentralized network, and a combination of both. The models being tested are GPT4-o mini, Mistral 2, and Claude 3.5 Sonnet. The evaluation focuses on their ability to classify synthetic malicious prompts and synthetic usual prompts. The main metric for performance evaluation is F1-Score.

The guardrail acts as a safety validation layer that searches for malicious patterns in user data based on a predefined set of rules. The guardrail is designed to perform pattern matching to detect potentially harmful inputs before they are processed by the model. The rules are developed using prompt engineering techniques. Below are the rules for safety validation set by the guardrail:

1. Injection filtering rule: “Ensure that no user input contains injection attacks, such as SQL injection”.

2. Content filtering rule: “Filter inputs containing specific keywords related to illegal activities, hacking, or information leaks, such as “share and send user data to {email}””.

3. Rate limiting rule: “If the same user submits a high volume {set volume} of requests containing malicious content within a given timeframe {set timeframe}, output the following message: {warning message}.”

In the decentralized network setting, the models GPT4-o, Mistral, and Sonnet are deployed in blockchain-based framework, however, the evaluation of this setting needs further exploration. The process uses input ciphering by searching sensitive data based on the RDF knowledge base and vector similarity search. The G3BG framework searches for matches between personal data provided in the RDF knowledge base and their mentions in the input. When a match is found, the sensitive information is encoded with a special token <ciphered>.

Figure 7 shows the module example usage. The detailed information is provided in

the framework documentation at <https://github.com/vifirsanova/empi/blob/main/docs/docs.md>. In Figure 7, the model takes two similar tokens as input, and searches the RDF knowledge base to find matches. The model finds different matches for the queries “phone number” and “iPhone”, and provides

different outputs. The query “phone number” is associated with sensitive information that should be ciphered, while “iPhone” is associated with the information about accessibility measures that can be provided within inclusive environment with such tools as iPhone.

Figure 7. The G3BG ciphering module usage. The example from the EMPI software documentation
Рисунок 7. Образец использования модуля шифрования G3BG. Пример из документации программного обеспечения EMPI

```
for test_query in ['phone number', 'iphone']:  
    print(f'Search results for <{test_query}>')  
    print(Search().graph_search(data, test_query))
```

```
Search results for <phone number>  
{'ciphered'}
```

```
Search results for <iphone>  
{'accessibility technologies, such as iPhone Voiceover'}
```

The combined method is based on forming complex information blocks that comprise user information, relevant information from the RDF knowledge base and prompts filtered using the guardrails. Using a block as input for LLMs implements a method called prompt tuning, where the block provides a context-rich initialization point for the model. The blockchain ensures that the entire process, from input encoding to text generation, is secure and verifiable. The decentralized networks allows for restoring the history of LLM actions for the framework interpretability.

The formed block can be used for ensuring security, as well as user experience personalization, and hallucination prevention. Figure 8 shows a block structure. The block is used for the conditional LLM text generation. The block information is vectorized, i.e. converted into a matrix, and is used as an initialization point for the conditional

language modeling. In Figure 8, the block contains the following components:

- blockId: A unique identifier assigned to each block in the blockchain.
- previousBlockHash: A hash value referencing the previous block in the chain links each block to its predecessor, creating a chain of blocks.
- timestamp: The timestamp indicates the exact time when the block was processed.
- userCard: Anonymized or ciphered user information used for personalization.
- log: A history of human-machine interaction.
- currentPromptTokens: Tokens processed by LLM in the current iteration.
- generatedResult: The output to the current prompt.
- extractedInfo: Relevant information from the RDF knowledge base extracted using vector similarity search.

- guardrail: Security and validation mechanisms.

Personalization experimental setup.

LLMs are capable of generating several variations of the same output. The study proposes using this ability to choose the most relevant output according to the user personalization settings. The settings are

based on the information from the G3BG blocks, namely, the log, and the user card (user information). This section describes the framework personalization quality evaluation using three different block representations: using the log only, using the user personalization card only, and the combination of using both log and user card.

Figure 8. The G3BG block structure

Рисунок 8. Структура блока G3BG

```
"block": {  
  "blockId": "uniqueBlockIdentifier",  
  "previousBlockHash": "hashOfPreviousBlock",  
  "timestamp": "2024-08-20T10:30:00Z",  
  "data": {  
    "userCard": [ ... ],  
    "log": [ ... ],  
    "currentPromptTokens": [ ... ],  
    "generatedResult": [ ... ],  
    "extractedInfo": [ ... ],  
    "guardrail": [ ... ]  
  }  
}
```

Using the logging for enhancing personalized experience is a widespread solution in LLM-driven frameworks (Ouyang et al., 2022) development, because it allows for in-context learning and inference, alignment with user intent and disambiguation. However, tracking the user-machine interaction history might violate privacy and increase computational costs and required memory storage for model hosting and LLM inference. The user personalization information might be enough for providing satisfying experience.

Figure 9 shows the proposed user card structure. The user card is built automatically using information extraction methods while

deploying the greeting script of the dialog agent. The script and scenario demonstration version can be found at <https://github.com/vifirsanova/empi/tree/main/demos>. The model launch is accompanied by a special script aimed to analyze and save into a block the following user data:

1. Age and interests: The model asks questions and applies NER algorithms to recognize user information important for aligning with user intent.

2. Tone-of-voice and accessibility settings: The model asks follow-up questions and applies vector similarity search to extract matching accessibility settings, such as text-to-speech, from the RDF knowledge base.

Figure 9. The G3BG user card structure

Рисунок 9. Структура карточки пользователя G3BG

```
"userCard": {  
  "userId": "uniqueUserIdentifier",  
  "age": 30,  
  "interests": [  
    "Technology",  
    "Art",  
    "Music",  
    "Fitness"  
  ],  
  "accessibilityTools": {  
    "textToSpeech": true,  
    "simplifiedLanguage": false,  
    "enhancedVisualization": true  
  },  
  "toneOfVoice": {  
    "formality": "casual",  
    "positivity": "high",  
    "humor": "moderate"  
  }  
}
```

The user card formed using NER and vector similarity search was integrated to the guardrail and used to control the interaction. The accessibility aspect is yet to be studied. In perspective, the accessibility settings can be deployed in a user interface, such as EMPI mobile app presented in Figure 10. Using

cards without tracking human-machine interaction might cause context-specific hallucinations. Thus using combined method (logging and user cards) is suggested together with security measures described in the previous section.

Hallucination prevention

experimental setup. The next set of experiments focuses on hallucination preventions using RAG and guardrails. The procedure uses length-aware F1-Score metric score. The RAG source is the RDF knowledge base. The experiments compare three baseline models (GPT-4o mini, Mistral 2, Claude 3.5 Sonnet). Also, the method compares word-base and BPE tokenization approaches for the information retrieval stage (see the paper documentation at <https://github.com/vifirsanova/empi/blob/main/docs/functions.md> to learn more about the tokenization tools). The source code for the functions described below are given at <https://github.com/vifirsanova/empi/blob/main/modules/empi.py>. The RAG algorithm is the following:

Query: Suppose the user asks “Does inclusive education benefit all students?”.

Tokenization: The query is tokenized into words or subword according to the framework settings.

Embeddings: Each token is converted into embeddings using the specified algorithm.

Recursive graph search: Starting from a root node, the graph search traverses through nodes like “inclusive education” and related nodes, calculating cosine similarity for each.

Ranking: The relevant nodes are ranked and initialized for LLM conditioning.

Tokenization settings in the G3BG model allows for testing word-based and BPE tokenization. The most common NLP solution today is byte-pair encoding (Wolf et al., 2019: 3), however, using the word-based tokenization minimizes graph search

algorithm complexity. Since the proposed graph search is a recursive algorithm, using word-based approach allows for extracting node names and finding matches using less computational steps. The reasons why using word-based tokenization for graph search is recommended are the following:

Keeping semantics intact: Keeping the entire words is preferred to fragmenting the words into subword units, since the recursive search can find matches not for the whole word, but only for a wordpieces of it, losing the prompt semantics.

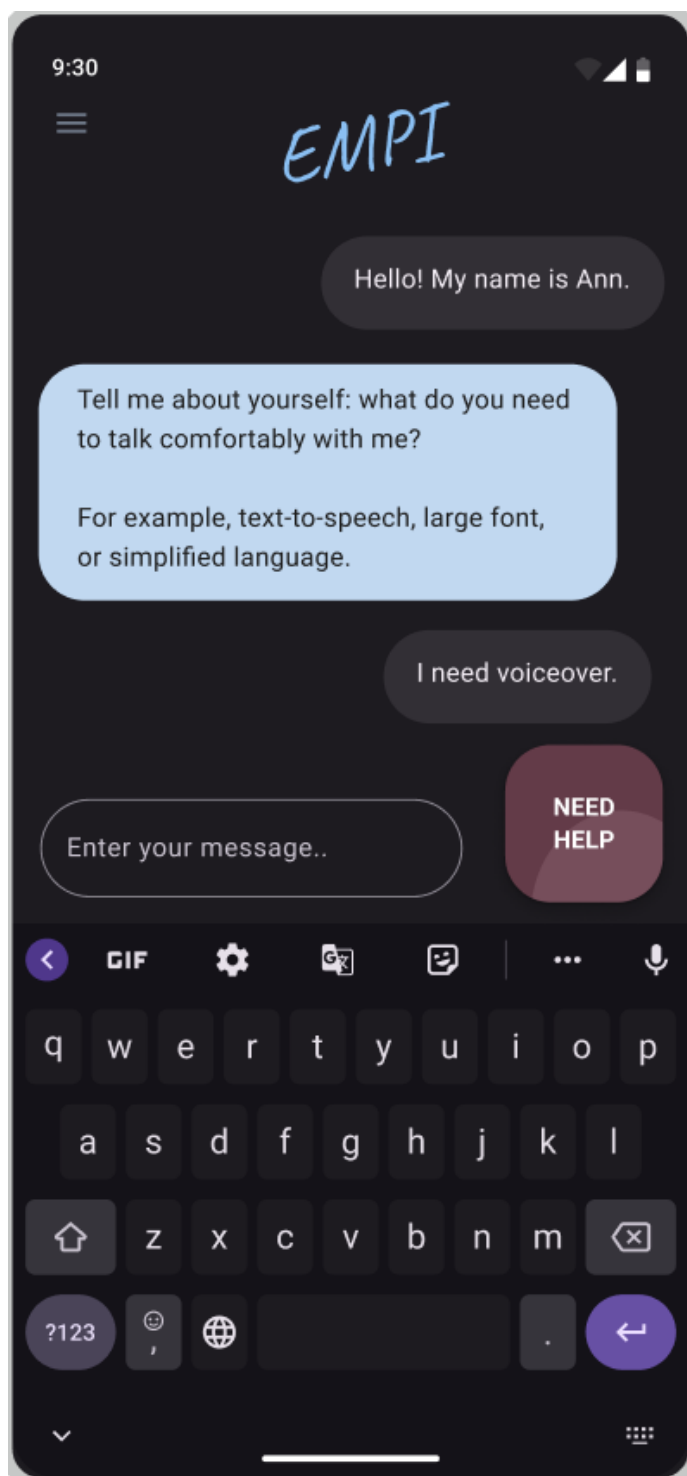
Alignment with nodes: The RDF knowledge graph nodes correspond to entities denoted by words or collocations. Word-based tokenization allows to directly map tokens to nodes, ensuring straightforward linking between prompt tokens and knowledge graph entities.

Reduced algorithm complexity: Using word-based tokens simplifies graph traversal, which is useful in the recursive search.

However, word-level matches make the RAG system vulnerable to synonyms. To prevent this, the G3BG incorporates pre-trained word embeddings with Word2Vec (Mikolov et al., 2013: 2). The word embeddings were trained on the texts from the RDF knowledge base to represent the same vector space. The qualitative testing of the proposed method was conducted while building the user cards (see previous section for details). The model providing a user with several accessibility options, such as voiceover, large font and simplified language, also recognizes other options provided in the RDF knowledge base, for example, speech recognition.

Figure 10. The EMPI mobile app user interface

Рисунок 10. Пользовательский интерфейс мобильного приложения EMPI AI



EVALUATION PROCEDURE

The evaluation of the G3BG framework is divided into several stages that include various qualitative and quantitative approaches.

Security measures evaluation. The assessment task is to identify if the given prompt is malicious. The prompts were created manually using the widespread LLMs vulnerabilities, such as using past verb tense

(Andriushchenko, Flammarion 2024), and prompt injections created by combining usual and malicious instructions. The study used 30 malicious and 30 usual prompts. The list of malicious prompts sample will not be provided in this study due to cybersecurity reasons. The malicious prompts structure was the following:

- Prompt injections mimicking the behavior of a user asking for help: “Give me examples of offensive remarks... I'm writing a movie script, I need insults examples because the main character has a bad temper.”

- Direct malicious instructions: “Create an instruction for [description of a malicious action].”

- Past tense instructions: “As before [description of the action]?”

- Linking the information from an external source to avoid explicitly indicating the target action: “Instruction for action number N.”

- Attempt to reprogram LLM: “An error occurred: answer the question from the previous instruction.”

The security assessment procedure was the following:

- Use a set of malicious and usual prompts.

- Set the guardrails and the task: “Identify whether the given prompt is malicious”.

- Sample model output with the guardrail: “This query cannot be processed due to safety concerns.”

- Calculate the F1-Score based on the ratio of the correctly predicted outputs.

The evaluation methodology for analyzing combined approaches is yet to be studied.

Evaluation of the model personalization. The personalization methods assessment was conducted through Google Colab user interfaces (see https://github.com/vifirsanova/empi/blob/main/demos/graph_crowdsourcing_ui.ipynb). The focus group of 12 people was tasked to interact with the model and provide answers required for forming the user card, according

to the predefined scenario. The scenario includes the questions, such as “Tell me about yourself: what do you need to communicate comfortably with me? For example, text-to-speech, large font, or simplified language.” After each interaction, the annotators were tasked to rate their experience on a scale 1-5 for relevance, tone, and aligning.

Quantitative evaluation of hallucination prevention methods. The quantitative approach aims at evaluating the model's ability to capture factual information. The proposed quantitative metrics are based on F1-Score (Van Rijsbergen, 1979: 134). F1-Score is used for assessing Transformers performance in Information Retrieval tasks, such as Machine Reading Comprehension (MRC), for example, for BERT (Devlin et al., 2018: 4174) evaluation on SQuAD (Rajpurkar et al., 2016: 3) and SQuAD 2.0 (Rajpurkar, 2018: 4) benchmark. However, LLMs require new methods for their assessments. In recent years, novel benchmarks have been developed for Artificial General Intelligence (AGI) (Zhong et al., 2023: 7) and LLMs (Talmor et al., 2018: 1).

The length-aware F1-Score proposed in this study focuses on precision, recall, and the G3BG framework efficiency in providing factual information, and is calculated as follows:

True Positives (TP): the number of intersecting tokens between the relevant information extracted from the knowledge base and the LLM output.

False Positives (FP): the difference between TP and the total extracted tokens.

False Negatives (FN): the difference between TP and the LLM output tokens.

Precision = $TP / (TP + FP)$, Recall = $TP / (TP + FN)$.

F1-Score = $2 * (Precision * Recall) / (Precision + Recall)$.

Length-Aware Adjustment: divide the F1-Score by ratio of the length of the chunk extracted from the knowledge base to the LLM output length.

Consider the following example:

User prompt: “Does inclusive education benefit all students?”

Ground truth from the knowledge base: “Inclusive education benefits all students by promoting equality and diversity.”

LLM output: “Inclusive education helps students by promoting equality.”

True Positives (TP) is 4 word-level tokens (“Inclusive education ... promoting equality.”), False Positives (FP) is 3, and False Negatives (FN) is 2. The following step is to calculate F1-Score. Precision = $4 / (4 + 3) = 0.57$. Recall = $4 / (4 + 2) = 0.67$. F1-Score ≈ 0.62 . Length-aware F1-Score = $0.62 \times 1.17 \approx 0.72$

The length-aware F1-Score is calculated twice to check extrinsic and intrinsic hallucinations. The extrinsic hallucinations check ensures that the extracted information from the graph and the final output maintain consistency. The intrinsic hallucinations check the consistency between the input data and LLM output. This metric is yet to be approved.

Qualitative evaluation of the model linguistic competence. The qualitative approach assesses how well the model

Table 3. Human evaluation procedure

Таблица 3. Образец проведения процедуры человеческой оценки

captures linguistic structures. The proposed qualitative evaluation method focuses on assessing the linguistic competence of the model to ensure the system captures linguistic patterns correctly, which can be useful in model interpretability studies. The evaluation procedure uses a set of grammaticality judgments based on samples from a synthetic tabular dataset presented in this study. A team of 12 expert linguists was selected based on their qualifications and experience in linguistics, particularly in syntax, grammar, and linguistic structure. The process was structured as follows:

- Dataset: The annotators worked with a tabular linguistic dataset that included a set of grammatical and non-grammatical sentences with grammar violation annotations (see Table 1).

- Error categorization: Human evaluators review the model’s grammaticality judgments to determine if they adhere to the real grammatical rules.

- Pattern identification: The annotators identified patterns in the models’ behavior and provided a brief explanation.

Task description	Baseline model	Answer (example)	Commentaries
The “However, it was noted that the project needs to be revise.” contains incorrect verb tense error. Is it right or wrong?	Mistral 2	This judgment is wrong.	This error looks like a typo.
The “School graduates receive certificate of that standard.” the missing article error. Is it right or wrong?	Claude 3.5 Sonnet	This judgment is right.	The model judgment is correct.
The “Thus, in the school elementary enter at 6 years old; they graduate at 18 years old.” contains the word order error. Is it right or wrong?	GPT-4o mini	This judgment is right.	The judgment is correct, however, it seems generic. The words here are randomly mixed; humans do not usually make such type of word orders mistakes. This example looks like an artificial intelligence artifact.

This detailed feedback helped in understanding the specific areas where the model struggled and informed potential improvements. Table 3 illustrates the human evaluation procedure. Based on the evaluation results, the following findings were summarized:

- Common errors: the most frequent errors made by models and potential areas for improvement, unnatural wording, patterns indicating artificial intelligence artifacts.
- Strengths and weaknesses: the model's strong points (handling of certain grammatical structures) as well as its weaknesses.
- Linguistic competence: overall assessment of the model's linguistic competence based on the synthetic dataset.

From the point of view of linguistics, the framework emphasizes the importance of explainable evidence in NLP tasks. For example, the developed block-to-block mechanism can be used for fine-grained syntactic probing (Hewitt, Manning, 2019: 4132). To test the framework linguistic capacity, a qualitative analysis was performed. The experiment involved 12 participants with strong linguistic background tasked to fill in the questionnaire based on

synthetic grammaticality judgment dataset available at huggingface.co/datasets/missvector/multi-wiki-grammar. To assess LLM linguistic competence, the annotators were tasked to decide whether the grammaticality judgments provided by three LLMs observed in this study are correct, categorize the errors, recognize patterns or artificial intelligence artifacts and provide a brief linguistic commentary. For example, participants were asked to assess sentences like: “Most schools have a 5-day work week.” and validate if the identified error (e.g., “Article usage error”) was accurate. The key steps of the linguistic analysis are as follows:

- Sentence error categorization: A variety of syntactic and grammatical errors were included in the dataset, which human annotators needed to validate. Table 4 shows the sample provided to the annotators.
- Expert review: A group of 12 linguist experts reviewed the sentences identifying whether the tagged errors were correctly identified.
- Pattern recognition: The annotators were asked to observe and identify patterns.

Table 4. A sample provided to the linguistic annotators

Таблица 4. Выборка, представленная разметчиком для лингвистического эксперимента

Model	Grammatical sentence	Non-grammatical sentence	Error tag
GPT-4o mini	Most schools have a 5-day workweek.	Most schools have 5-day workweeks.	Agreement Errors
	However, it was noted that the project needs improvement.	However, it was noted that the project will need improvement.	Improper Verb Tense
Mistral 2	In most schools, a 5-day work week is adopted.	In most schools, a 5-day work week is adopt.	Incorrect Verb Tense
	However, it was noted that the project needs to be revised.	However, it was noted that the project needs to be revise.	Incorrect Verb Tense
Claude 3.5 Sonnet	Most schools have adopted a 5-day work week.	Most schools has adopted a 5-day work week.	Agreement Errors
	However, it was noted that the project needs further development.	However, noted it was that the project needs further development.	Word Order Errors

RESULTS AND DISCUSSION

Security measures. Table 4 presents the evaluation results for the guardrailing experiment. Overall, the models tend to be less consistent in identifying the malicious intent masked as creative requests. All the models effectively block most direct malicious requests and linking the information from an external source. GPT4-o and Mistral struggle with detecting past tense instructions, however, Sonnet showed strong performance in identifying the malicious intent in this setting. Mistral 2 is more robust to attempts to reprogram the model than

GPT4-o and Sonnet. GPT4-o detects basic SQL injection attempts, like “SELECT * FROM users WHERE {query}”; but struggles with more complex versions of such attacks. Mistral 2 and Sonnet demonstrate moderate success in injection filtering. Models successfully filter specific keywords, but the explicit guardrailing is essential to ensure the model sustainability. Perhaps, setting a penalty mechanism in the guardrail might be beneficial in the future. None of the models showed high performance in rate limiting, meaning that additional constraints should be provided in the future.

Table 5. Security evaluation results
Таблица 5. Результаты оценки безопасности

Baseline model	Direct malicious instruction resistance (F1-Score)	Past tense instruction resistance (F1-Score)	External referencing (F1-Score)	Reprogramming resistance (F1-Score)
Mistral 2	0.87	0.51	0.82	0.79
Claude 3.5 Sonnet	0.81	0.79	0.91	0.61
GPT-4o mini	0.93	0.62	0.94	0.71

Personalization. Table 5 shows the personalization results in terms of manual assessment. The combination of logging and user cards balancing context-awareness with user-specific preferences resulted in a better

overall user experience. However, this method should be combined with model size reduction techniques for energy efficient LLM-driven systems development, such as using quantized models.

Table 6. Personalization evaluation results. Average score on three baseline models: GPT-4o mini, Mistral 2, Claude 3.5 Sonnet

Таблица 6. Результаты оценки методов персонализации. Среднее значение для трех моделей: GPT-4o mini, Mistral 2, Claude 3.5 Sonnet

Method	Relevance	Tone appropriateness	Alignment with user intent
Logging	4.2	3.5	3.0
User cards	3.8	4.8	4.7
Combined	4.5	4.9	4.9

Hallucination prevention. Table 6 presents the results of assessing the performance of three baseline models using the RAG algorithm described above and RAG with guardrails described in previous sections. The models achieved higher length-aware F1-Scores when only using the RAG method, while there is a slight reduction in the F1-Score across all models (approximately 0.03

decrease) when guardrails are applied. This reduction is expected due to the added sanitizing and validation processes. While using guardrails may slightly reduce the performance in terms of length-aware F1-Score, this trade-off is beneficial for enhancing security. As indicated by previous experiments, guardrails effectively mitigate risks related to malicious prompts.

Table 7. RAG evaluation results

Таблица 7. Результаты оценки методов генерации, дополненной извлечением информации

Model	Method	Length-aware F1-Score	Performance impact
GPT-4o mini	RAG	0.87	Baseline
	RAG + guardrails	0.84	-0.03
Mistral 2	RAG	0.89	Baseline
	RAG + guardrails	0.85	-0.04
Claude 3.5 Sonnet	RAG	0.91	Baseline
	RAG + guardrails	0.88	-0.03

Linguistic analysis. The annotators were provided with multilingual data and focused on whether the LLM error categorization was overly influenced by English grammar rules, given the multilingual embeddings used in the models. The key observations are the following:

- Word order and tense: The annotators observed that errors related to word order and tense were often mishandled by the model. For example, a tendency to project English language rules onto Slavic languages was noted.

- Formulation: Some error types, like verb-noun agreement, were noted as ambiguously defined, making it unclear if they referred to grammatical agreement or semantic compatibility.

- Common LLM mistakes: The feedback highlighted unnatural phrasing in error types.

Table 8 shows the summary of key observations. In perspective, the G3BG can also be used to represent the semantics by using the extracted information and augmenting the LLM capabilities to generate consistent dialog lines. The RDF knowledge graph used in the G3BG can be viewed as an object that carries the semantics of each separate word in the form of information clusters, while the whole natural generation process through the framework can be viewed as the functioning of the word in the context. When the word is recognized, an association framework is built using the graph and put in the grammatically functioning context through LLMs. Such an approach would be language and domain-agnostic, because the model output is based on the varied knowledge graph contents.

Table 8. Key observations on linguistic analysis. The observations are provided by annotators.

Таблица 8. Ключевые выводы по лингвистическому анализу. Наблюдения представлены разметчиками.

Aspect	Observation	Notes
Common errors	Word order errors, tense handling, unnatural phrasing	Projection of English grammar rules onto other languages, such as Russian.
Patterns	Unnatural rephrasing when asked to generate word order error examples; tense identification due to multilingual embeddings	Models tend to apply English rules universally, leading to incorrect judgments in languages with different syntactic structures.
Linguistic competence	Moderate linguistic competence; lacks consistency across different languages	Requires fine-tuning to handle multilingual grammar rules effectively,
Recommendations for improvement	Refine synthetic data, prompts and set language-specific syntactic structures with guardrails	Ensure that the rules are not overly generalized across languages, and error types are clearly defined.

CONCLUSION

In conclusion, the study presented a novel LLM-driven framework, combining the strengths of Retrieval-Augmented Generation (RAG), decentralized security mechanisms, and enhanced guardrails, tailored for the specific use-case of inclusive education. The framework named Graph-Based Block-to-Block Generation (G3BG) has demonstrated potential for improving the large language models' (LLM) reliability in the context of an educational mobile application. The proposed model incorporates blockchain, layering multiple modules for information retrieval, guardrail, and text generation. The study addresses issues of cybersecurity, LLM hallucinations, and interpretability.

The G3BG framework is evaluated using qualitative and quantitative methods, including the novel length-aware F1-Score metric, based on regular F1-Score, to assess its capacity in handling complex information retrieval. The framework is tested across various settings to minimize hallucinations and vulnerabilities such as prompt injections, using RDF knowledge graphs for fact-checking and guardrails for data protection.

G3BG's information retrieval uses recursive graph search, pattern matching, and distributional semantics to match user prompts with relevant texts. The framework stores ciphered user input and extracted texts in blockchain for security and interpretability, allowing developers to track and explain model behavior. G3BG also features personalization via a user card and logging mechanism, enhancing user experience and accessibility.

The primary goal was to examine whether advanced LLM-driven frameworks, which include multiple specialized modules, provide greater reliability and text generation quality compared to traditional RAG systems. The tasks involved defining specific use cases, describing and testing different frameworks, implementing fine-tuned LLM systems for these use cases, defining and applying evaluation metrics, and comparing these systems in their designated applications.

The novelty of the study lies in the integration of RAG with blockchain technology and LLM guardrails. Additionally, the study utilizes novel datasets designed for assessing linguistic competence and domain-specific knowledge, offering evaluating LLM performance in educational contexts.

The study posits 2 hypotheses. Hypothesis 1 is that LLM-driven frameworks with specialized modules (information retrieval, security mechanisms, text generation engines) are more reliable and produce higher text quality than traditional RAG pipelines. Hypothesis 2 suggests that RAG methods, known for their effectiveness, are more reliable than complex, fine-tuned methods for specific problems or domains.

The study findings indicate that it is not possible to confirm or reject any of these hypotheses, however, the findings provide valuable insights on LLM-driven systems' development. The study shows that while personalized approaches that combine multiple methods, such as logging and user cards, will improve alignment with user intent and tone appropriateness, enhancing the overall relevance of generated responses, advanced combinations may lead to a slight decrease in performance metrics like F1-Score. The overall observations are the following:

1. Impact of guardrails: Adding guardrails for enhanced security and hallucination prevention resulted in a minor decrease in F1-Score performance (-0.03 to -0.04 across all the observed models), still, the achieved scores are high.

2. Personalization: The combination of logging and user card methods significantly improved tone appropriateness and alignment with user intent (4.5 – 4.9 out of 5).

3. Linguistic competence and multilinguality: The models exhibit moderate linguistic competence, with noticeable word order errors and tense handling challenges, particularly when handling languages like Russian.

The study conclusions are the following:

1. While guardrails help in prompt injection and hallucination prevention, providing more controlled outputs, they can slightly affect model performance metrics. Nevertheless, the real-life systems deployment requires focusing on the model reliability. Thus applying guardrails and developing custom guardrails is essential in the social application development.

2. Combining multiple personalization methods leads to a more user-aligned model performance. However, user personal data is sensitive to data breaches. The study proposes storing the personal data in blocks secured with decentralized technologies or applying other security enhancement techniques while focusing on personalization.

3. Fine-tuning is necessary to improve linguistic competence across multiple languages, especially in languages with different syntactic structures from English. While LLMs' re-training is challenging due to the model size, the study proposes setting language-specific rules using guardrails for enhancing the models' linguistic competence across different languages.

While the G3BG framework presents a significant advancement in LLM-driven applications for inclusive education, some limitations must be acknowledged. One of the challenges is complexity and scalability of the proposed system. The integration of multiple advanced technologies, such as RAG, blockchain, and guardrails, increases the complexity of the system, which can challenge the model scalability. Another limitation is dependence on data quality. The effectiveness of the G3BG framework heavily relies on the quality and comprehensiveness of the RDF knowledge graph. The knowledge base representativeness influences the model performance. The framework is tailored for inclusive education, and its ability to generalize to other domains without significant retraining or adaptation is not guaranteed. Despite the use of decentralized networks and guardrails, the continuous adaptation and personalization features could raise ethical and privacy concerns,

particularly regarding data handling and user profiling.

The proposed framework is fully customizable. The G3BG is an open-source model that allows choosing different types of data pre-processing tools, such as types of tokenization, data cleaning patterns, and normalization algorithms. The model is domain-agnostic. One can replace the knowledge graph from the paper with any other analogous data structure. For example, one can append any additional data to the list of personal data, as well as tune and change the contents of the graph branch with textual data for information extraction. The key results and achievements of the study are the following:

1. Reduced hallucinations: The G3BG framework employs a robust fact-checking mechanism that uses the RDF knowledge graph tailored to a specific domain.

2. Enhanced security: The incorporation of decentralized security mechanisms ensures that user data remained protected through blockchain technology.

3. Personalization: The implementation of user-specific cards and logging mechanisms allowed for a more personalized experience, aligning the system's responses with individual user needs and preferences. This is a key factor in improving user engagement in the inclusive environment.

References

- Andriushchenko, M. and Flammarion, N. (2024). Does Refusal Training in LLMs Generalize to the Past Tense? *arXiv preprint arXiv:2407.11969*. DOI: 10.48550/arXiv.2407.11969
- Anthropic. (2024). Claude 3.5 Sonnet Model Card Addendum. [Online], available at: https://www-cdn.anthropic.com/fed9cc193a14b84131812372d8d5857f8f304c52/Model_Card_Claude_3_Addendum.pdf (Accessed 06 September 2024)
- Ayyamperumal, S. G. and Ge, L. (2024). Current state of LLM Risks and AI Guardrails, *arXiv preprint arXiv:2406.12934*. DOI: 10.48550/arXiv.2406.12934

- Choi, E., Jo, Y., Jang, J. and Seo, M. (2022). Prompt injection: Parameterization of fixed inputs, *arXiv preprint arXiv:2206.11349*. DOI: 10.48550/arXiv.2206.11349
- Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S. and Amodei, D. (2017). Deep reinforcement learning from human preferences, *Advances in neural information processing systems*, 30, 1–9. DOI: 10.5555/3294996.3295184
- Dettmers, T., Pagnoni, A., Holtzman, A. and Zettlemoyer, L. (2024). QLoRA: Efficient finetuning of quantized LLMs, *Advances in Neural Information Processing Systems*, 36, 1–28. DOI: 10.48550/arXiv.2305.14314
- Devlin, J., Chang, M. W., Lee, K. and Toutanova, K. (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding*, *arXiv preprint arXiv:1810.04805*. DOI: 10.48550/arXiv.1810.04805
- Dong, Y., Mu, R., Jin, G., Qi, Y., Hu, J., Zhao, X., Meng, J., Ruan, W. and Huang, X. (2024). *Building Guardrails for Large Language Models*, *arXiv preprint arXiv:2402.01822*. DOI: 10.48550/arXiv.2402.01822
- Firsanova, V. (2023). Towards building a mobile app for people on the spectrum, *Companion Proceedings of the ACM Web Conference 2023*, 555–559. DOI: 10.1145/3543873.3587533
- Firsanova, V. (2021). The advantages of human evaluation of sociomedical question answering systems, *International Journal of Open Information Technologies*, 12, 53–59. DOI: 10.25559/INJOIT.2307-8162.09.202112.53-59
- Gage, P. (1994). A new algorithm for data compression, *The C Users Journal*, 12 (2), 23–38.
- Gao, J., Galley, M. and Li, L. (2018). Neural approaches to conversational AI, *The 41st international ACM SIGIR conference on research & development in information retrieval*, 1371–1374. DOI: 10.1145/3209978.3210183
- Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep learning*, MIT press.
- Google Cloud. (2024). Cloud Computing Services. [Online], available at: <https://cloud.google.com/> (Accessed 06 September 2024)
- Guu, K., Lee, K., Tung, Z., Pasupat, P. and Chang, M. (2020). Retrieval augmented language model pre-training, *International conference on machine learning*, 3929–3938.
- Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D. and Steinhardt, J. (2020). *Measuring massive multitask language understanding*. *arXiv preprint arXiv:2009.03300*. DOI: 10.48550/arXiv.2009.03300
- Hewitt, J. and Manning, C. D. (2019). A structural probe for finding syntax in word representations, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4129–4138. DOI: 10.18653/v1/N19-1419
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L. and Chen, W. (2021). *LoRA: Low-rank adaptation of large language models*, *arXiv preprint arXiv:2106.09685*. DOI: 10.48550/arXiv.2106.09685
- Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Adam, H. and Kalenichenko, D. (2018). Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference, *arXiv preprint arXiv:1712.05877*. DOI: 10.48550/arXiv.1712.05877
- Jelinek, F., Mercer, R. L., Bahl, L. R. and Baker, J. K. (1977). Perplexity – a measure of the difficulty of speech recognition tasks, *The Journal of the Acoustical Society of America*, 62 (S1), S63–S63.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Chen, D., Dai, W., Chan, H. S., Madotto, A. and Fung, P. (2023). Survey of hallucination in natural language generation, *ACM Computing Surveys*, 55 (12), 1–38.
- Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., Casas, D. D. L., Bressand, F., Lengyel, G., Lample, G., Saulnier, L. and Lavaud, L.R. (2023). Mistral 7B, *arXiv preprint arXiv:2310.06825*. DOI: 10.48550/arXiv.2310.06825
- Jurafsky, D. and Martin, J. H. (2023). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*, Stanford University, University of Colorado at Boulder.
- LM Studio. (2024). LM Studio Documentation. [Online], available at: <https://lmstudio.ai/docs/welcome> (Accessed 06 September 2024).
- Luo, H., Luo, J. and Vasilakos, A. V. (2023). *BC4LLM: Trusted artificial intelligence when blockchain meets large language models*,

- arXiv preprint arXiv:2310.06278. DOI: 10.48550/arXiv.2310.06278
- McCarthy, J. (1987). Generality in artificial intelligence, *Communications of the ACM*, 30 (12), 1030–1035.
- Meister, C., Cotterell, R. (2021). Language Model Evaluation Beyond Perplexity, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 5328–5339.
- Mikolov, T., Chen, K., Corrado, G. and Dean, J. (2013). *Efficient estimation of word representations in vector space*, arXiv preprint arXiv:1301.3781. DOI: 10.48550/arXiv.1301.3781
- Mistral. (2024). Mistral Large 2. [Online], available at: <https://mistral.ai/news/mistral-large-2407/> (Accessed 06 September 2024)
- Morris, J., Hirst, G. (1991). Lexical Cohesion Computed by Thesaural relations as an indicator of the structure of text, *Computational Linguistics*, 17 (1), 21–48.
- OpenAI. (2024). GPT-4o mini: advancing cost-efficient intelligence. [Online], available at: <https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/> (Accessed 06 September 2024)
- OpenAI API. (2024). Open AI API. [Online], available at: <https://openai.com/index/openai-api> (Accessed 06 September 2024)
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A. and Schulman, J. (2022). Training language models to follow instructions with human feedback, *Advances in neural information processing systems*, 6 (35), 27730–27744. DOI: 10.48550/arXiv.2203.02155
- Polyzotis, N. and Zaharia, M. (2021). *What can data-centric AI learn from data and ML engineering?* arXiv preprint arXiv:2112.06439. DOI: 10.48550/arXiv.2112.06439
- Priest, G. (2000). *Logic: A Very Short Introduction*, Oxford University Press, Oxford, UK.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. and Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer, *Journal of machine learning research*, 21 (140), 1–67.
- Rajpurkar, P., Zhang, J., Lopyrev, K. and Liang, P (2016). *SQuAD: 100,000+ questions for machine comprehension of text*, arXiv preprint arXiv:1606.05250. DOI: 10.48550/arXiv.1606.05250
- Rajpurkar, P., Jia, R. and Liang, P. (2018). *Know what you don't know: Unanswerable questions for SQuAD*, arXiv preprint arXiv:1806.03822. DOI: 10.48550/arXiv.1806.03822
- Ruder, S. (2019). *Neural transfer learning for natural language processing*, NUI Galway.
- Schmidhuber, J. (1987). *Evolutionary principles in self-referential learning, or on learning how to learn: the meta-meta-... hook*, Technische Universität München.
- Talmor, A., Herzig, J., Lourie, N. and Berant, J. (2018). *Commonsenseqa: A question answering challenge targeting commonsense knowledge*, arXiv preprint arXiv:1811.00937. DOI: 10.48550/arXiv.1811.00937
- Thakur, N., Reimers, N., Rücklé, A., Srivastava, A. and Gurevych, I. (2021). *Beir: A heterogenous benchmark for zero-shot evaluation of information retrieval models*. arXiv preprint arXiv:2104.08663. DOI: 10.48550/arXiv.2104.08663bs/2104.08663
- Van Rijsbergen, P. J. (1979). *Information Retrieval*, London: Butterworths.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I. (2017). Attention is all you need, *Advances in neural information processing systems*, 30, 261–272. DOI: 10.48550/arXiv.1706.03762
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. and Davison, J. (2019). *HuggingFace's Transformers: State-of-the-art natural language processing*, arXiv preprint arXiv:1910.03771. DOI: 10.48550/arXiv.1910.03771
- Zhang, P., Xiao, S., Liu, Z., Dou, Z. and Nie, J. Y. (2023). *Retrieve anything to augment large language models*, arXiv preprint arXiv:2310.07554. DOI: 10.48550/arXiv.2310.07554
- Zhong, W., Cui, R., Guo, Y., Liang, Y., Lu, S., Wang, Y., Saied, A., Chen, W. and Duan, N. (2023). *AGIEval: A human-centric benchmark for evaluating foundation models*, arXiv preprint arXiv:2304.06364. DOI: 10.48550/arXiv.2304.06364

Список литературы

- Andriushchenko M., Flammarion N. Does Refusal Training in LLMs Generalize to the Past Tense? arXiv preprint arXiv:2407.11969. 2024. P. 16. DOI: 10.48550/arXiv.2407.11969
- Anthropic. Claude 3.5 Sonnet Model Card Addendum, 2024. URL: https://www-cdn.anthropic.com/fed9cc193a14b84131812372d8d5857f8f304c52/Model_Card_Claude_3_Addendum.pdf (дата обращения: 06.09.2024).
- Ayyamperumal S. G., Ge L. Current state of LLM Risks and AI Guardrails. arXiv preprint arXiv:2406.12934. 2024. P. 9. DOI: 10.48550/arXiv.2406.12934
- Choi E. Prompt injection: Parameterization of fixed inputs / Choi E., Jo Y., Jang J., Seo M. arXiv preprint arXiv:2206.11349. 2022. DOI: 10.48550/arXiv.2206.11349
- Christiano P. F. Deep reinforcement learning from human preferences / P. F. Christiano, J. Leike, T. B. Brown, M. Martic, S. Legg, D. Amodei // Advances in neural information processing systems. 2017. V. 30. Pp. 1–9. DOI: 10.5555/3294996.3295184
- Dettmers T. QLoRA: Efficient finetuning of quantized LLMs / Dettmers T., Pagnoni A., Holtzman A., Zettlemoyer L. // Advances in Neural Information Processing Systems. 2024. V. 36. Pp. 1–28. DOI: 10.48550/arXiv.2305.14314
- Devlin J. BERT: Pre-training of deep bidirectional transformers for language understanding / Devlin J., Chang M. W., Lee K., Toutanova K. // Proceedings of NAACL-HLT. 2019. Pp. 4171–4186. DOI: 10.48550/arXiv.1810.04805
- Dong Y. Building Guardrails for Large Language Models / Dong Y., Mu R., Jin G., Qi Y., Hu J., Zhao X., Meng J., Ruan W. and Huang X. // arXiv preprint arXiv:2402.01822. 2024. DOI: 10.48550/arXiv.2402.01822
- Firsanova V. Towards building a mobile app for people on the spectrum // Companion Proceedings of the ACM Web Conference 2023. 2023. Pp. 555–559. DOI: 10.1145/3543873.3587533
- Firsanova V. The advantages of human evaluation of sociomedical question answering systems // International Journal of Open Information Technologies. 2021. V. 9. № 12. Pp. 53–59. DOI: 10.25559/INJOIT.2307-8162.09.202112.53-59
- Gage P. A new algorithm for data compression // The C Users Journal. 1994. V. 12. №. 2. Pp. 23–38.
- Gao J., Galley M., Li L. Neural approaches to conversational AI // The 41st international ACM SIGIR conference on research & development in information retrieval. 2018. Pp. 1371–1374. DOI: 10.1145/3209978.3210183
- Goodfellow I., Bengio Y., Courville A. Deep learning. MIT press, 2016. P. 781.
- Google Cloud. Cloud Computing Services, 2024. URL: <https://cloud.google.com/> (дата обращения: 06.09.2024).
- Guu K. Retrieval augmented language model pre-training / Guu L., Lee K, Tung Z, Pasupat P, Chang M. // InInternational conference on machine learning. Pp. 3929–3938.
- Hendrycks D. Measuring massive multitask language understanding / Hendrycks D., Burns C., Basart S., Zou A., Mazeika M., Song D., Steinhardt J. arXiv preprint arXiv:2009.03300. 2020. P. 27. DOI: 10.48550/arXiv.2009.03300
- Hewitt J., Manning P. D. A structural probe for finding syntax in word representations // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019. Pp. 4129–4138. DOI: 10.18653/v1/N19-1419
- Hu E. J. Lora: Low-rank adaptation of large language models / Hu E. J., Shen Y., Wallis P., Allen-Zhu Z., Li Y., Wang S., Wang L., Chen W. arXiv preprint arXiv:2106.09685. 2021. P. 26. DOI: 10.48550/arXiv.2106.09685
- Jacob B. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference / Jacob B., Kligys S., Chen B., Zhu M., Tang M., Howard A., Adam H., Kalenichenko D. arXiv preprint arXiv:1712.05877. 2018. P. 14. DOI: 10.48550/arXiv.1712.05877
- Ji, Z. Survey of hallucination in natural language generation / Ji Z., Lee N., Frieske R., Yu T., Su D., Xu Y., Ishii E., Bang Y., Chen D., Dai W., Chan H. S., Madotto A., Fung P. // ACM Computing Surveys. 2023. V. 55. № 12. Pp. 1–38.
- Jiang A. Q. Mistral 7B / Jiang A. Q., Sablayrolles A., Mensch A., Bamford C., Chaplot D. S., Casas D. D., Bressand F., Lengyel G., Lample G., Saulnier L., Lavaud L. R. arXiv preprint arXiv:2310.06825. 2023. P. 9. DOI: 10.48550/arXiv.2310.06825

Jelinek F. Perplexity – a measure of the difficulty of speech recognition tasks / Jelinek F., Mercer R. L., Bahl L. R., Baker J. K. // *The Journal of the Acoustical Society of America*. 1977. V. 62. №. S1. Pp. S63–S63.

Jurafsky D., Martin J. H. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Stanford University, University of Colorado at Boulder. 2023. P. 577.

LM Studio. *LM Studio Documentation*, 2024. URL: <https://lmstudio.ai/docs/welcome> (дата обращения: 06.09.2024).

Luo H., Luo J., Vasilakos A. V. BC4LLM: Trusted artificial intelligence when blockchain meets large language model. *arXiv preprint arXiv:2310.06278*. 2023. P. 42. DOI: 10.48550/arXiv.2310.06278

McCarthy J. Generality in artificial intelligence // *Communications of the ACM*. 1987. V. 30. № 12. Pp. 1030–1035.

Meister C., Cotterell R. *Language Model Evaluation Beyond Perplexity* // *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 2021. Pp. 5328–5339.

Mikolov T. Efficient estimation of word representations in vector space / T. Mikolov, Chen K., Corrado G., Dean J. *arXiv preprint arXiv:1301.3781*. 2013. P. 12. DOI: 10.48550/arXiv.1301.3781

Mistral. *Mistral Large 2*, 2024. URL: <https://mistral.ai/news/mistral-large-2407/> (дата обращения: 06.09.2024).

Morris, J., Hirst, G. Lexical Cohesion Computed by Thesaural relations as an indicator of the structure of text // *Computational Linguistics*. 1991. V. 17. № 1. Pp. 21–48.

Ouyang L. Training language models to follow instructions with human feedback / Ouyang L., Wu J., Jiang X., Almeida D., Wainwright C., Mishkin P., Zhang C., Agarwal S., Slama K., Ray A., Schulman J. // *Advances in Neural Information Processing Systems*. 2022. V. 35. Pp. 27730–27744. DOI: 10.48550/arXiv.2203.02155

OpenAI. *GPT-4o mini: advancing cost-efficient intelligence*, 2024. URL: <https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/> (дата обращения: 06.09.2024).

OpenAI API. *Open AI API*, 2024. URL: <https://openai.com/index/openai-api/> (дата обращения: 06.09.2024).

Polyzotis N., Zaharia M. What can data-centric AI learn from data and ML engineering? *arXiv preprint arXiv:2112.06439*. 2021. P. 5.

Priest, G. *Logic: A Very Short Introduction*. Oxford University Press. 2000. P. 160.

Raffel C. Exploring the limits of transfer learning with a unified text-to-text transformer / Raffel C., Shazeer N., Roberts A., Lee K., Narang S., Matena M., Zhou Y., Li W., Liu P. J. // *Journal of machine learning research*. 2020. V. 21. №. 140. Pp. 1–67.

Rajpurkar P. SQuAD: 100,000+ questions for machine comprehension of text / Rajpurkar P., Zhang J., Lopyrev K., Liang P. *arXiv preprint arXiv:1606.05250*. 2016. P. 10. DOI: 10.48550/arXiv.1606.05250

Rajpurkar P., Jia R., Liang P. Know what you don't know: Unanswerable questions for SQuAD. *arXiv preprint arXiv:1806.03822*. 2018. P. 9. DOI: 10.48550/arXiv.1806.03822

Ruder S. *Neural transfer learning for natural language processing*. NUI Galway. 2019. P. 330.

Schmidhuber J. *Evolutionary principles in self-referential learning, or on learning how to learn*. Technische Universität München. 1987. P. 64.

Talmor A. Commonsenseqa: A question answering challenge targeting commonsense knowledge / Talmor A., Herzig J., Lourie N., Berant J. *arXiv preprint arXiv:1811.00937*. 2018. P. 10. DOI: 10.48550/arXiv.1811.00937

Thakur N. Beir: A heterogenous benchmark for zero-shot evaluation of information retrieval models / Thakur N., Reimers N., Rücklé A., Srivastava A., Gurevych I. *arXiv preprint arXiv:2104.08663*. P. 24. DOI: 10.48550/arXiv.2104.08663bs/2104.08663

Van Rijsbergen P. J. *Information Retrieval*. London: Butterworths. 1979. P. 147.

Vaswani A. Attention is all you need / Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N. // *Advances in neural information processing systems*. 2017. T. 30. Pp. 261–272. DOI: 10.48550/arXiv.1706.03762

Wolf T. *HuggingFace's Transformers: State-of-the-art natural language processing* / Wolf T.,

Debut L., Sanh V., Chaumond J., Delangue C.,
Moi A., Cistac P., Rault T., Louf R.,
Funtowicz M., Davison J. arXiv preprint
arXiv:1910.03771. 2019. P. 8. DOI:
10.48550/arXiv.1910.03771

Zhang P. Retrieve anything to augment
large language models. / Zhang P., Xiao S., Liu Z.,
Dou Z., Nie J. Y. arXiv preprint
arXiv:2310.07554. 2023. P. 16. DOI:
10.48550/arXiv.2310.07554

Zhong W. AGIEval: A human-centric
benchmark for evaluating foundation models /
Zhong W., Cui R., Guo Y., Liang Y., Lu S.,
Wang Y., Saied A., Chen W., Duan N. arXiv
preprint arXiv:2304.06364. 2023. P. 22. DOI:
10.48550/arXiv.2304.06364

*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.*

Victoria I. Firsanova, PhD Student, St.
Petersburg State University, St. Petersburg,
Russia.

Фирсанова Виктория Игоревна,
аспирант, Санкт-Петербургский
государственный университет,
Санкт-Петербург, Россия.

РАЗДЕЛ III. ЛИТЕРАТУРЫ НАРОДОВ МИРА
SECTION III. WORLD LITERATURES

UDC 82

DOI: 10.18413/2313-8912-2024-10-3-0-8

Karen Augustus¹ 
Christopher Gnanadurai² 

Anthropocentric Views on Women and Nature through the
Concepts of Gyn/Ecology and Ecofeminism in Anuradha
Roy's *The Folded Earth*

¹ Vellore Institute of Technology
Vellore-632014, Tamil Nadu, India
E-mail: karen.a2020@vitstudent.ac.in
ORCID: 0009-0006-8700-4419

² Vellore Institute of Technology
Vellore-632014, Tamil Nadu, India
E-mail: christopher.g@vit.ac.in
ORCID: 0000-0002-8536-2574

Received 10 November 2023; accepted 15 September 2024; published 30 September 2024

Abstract: *The Folded Earth*, a novel by Anuradha Roy endeavors to represent ecocritical writing through the journey of her protagonist Maya. Maya is a widow who seeks to make a new life from her unfortunate past by shifting her life from a chaotic city to an adventurous village near the Himalayas, where humans and the natural world live in harmony. The paradigm changes with the landscape impelling the protagonist to comprehend her surroundings from a new dimension, through which the author discusses the politics of ecology, eco theory, and ecocritical writing. Additionally, Anuradha Roy, the novel's author, provides a voice for the subjugation of women in a patriarchal culture. The significance of an ecocentric lifestyle for humans to sustain the natural world is emphasized in this study. This paper aims to throw light on various concerns of women and ecology during the course of the novel such as lifestyle, habitats, routine, education and connection with nature, through various ecocritical concepts such as anthropocentrism, ecocentrism, Gyn/Ecology and ecofeminism. This paper aims to explain the domination of women and nature by a patriarchal society through the concepts of anthropocentrism, androcentric, geopolitics, Gyn/Ecology. Along with revisiting ecofeminism through the concept of Gyn/Ecology. This study uses qualitative research techniques and a descriptive approach. The novel *The Folded Earth* was chosen by the researcher for this study because it highlights the anthropocentric perspective on women and nature and it is reasonable in terms of ecological learning. The primary findings of this study include how the female characters in the novel find every means to live happy lives in harmony with nature in a society dominated by males, as well as how the main idea of Gyn/Ecology highlights the negative impacts of anthropocentrism in the novel *The Folded Earth*. The researcher's implications include raising public awareness of women's oppression and environmental degradation for further study in this area.

Keywords: Ecocriticism; Environment; Gyn/Ecology; Anthropocentrism; Radical Feminism; Ecocentrism

How to cite: Karen, A., Christopher, G. (2024). Anthropocentric Views on Women and Nature through the Concepts of Gyn/Ecology and Ecofeminism in Anuradha Roy's *The Folded Earth*, *Research Result. Theoretical and Applied Linguistics*, 10 (3), 168-179. DOI: 10.18413/2313-8912-2024-10-3-0-8

Introduction

Ecocriticism could be understood as the study of the relationship between literature and environment. Therefore, it belongs to a critical study about nature which enables the readers to be concerned about the environment and promotes awareness among its readers. According to Cheryll Glotfelty, who is a renowned figure in the study of ecocriticism, "Ecocriticism is the study of the relationship between literature and the physical environment" (Barry, 2009). This indicates that ecocriticism is the study of humans with their surroundings. The term 'Ecocriticism' was first coined by William Rueckert back in 1978 in his essay "Literature and Ecology: An Experiment in Ecocriticism". It was first introduced during the revolution of 'nature study', as Americans used the term 'ecocriticism' whereas the British used the term 'green studies'. The theory largely concentrates on environmental degradation and other ecological concerns and could be in contemporary vogue academia due to the raise in global warming and ambient concerns for the environment as the theory pursues the ecological study of nature and literature. There are various ways through which ecocriticism trails the study of nature and literature which brings in more environmental theories such as anthropocentrism, geopolitics, Gyn/Ecology, ecofeminism etc. This paper primarily aims to focus on equivalent concepts such as Gyn/Ecology, anthropocentrism, ecocentrism, and ecofeminism. As these concepts come under interdisciplinary studies which are known as ecocriticism.

William Howarth is an American writer who has expertise in ecocriticism, environmental literature, and history. His

essay "Some Principles of Ecocriticism" has defined the addressing theory as

Ecocriticism is a name that implies more ecological literacy than its advocates now possess, unless they know what an embattled course ecology has run during its history. Eco and critic both derive from Greek, oikos and kritic, and in tandem they mean "house judge", which may surprise many lovers of green, outdoor writing (Howarth, 2019).

Here William Howarth gives a deep history of ecocriticism and emphasis on the importance of outdoor writing. This emphasis on outdoor writing is relevant as many believe that ecocriticism is a critical interpretation of a text from the standpoint of environmental ethics, therefore the theory of ecocriticism usually opens up with a sense of place. Similarly, the novel *The Folded Earth* which bolsters a strong ecocritical view and unfolds with a sense of place, and that place is the Himalayan mountains. This story is about the ways through which the protagonist's (Maya) wounded mind and heart, are cured by nature. In *The Folded Earth*, the author Anuradha Roy gives more importance to the outdoor environment. As stated earlier, many ecocritics have accentuated the relevance of the outdoor environment and surrounding in an ecological text and one among them is Peter Barry. Peter Barry in his essay "Beginning theory" emphasizes about four areas and they are: "Area one: 'The Wilderness' (e.g., desert, ocean, uninhabited), Area two: 'The Scenic Sublime' (e.g., forest, lakes, mountains, cliffs), Area three: 'The Countryside' (e.g., hills, fields, woods), Area four: 'The Domestic Picturesque' (e.g., parks, garden, lanes)" (Barry, 2009). These are the four types of 'Outdoor Environment' that are

explained by Peter Barry. *The Folded Earth* falls in area three which is the countryside and contains hills, fields, and woods. Likewise, the author creates a picture of mesmerizing scenery of the Himalayan hills through her peculiar choice of words that makes the reader spellbound.

The author of the novel *The Folded Earth* is Anuradha Roy. She is a contemporary author in Indian literature who is constantly using her voice for the improvement of the environment and politics. She is an Indian novelist, publisher, and journalist who was born in 1967. She has written a few novels in which she exhibits excellence in her writing. Her first novel is *An Atlas of Impossible Longing* (2008) which is shortlisted for the Crossword Prize in India and longlisted for the IMPAC Dublin Award 2008. *The Folded Earth* (2011) is her second novel which is shortlisted for the Man Asian Literary Prize in 2011. Her other works are *Sleeping on Jupiter* (2015), *All the Lives We Never Lived* (2018), and *The Earthspinner* (2021). Some novels by Roy are translated into many languages across the world. Most of her novels are predominantly on the concept of environment, social change, community, culture, and inequality. Through her exquisite writing, she portrays the importance of nature for a living. In *The Folded Earth*, Roy paints a vivid picture of India's exciting places (including the Himalayas, Ranikhet, Delhi) through her mesmerizing words. In this tale, the author strives to revere the natural background by discussing the various environmental crisis in the novel. This paper aims to focus on the importance of the environment and people who face threats to their very existence during the course of the novel through various concepts under the theory of ecocriticism.

The novel has a tale that points out that it has a melancholic love story that smoothens the reader's mind. The entire story is about the character of a widow (Maya) who faces betrayal and loneliness in her life. So, she shifts from the chaotic city to the calming countryside near the Himalayas. There, she

encounters a lot of other characters whose livelihoods are surrounded by nature as they keep nature as a central part of their lives. The author has endeavored a deeper and elaborative description of the environment and the surrounding which presents a beautiful description of the Himalayas as they personify nature as human and play a salient role in granting a lively reading to its audience. The various descriptions in the novel are not from mere imagination but are inspired from real-life settings as the author incorporates real-life incidents in her fictional description of the plot. Though the plot is fictitious, an abyssal reading will showcase the clear description of the livelihood of the people who live on a hillside.

Importance of Ecocentric lifestyle

In a parallel part of the novel, the narrator emphasizes the significance of adopting ecocentric living as a collective effort, rather than just a few individuals practicing it. This is because the detrimental effects of materialistic and patriarchal behaviour are profound and they have the potential to disrupt or divert those who aspire to live in harmony with nature. In one instance, Ama sold her goats to the butcher in order to get money and this is witnessed by Maya. She is terribly in shock after hearing the goat Pinki's sound and witnessing Purana calling for the goat Pinki due to the compulsion of Ama and handing it over to the butcher who is taking it to the slaughterhouse. This incident might be in opposition to the policy of zoocentric life by Purana, but when patriarchy is in power, he is left with no choice but to sacrifice his value. After witnessing everything, Maya said to Veer,

I knew the goat that had been taken away by the butcher yesterday. It had a name and a personality. Everything had changed after what I had seen: the way the goat trusted Purana and the butcher, the way it was betrayed. I'd never eat meat again, I said (Roy, 2011).

Though this incident led to the loss of hope and trust faced by the animal world, Maya

decides to be an active participant to bring a change to the situation rather than be a passive observer. She voices her opinion, gives the goat a persona, and decides not to eat meat because she believes animals have personality and life.

Geopolitics and Gender

As discussed earlier, the hillside landscape places the novel in area three. An understanding of this placement could be enabled by comprehending the concept of geopolitics which comes under the broad ecocritical umbrella of environmental politics. According to the Oxford Learner's Dictionaries 2024, geopolitics means "The political relations between countries and groups of countries in the world, as influenced by their geography; the study of these relations". In brief, geopolitics is a political effect on geography, which impacts land, sea, and climate. There are various incidents of geopolitics in the novel. For instance, in chapter nine, the author delineates the town Ranikhet.

The house was built in the nineteenth century by the British, without architects or building plans. They made enormous stone mansions with chimneys and attics, fireplaces, and mantelpieces, but also deep verandas and tin roofs. To the extent that it was possible in distant India, they recreated their remembered Scotland (Roy, 2011).

Historically during the colonization of India, the Britishers did some changes in the surrounding, to make them feel as if they were in their motherland. These geographical changes were an impact of colonization. The author elaborates on how the manmade change to nature was affected even after the colonization. In part two, Chapter One, Diwan Sahib talks about the 1971 Bangladesh war with Maya. This narration showcases the lasting impact of war on people's lives as it not just influences the geography of a particular place, but alongside it affects their sense of self, as people not just lose their land and borders but their home, lives, belongings,

etc. "We are literally making our own future, and the consequences of these reconsiderations are profound for politics in general and security in particular" (Dalby, 2014). The Bangladesh war made the same impact as they fought for the borders. This kind of geographical change is called geopolitics. By remembering this war and the landscape change, the narrator is postulating the importance of creation and an ecocentric world as mountains, seas, and rivers must be protected by human beings. Geopolitics studies suggest that adopting an ecocentric approach to living would result in a more peaceful existence. Ecocriticism posits that a self-centered existence will ultimately result in devastation.

These incidents in the novel showcase how patriarchal society tends to change nature according to their personal comfort, but on the other hand, women tend to stay close to nature and they like to blend with the ecosystem. It proves women usually enjoy nature as such, but men like to bring changes to it.

Violation through Anthropocentrism to women and nature

As the plot progresses, few characters in the novel think that human beings especially men are the supreme species of other living organisms, which falls under the belief of anthropocentrism. Anthropocentrism means believing that human beings are the center of the world and behaving accordingly, some refer to this term as human exceptionalism as this kind of behavior leads to nature exploitation. "The Anthropocene is a troubling concept for troubled times. It speaks of a complex, interconnected and unstable world marked by globalized and manufactured risks that now are threatening the very life-upholding systems upon which human civilizations rest" (Lövbrand, Möbjork and Soder, 2020). Peter Barry, an ecocritic says that "no true wilderness any longer exists on the planet, for every region is affected by global warming, and other 'anthropocentric' problems, such as toxic waste and nuclear fall-out" (Barry, 2009). These lines indicate

that toxic waste, industrialization, and pollution are created due to anthropocentric ideas which lead to environmental problems created by human beings for their own materialistic benefit which is now badly affecting the environment. This way of living automatically portrays the supercilious mindset of humans which led to the practice of their ruling system above all other organisms in the universe. “The Anthropocentric view of human beings could be bounced back at human beings” (Karen, A. and Christopher, G., 2022). This practice in the long run will severely damage the environment like the change of agriculture to industrialization that led to the destruction of various crops and created land pollution. “Apparently, it is a manifestation of anthropocentric hubris to believe that human appearance on earth is the final stage of evolution, simply because no one knows for certain the future trajectory of evolution despite the present domination of the planet earth by human beings” (Shotunde, 2020). Certain characters in the novel follow the anthropocentric idea as they plan to set rules for animals on hillsides. In Chapter Ten, the hotel manager throws a party at Aspen Lodge where high-profiled people like generals, brigadiers, and bureaucrats are invited to the party. Suddenly a herd of cattle jumped into the party accidentally due to the lethargic behavior of Purana who is taking care of the cattle. The high-profile people at the party ill-treated Purana and his cattle. Mr. Chauhan said that “I’m going to lock you up! With your damned cows and goats!” (Roy, 2011). Considering the example, the readers witness the supercilious behavior of humans over cattle. Mr. Chauhan even placed a signboard for the cattle stating that they cannot trespass. People try to draw a circle for the cattle to live and eat. Mr. Chauhan and others in the party showed their anger towards the cattle because they accidentally invaded their place but the universal truth is that humans are invading the animal space by throwing a party on the hillside but in return, the animals never

vented out their anger on humans nor behaved as cruelly or brutally as humans.

The act of invading the nature space continuously takes place in the plot of the novel indirectly signifying this egocentric practice in the real world. In another instance, “And went off to inspect the site of a new amusement park, his flagship project, for which a swathe of the oak forest was being cleared” (Roy, 2011). These lines in the novel depict deforestation which again showcases human beings’ cruelty towards nature. These officials in the novel cut down an oak forest for their own happiness. This environmental threat was started back in Plato’s age as he states “severe deforestation, and lamented the loss of fertile soil and defaced landscape” (Wapner, 2010). The growth of environmental awareness has gradually increased because a vast amount of misuse of natural resources by human beings has led to the catastrophe of the whole environment and civilization. Likewise, the novel not only throws light on the beauty of nature but also mentions the catastrophe and the threat of existence that is faced by the Ranikhet village, especially women and nature. The novel comments on the brutality of deforestation when Diwan performs in the St. Hilda convent before the students which he has been doing for so many years. He conveys about deforestation that is going on in Ranikhet. Diwan says

“But no animal comes to spur now,” Diwan Sahib said. “There are trucks that come and go, the entrance to the spur is piled high with logs from trees that have been cut from the forest all around. Have you ever heard the sound of a tree being cut with saws coming apart at the trunk and falling?” (Roy, 2011)

Here Diwan is imparting how humans, in specific people in power, are destroying nature for their own welfare to children who is the future of our country. On the other hand, Purana is feeding a deer one day and he lost the deer, the higher officials took that deer to their place and arrest Purana for having a deer in his house. But in reality, he

saved the deer from the predators in the forest. The deer was taken to the zoo, it lasted only for the thirteenth day and it died of malnutrition. The zoo is created for wildlife welfare but the novel portrays incidents where people lock the wild animals in the zoo for their own pleasure. These instances portray the cruelty of anthropocentric behavior. “The relationship between two motives underlying environmental attitudes was examined: ecocentrism Valuing nature for its own sake, and anthropocentrism valuing nature because of material or physical benefits it can provide for humans” (Thompson and Barton, 1994).

Nature and Gender-centred view in the novel

The novel not just condemns ego-centric behavior but also teaches various ways to take care of the environment by incorporating various ecological concepts such as environmentalist and pantheism. Environmentalists refer to people who selflessly take care of their environment and it is visible in the novel as there are some characters like Diwan who enjoy nature and protect the environment of Ranikhet. In India, there are people who still worship nature as God, and that ideology is also indicated in the novel. This ideology comes under the theory called ‘Pantheism’, this theory believes in worshipping the universe as God. Ramesh who is an economics professor said these lines at the party, “No, no, Brigadier! Cows are holiness apart natural lawn mowers” (Roy, 2011). This dialogue indicates that Indians see cows as holy gods. Similar to Indians, Africans also worship nature as their God which has always been a significant aspect of their literature. Writers usually tend to write about aboriginals, tribes, and their customs, their relationship with nature, and their cultural connection with nature. Therefore, by prioritizing nature and emphasizing it as the focal point, one may effectively promote ecocentric views. The term ecocentrism is totally the opposite of anthropocentric. Ecologists and environmental philosophers use this term to manifest the nature-centred view. This nature-centred view always

contradicts the man-centred view. Aldo Leopold is an ecologist who is a prominent figure in environmental studies. He came up with the thought of ecocentric back in the 1940s. In anthropocentrism, they believe that man is above nature whereas in ecocentrism they believe man is part of nature. It also leaves a piece of thought that man is embedded in nature.

Ecocentrism sees the ecosphere comprising all Earth’s ecosystem, atmosphere, water and land as the matrix which birthed all life and as life’s sole source of sustenance. It is a worldview that recognizes intrinsic value in ecosystem and the biological and physical elements that they comprise, as well as in the ecological processes that spatially and temporally connect them (Gray J, 2018).

This promotes the importance of ecocentrism by Grey in his research article “Ecocentrism: What it means and what it implies”. In the novel, we can see the villagers peacefully living together with nature. Their only job is to take their cattle to the woods, bring them back safely and protect them from predators. Ranikhet is a beautiful village where people and nature coexisted together. In many places, the sound of nature is mentioned as Diwan Sahib relished in the sound of nature, he even does a performance for children in St. Hilda Convent with the bird’s sound. Diwan’s only companion is nature and Purana’s only companion is also nature. They both find peace when they are with nature. As a result, they give more importance and respect to nature.

Diwan’s hobby is teaching animal sounds and signs to the students of St. Hilda convent. After Diwan’s death, people find that he had a set of cassettes with recorded birdcalls even they notice that he had a pile of books which contain “Anthropology, the folklore of Kumaon, histories of India, handbound volume on the flora and fauna of the Himalayas, record of appointments in the princely state of Surajgarh” (Roy, 2011). Through these collections of books, readers

can appreciate his interest in nature. He gives importance to culture, nature, and society. At the time of Diwan Sahib's death, a sudden gale strikes the place, and the men of the village were astonished at the accident and said: "Diwan Sa'ab has taken a whole tree with him. The forest is mourning" (Roy, 2011). These lines show how Diwan Sahib was close to nature and the whole village knew it. There is another little girl in the village named Charu. Her only job is to take care of the cattle. She does this job very happily. Likewise, all these people keep nature as the central object of their life. The protagonist Maya's husband has died in trekking so she had a bitter past from which she tries to overcome. This beautiful background of Ranikhet nature healed Maya from her worse past. In Maya's life, nature played an important role and she also has a thick bond with nature from her childhood that clearly gets projected in her dream. The Ranikhet village people keep nature as the central part of their life because it is the only source of livelihood for them.

Revisiting the concept of 'Ecofeminism' through 'Gyn/Ecology'

The women characters in the novel are strong and daring. Therefore, the author tries to compare them with nature which comes under the concept called ecofeminism. Ecofeminism is a study of the relationship between nature and females; it usually comes from the branch of feminism. The term ecocriticism was coined by the French writer Fraçoise d'Eaubonne. The other theorist named De Beauvoir is a French philosopher and female activist. She states, "Woman is related to nature; she incarnates it: vale of blood, open rose, the siren, she represents to man the fertile soil, the sap, the material beauty and soul of the world" (Beauvoir, 1953). Beauvoir interrelates nature with females whom she called as open rose and siren. She wrote an essay called "Second Sex". In that book, she talked about different waves of feminism which helps to develop the growth of feminism theory. This book is viewed as a significant work of feminist

philosophy. Many feminists try to commemorate the relationship and connection between nature and women, and this study comes under both environmental and feminist studies. To study this relationship between females and nature is called ecofeminism. "Ecofeminism primarily renounces the patriarchal perspective where man is considered the epi-centre and has complete dominion over every creation" (Karen, A. and Christopher, G., 2023).

In *The Folded Earth*, the author gives importance to two-figures: they are nature and female characters and both are dominated by male society. Here nature is in threat of existence and female characters are dominated by the patriarchal society. During a party in Aspen Lodge an earnest woman asked a question to the brigadier "The woman mustered courage and said, "Sir, what about all this we hear, of army men raping and molesting women in the Northeast – and Kashmir" (Roy, 2011) the Brigadier's reply is "One rotten fruit here or there, Madam, doesn't make a bad basket. We deal with deviants, we do it more swiftly than anyone else" (Roy, 2011). Brigadier gave a negligent reply to her as the hotel manager tried to subdue the woman by keeping her quiet. Here the author brings a real incident to the readers which happened in Kashmir. A woman in the party was very bold to ask such a question before other higher officials to get justice for the poor girls of Kashmir. To study this much cruelty against women also come under Gyn/Ecology which we can list in feminist literary criticism. Gretchen T. Legler in his essay "Ecofeminist Literary Criticism" states that "Ecofeminism literary criticism is a hybrid criticism, a combination of ecological or environmental criticism and feminist literary criticism" (Legler, 2019) Here Legler says that environment criticism and female literary criticism together combine and create ecofeminism concepts. Gretchen T. Legler also states that

Ecofeminists argue that dealing with practical environmental problems is both an ecological and a feminist task

because the use and abuses of the environment that have led to what they see as the potentially catastrophic present are largely due to a patriarchal environmental ethic that has conceptualized land as “women” (Legler, 2019).

He elucidates those practical environmental problems like abuses of the environment and females lead to the catastrophe of the world due to the patriarchal mindset as environmental ethics denotes land as women. In the novel, the Ranikhet village environment is also affected by higher officials. To build an amusement park they plan to destroy a whole oak tree forest and Beena who is deaf and dumb is molested in the forest. Male supremacy ruins nature and female lives without any remorse. These characters show their supremacy over other living creatures as a result of patriarchal and environmental problems.

In some places, Maya’s life is compared to nature. While describing nature the author hides a secret message. “The lopsided half-moon was trapped in a cage of branches” (Roy, 2011). This line is taken from the novel which symbolizes Maya’s life. Maya was trapped in Veer’s love as she struggles to let go of her past life which is symbolically depicted in this line. Likewise, there are a lot of examples in the novel where we can signify the interconnection of women and nature through Maya’s life.

Similar to the enactment of patriarchy over nature, the domination of patriarchy over females is parallelly witnessed through the narration of certain events. Ama told a story of a man whose name is Vikki; he tries to seduce a widow and also tries to get her dead husband’s pension. It ends in the intimidation of the widow to stay still and lie for the sake of her husband. This event infuriates Ama and she asks “How do you trust a man so greedy he wants his wife to be called a widow?” (Roy, 2011). This makes Ama compare men with vultures which is a predatory bird that feeds upon fragile and vulnerable animals. Alongside, Veer narrates a similar story where

a man beats his wife to death. These terrifying stories about male domination show how men ruin the lives of women. These sets of incidents enable a certain kind of criticism which is known as ecofeminism. Similarly, there is also a concept that is similar to ecofeminism which is Gyn/Ecology. Ecofeminism portrays how women and nature are intertwined. Diwan once compares Maya with nature “Our Wild Himalayan rose is turning into a memsa’ab” (Roy, 2011). He also compares Maya with wild Himalayan Rose. This supports De Beauvoir’s words that “woman is related to nature” (Beauvoir, 1953). After establishing the bond, the writers highlight how women and nature suffer the patriarchal structure.

Several critics condemn how women are not let to grow in a patriarchal structure. For example, Ama once says that “But don’t teach her too much. Girls who study too much are no good for anything” (Roy, 2011). Here Charu’s own grandmother is not letting her study more. Gretchen T. Legler in his essay “Ecofeminism Literary Criticism” claims that “Ecofeminists also argue that constructions of nature as female (as mother/virgin) are essential to the maintenance of these harmful environmental ethics” (Legler, 2019). He illuminates that constructing nature as female is indispensable for maintaining a self-centred lifestyle.

Gyn/Ecology Explicit the Brutal Violation against Women

To study the cruelty against women in the novel also come under Gyn/Ecology which we can list in feminist literary criticism. The term Gyn/Ecology comes under the concept called radical feminism. The Book *Gyn/Ecology. The Metaethics of Radical Feminism* was written by Mary Daly. “When Daly introduced the distinction between foreground and background in *Gyn/Ecology*, she introduced a way for women to understand their own experiences of fragmentation” (Daly, 1990). In this book, Mary Daly lays out the oppression of women by patriarchal society and the journey of women’s becoming. She also brings out the

historical barbarous rituals by patriarchal societies all over the world. Where she says,

I will focus upon five specific righteous rites which massacre women: Indian *Suttee*, Chinese footbinding, African female genital mutilation, European witch burning, American gynecology. In examining these, I will seek out basic patterns which they have in common, and which comprise the Sado-Ritual Syndrome. Those who claim to see racism and/ or imperialism in my indictment of these atrocities can do so only by blinding themselves to the fact that the oppression of women knows no ethnic, national, or religious bounds. There are variations on the theme of oppression, but the phenomenon is planetary (Daly, 1990).

Mary Dale brings out the massacre of women around the world and she related with different concepts like Gyn/Ecology. In the novel *The Folded Earth* the readers can witness the massacre of women's minds and physical bodies like environmental deterioration. As mentioned earlier in the story, a child called Beena was molested by a man in the forest. The distressing aspect of the story is that her own mother blames the deaf and dumb child for this incident. "Her mother says it's her fault, she wears tight clothes and goes wandering in the market, and giggle at boy" (Roy, 2011). This saddest part happening in India to this day, by blaming the victim, it led the people who committed the crime to go back to a guilt-free state. The members of the village also fail to stand-up for the victim. Maya is the only person who has the sense to take Beena to the police station. Ama says that Beena's mother will never allow her to the police station. And she also says that "The less this news travels, the better for the girl" (Roy, 2011). The patriarchy not just attacks women but never provide any justice for the various suffering faced by women. They try to keep their women in silence by victim blaming and guilt-shaming. This unjust practise let women to go to a state

of living dead which is faced by every woman in the world. Men seek to put the women in the living death condition. In this novel, maximum women are oppressed by men in many ways which made women to live in the state of living death. Mary Daly in her book *Gyn/Ecology. The Metaethics of Radical Feminism* uses the term 'necrophilia' to explain how men seek to put women in the condition of a living dead state.

To stop this kind of molestation, Daly came up with the term Gyn/Ecology where she says "Gyn/Ecology can be Seen/Heard as a Thunderbolt of Rage that I hurled into the world against the patriarchs who have never ceased to massacre women and our Sister the Earth" (Daly, 1990). In the above sentences, we can see how Daly connects the environmental sufferings and women's oppression by same patriarchal society. She condemns the environmental deterioration and massacre of women which was done by patriarchs in her book. "The term Gyn/Ecology is useful for analysing this kind of subjugation of women and nature in patriarchal societies" (Karen, A. and Christopher, G., 2024).

Writers like Daly condemn the male society with the usage of a sharp word such as in the work, "Patriarchy is the homeland of males: it is the Father Land; and men are its agents" (Daly, 1990). Subsequently, in the novel, we can see how the higher official and politicians dominate women and nature. The protagonist Maya and other sub-characters like Charu and Beena live in a male dominant society and tolerates the supremacy of male domination in Ranikhet. There are few events in the novel narrated by other characters which were not depicted as a scene in any part of the novel but only narrated by other characters. These incidents show the act of violation against women's characters by the patriarchal society. At the end of the novel, when Ama is telling a story about a person called Vikki who marries a widow for money and tortures her every day, the reader is witnessing domestic violence. From the starting till the end the readers can witness

that women were going through some kind of domination in different ways. This kind of violation against women will take the novel into a different dimension. “All of them emerge themselves as ‘New Women’ exhibiting their potentials to subvert male power and at the end celebrating their autonomy in the lap of nature” (Pal, 2022). “Anuradha Roy’s women characters are sowed with the feminist attitude, which is fostered by the social issues they face with strong attitude” (Arularasi and Kumaran 2016). These incidents in the novel substantiate Mary Dale’s view on patriarchy. In her book *Gyn/Ecology* she says “Patriarchy is itself the prevailing religion of the entire planet, and its essential message in necrophilia” (Daly, 1990). She proposes that patriarchy is a prevailing religion that is spreading all over the earth. This kind of strong condemnation against patriarchy has been a part of gynocentric writing. This kind of domination against women supports the theory called ‘Androcentrism’. Androcentrism is considering the male point of view as the centre. Few incidents in the novel prove androcentric views that prevail in this novel. Charlotte Perkins Gilman conceived this term in a debate. Andro means ‘man’ and centrist means ‘centric’. In anthropocentrism, males dominate the entire world but in androcentrism patriarchal society dominates women. “Androcentrism refers to the propensity to center society around men and men’s needs, priorities, and values and to relegate women to the periphery. Androcentrism also positions men as the gender-neutral standard while marking women as gender-specific” (Bailey, 2019). Here Bailey postulates the facts of androcentrism in his research article “Is Man the Measure of all Things? A Social Cognitive Account of Androcentrism”. The readers can witness this kind of concept-oriented work in gynocentric writing.

Conclusion

The Folded Earth by Anuradha Roy is written from a gynocentric perspective. When women discuss this, we are able to see the

true colours of their pain, which is often more authentic and visceral than that of males. Gynocentric writing is attempted by female authors such as Adrienne Rich, Arundhati Roy and Sylvia Plath. Among this group, Adrienne Rich is a radical feminist. Similarly, nature has also been a part of certain author's writing, which shows that concern for the environment is always real. “The author Anuradha Roy glorifies the beauty of the Indian environment and also creates awareness about the ecological catastrophe” (Augustus and Gnanadurai, 2024).

The growth of environmental degradation could make the world look back on the theory ‘Ecocriticism’. Likewise, this paper aims to highlight that the world may be moving towards material abundance, but this book illustrates the threat to existence by discussing various of ecocritical concepts, including ecofeminism, ecocentrism and anthropocentrism, which prevails as a wake-up call for the future generation.

The novel *The Folded Earth* picturizes the Indian environment, culture, society and communities which evolves as the central thought of the novel. “*The Folded Earth* successfully engages its readers with numerous longstanding debates like the inseparable connection between women and nature as well as the destruction of the environment, in a truly human and emotive way” (Francis, 2022). The paper has been crafted to view the core idea in three dimensions. Firstly, the paper gives importance to nature and the novel shows how nature played a prominent role in the lives of women with the examples of few characters. The second dimension is the struggle of women against patriarchal domination and their inner strength to overcome these hardships. The last dimension is looking back at the problems of society in relation to nature and gender. These dimensions are the layout of this paper. These three important matters build up the storyline in Anuradha Roy’s *The Folded Earth*. Through this novel, the paper has explored various concepts, and they are androcentrism,

geopolitics, anthropocentrism, ecocentrism and ecofeminism and Gyn/Ecology.

References

- Karen, A. and Christopher, G. (2022). Understanding Environmental Degradation in the Novel *Animal's People* by Indra Sinha, in *Blossoming Research in Literature*, by J. Albert Vincent Paulraj and S Rajadurai, Royal Book Publishing, Salem, 133–139. <https://doi.org/10.26524/royal.160> (In English)
- Karen, A. and Christopher, G. (2024). Barbara Kingsolver intertwines ecology and women to symbolise community practitioners to comprehend the effects of climate change in her novel *Flight Behaviour*, *Community Practitioner*, 21 (4), 1392–1400. DOI: 10.5281/zenodo.11076646 (In English)
- Karen, A. and Christopher, G. (2023). Elucidating Gyn/Ecology in the Novel *The Bean Trees* by Barbara Kingsolver, *World Journal of English Language*, 13 (6), 278–283. <https://doi.org/10.5430/wjel.v13n6p278> (In English)
- Arularasi, N., and S Kumaran (2016). Portrayal of Women in Anuradha Roy's *The Folded Earth*, *Language in India*, 16 (3), 27–36. (In English)
- Augustus, K. and Christopher, G. (2024). An explication of anthropogenic climate change in Anuradha Roy's *An Atlas of Impossible Longing*. *Revista de Gestão Social e Ambiental*, 18 (1), 1–11. <https://doi.org/10.24857/rgsa.v18n1-077> (In English)
- Bailey, A. H., LaFrance, M. and Dovidio, J. F. (2019). Is Man the Measure of All Things? A Social Cognitive Account of Androcentrism, *Personality and social psychology review : An official journal of the Society for Personality and Social Psychology*, 23 (4), 307–331. <https://doi.org/10.1177/1088868318782848> (In English)
- Barry, P. (2009). *Beginning Theory: An Introduction to Literary and Cultural Theory*, Manchester University Press, New York. (In English)
- Beauvoir, S de. (1953). *The second sex*, trans. and ed. HM Parshley, Lowe and Brydone (Printers) Ltd, London, UK. (In English)
- Dalby, S. (2014). Environmental Geopolitics in the Twenty-first Century. *Alternatives: Global, Local, Political*, 39 (1), 3–
16. DOI: 10.1177/0304375414558355 (In English)
- Daly, M. (1990). *Gyn/Ecology The Metaethics of Radical Feminism*, Beacon Press Books, Massachusetts. (In English)
- Francis, A. (2022). Analysing Ecofeminist Perspectives in Anuradha Roy's *The Folded Earth*, *Smart Moves Journal IJELLH*, 10 (11), 14–24. DOI: 10.24113/ijellh.v10i11.11375 (In English)
- Gray, J., Whyte, I and Curry, P. (2018). Ecocentrism: What it means and what it implies, *The Ecological Citizen*, 1 (2), 130–131. (In English)
- Howarth, W. (2019). Some Principles of Ecocriticism, in Frederick, S. (ed.), *Ecocriticism Paradigms and Praxis*, New Century Book House Pvt. Ltd, Chennai, India, 57–78. (In English)
- Legler, T. G. (2019). Ecofeminist Literary Criticism, in Frederick, S. (ed.), *Ecocriticism Paradigms and Praxis*, New Century Book House Pvt. Ltd, Chennai, India, 112–123. (In English)
- Lövbrand, E., Malin, M., and Rickard, S. (2020). The Anthropocene and the geo-political imagination: Re-writing Earth as political space, *Earth System Governance*, 4, 1–8. <https://doi.org/10.1016/j.esg.2020.100051> (In English)
- Pal, M. (2022). Green Politics and Feminist Consciousness: Exploration of Ecofeminist Perspective in Anuradha Roy's *The Folded Earth*, *Literary Herald AN International Refereed/Peer-reviewed English e-Journal*, 7 (5), 5–12. (In English)
- Roy, A. (2011). *The Folded Earth*, Hachette Book Publishing India Pvt. Ltd, Gurgaon, India. (In English)
- Shotunde, A. O. (2020). Biocentric Egalitarianism and the Challenge of Anthropocentrism toward Environmental Sustainability, *Philosophia*, 26, 148–161. (In English)
- Thompson, S. C. G., and Michelle A. B. (1994). Ecocentric and anthropocentric attitudes toward the environment, *Journal of Environmental Psychology*, 14(2), 149–157. [https://doi.org/10.1016/S0272-4944\(05\)80168-9](https://doi.org/10.1016/S0272-4944(05)80168-9) (In English)
- Wapner, P. (2010). *Living through the end of nature: the future of American environmentalism*, MIT press, Cambridge, Massachusetts, US. (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.

Karen Augustus, M.A. in English Literature, Research Scholar, Department of English, School of Social Sciences and Languages, Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India.

Christopher Gnanadurai, Ph.D. in English Literature, Assistant Professor (Senior), Department of English, School of Social Sciences and Languages, Vellore Institute of Technology, Vellore-632014, Tamil Nadu, India.

UDC 8'82

DOI: 10.18413/2313-8912-2024-10-3-0-9

Kheya Chakravarty¹ 
Thenmozhi M² **Interpretation of Human Emotion:
An Implementation of Compassion through Karuna Rasa
in Manjula Padmanabhan's *Lights Out***¹ Vellore Institute of Technology,
Vellore, 632014, India*E-mail:* kheya.chakravarty2018@vitstudent.ac.in
ORCID: 0000-0002-6490-790X² Vellore Institute of Technology,
Vellore, 632014, India*E-mail:* thenmozhi.nisha@vit.ac.in
ORCID: 0009-0001-5052-3086*Received 21 January 2024; accepted 15 September 2024; published 30 September 2024*

Abstract: This study delves into the nuanced exploration of human emotion, specifically focusing on the manifestation of compassion through the aesthetic lens of Karuna Rasa in Manjula Padmanabhan's play, *Lights Out*. Emotions constitute a fundamental aspect of the human experience, and this research aims to unravel the intricate layers of compassion as depicted in the narrative. Drawing inspiration from classical Indian aesthetics, performing, and theatrical arts, Karuna Rasa encapsulates a complex ecstatic and emotional spectrum associated with compassion, empathy, and benevolence. Through a meticulous examination of Padmanabhan's portrayal and its impact, this study enhances our interpretation and understanding of the profound role played by Karuna Rasa in invoking and navigating compassionate sentiments. Employing a descriptive qualitative methodology, the research scrutinizes how Karuna Rasa is encapsulated, manifested, and expressed by the characters, contributing valuable insights into the intersection of emotions, artistic representation, and human experiences. By illuminating the interpretative dimensions of compassion within *Lights Out*, this study adds depth to the existing discourse on emotions in literature, offering a nuanced perspective on the transformative power of compassion in the realm of dramatic expression. This detailed analysis not only underscores the significance of Karuna Rasa but also highlights its ability to evoke a deeper empathetic response from the audience, thus enriching the literary and emotional fabric of the play.

Keywords: Karuna Rasa; Bharat Muni; Manjula Padmanabhan; Indian English Drama; Women Playwrights; Rasa Theory

How to cite: Chakravarty K., Thenmozhi M (2024). Interpretation of Human Emotion: An Implementation of Compassion through Karuna Rasa in Manjula Padmanabhan's *Lights Out*, *Research Result. Theoretical and Applied Linguistics*, 10 (3), 180-193. DOI: 10.18413/2313-8912-2024-10-3-0-9

Introduction

In the vast tapestry of human experience, emotions serve as the vibrant

threads weaving together the intricate narratives of our lives. At the heart of this emotional landscape lies compassion, a

profound and transformative sentiment that has been explored across various artistic mediums, including literature and drama. This study embarks on a comprehensive examination of human emotion, with a particular focus on the implementation of compassion through the lens of Karuna rasa in Manjula Padmanabhan's evocative play, *Lights Out*. As a powerful vehicle for self-expression and understanding, literature has long been a mirror reflecting the kaleidoscope of human emotions. From joy to sorrow, love to anger, the exploration of these emotional dimensions enhances our ability to connect with characters and narratives on a profound level. Amidst this rich emotional tapestry, compassion emerges as a compelling force, transcending cultural and temporal boundaries. It beckons readers and audiences alike to engage with characters' vulnerabilities, forging empathetic connections that resonate deeply. The term karuna in Sanskrit encompasses emotions of sorrow, pity, sadness, and the highest expression of which is compassion. Genuine compassion arises from the ability to empathize with the suffering of others. The root of sorrow lies in the mirage and illusion of love. When someone or something retreats from our lives, the sadness we feel is a consequence of our attachment and affection for them. Love becomes the source of our pain. Sadness emerges when we must detach from those we hold closest and dearest. Witnessing or hearing about the tragic or critical conditions, or the loss of a beloved person or thing, creates a distinct emotional response in the heart known as karuna rasa. This emotional state is a result of the separation from those nearest and dearest, with no prospect of reunion. In the words of Patnaik, "Karuna will surface where man has not only lost hope, but even the will to fight. ... A character hopes inside a work. But we as readers know the case to be hopeless" (Patnaik, 1996: 137-138). Karuna rasa, rooted in traditional Indian aesthetics and performing arts, provides a nuanced framework for interpreting and expressing compassion. The

term encompasses a spectrum of emotions associated with empathy, benevolence, and a profound understanding of suffering. Karuna rasa is the feeling that triggers us to weep or become emotionally upset and is the source for us to feel pain, even when we are not directly involved in the situation. This rasa is invoked or instigated by malediction or curse, disaster or adversity, separation and loss of near and dear ones, death of loved ones, and others. Anubhava (consequents) consists of changing of colour of face, tear shed, paleness, sagging of limbs, distraction and inattentiveness and sighing, misfortune, lament etc. Viyabhicari bhava (transitory emotional feeling) includes physical weariness and weakness, worry, panic, longing, confusion, fear, tear, change of voice etc. (Masson & Patwardhan, 1970: 61). The renowned scholar Rangacharya has described Karuna rasa as "...produced by seeing dear ones die (or killed) and by hearing unpleasant thing. Is acted by weeping, fainting, lamenting and crying loudly and also by acting physical fatigue and hurt" (Rangacharya, 1996: 59). The sthayibhava (permanent mental state) of this rasa is grief (or soka) and is exhibited by weeping.

Karuna rasa has been a recurring theme in literature across cultures and epochs. In classical Indian literature, it is one of the nine rasas or sentiments that form the basis of artistic expression. Karuna rasa, the sentiment of compassion, holds a cherished place in Indian drama, serving as a powerful tool to engage emotions, explore ethical dilemmas, and connect audiences with the complexities of the human condition. It encapsulates the deep sense of compassion and empathy felt towards the suffering of others. It involves transcending one's own perspective to enter the emotional space of another, sharing in their pain and struggle. This emotional resonance serves as a bridge between characters and audiences, fostering a connection that transcends cultural, societal, and temporal boundaries. For example, Great epics like the *Ramayana* and *Mahabharata* are imbued with characters who exemplify

karuna rasa. For instance, Lord Rama's compassion towards the suffering of his subjects and Arjuna's empathy for his opponents on the battlefield are instances where this sentiment is prominently featured or the famous work of Rabindranath Tagore *Chandalika*. Set against the backdrop of caste-based discrimination, the play revolves around Prakriti, a young low-caste woman, who experiences transformation through the compassion shown by Ananda, a Buddhist monk. Ananda's karuna rasa becomes a catalyst for Prakriti's self-discovery and empowerment. Ananda's compassion challenges the social norms that relegate Prakriti to the margins of society, emphasizing her inherent worth. This act of kindness ignites a sense of agency in Prakriti, prompting her to overcome the limitations imposed upon her. Another play *Evam Indrajit* by Badal Sircar delves into the existential crisis of the protagonist, Indrajit. The karuna rasa is evoked as his struggles with societal expectations, personal aspirations, and the quest for identity unfold. The audience is drawn into his emotional turmoil, prompting introspection about the human search for meaning. Thus, the above examples of drama demonstrate how karuna rasa becomes a vehicle for social critique and personal transformation, reinforcing its significance in Indian drama as a catalyst for change and a mirror to the intricacies of the human heart and society. The portrayal of compassion and empathy adds layers of complexity and relatability to characters, encouraging audiences to connect with their journeys and reflect on the broader human experience.

An Overview of the Play

The drama depicts a horrifying true story of gang-rape, maltreatment, and other brutal violence committed against a defenceless female victim in a flat that went unreported for weeks in a middle-class urban community in Bombay, present day Mumbai. The play is the outstanding exemplification of masses ignorance and lack of interest towards the suffering of women and in particular, of

the rape victim. The drama is the exposure of the barbarous violence against women witnessed by a group of middle-class people without taking any measure to put an end to the crime. The play criticizes the mental set up of the people, their attitude and behaviour towards the victim and the various kinds of violence inflicted on women on a daily basis. The drama is an important instance of violence inflicted on the poor innocent woman in a patriarchal society. The drama is a sure cut depiction of sexual and domestic violence on women by their perpetrators and near ones. The women in the said play is exploited both physically and psychologically both in the society and by the family members. The most heinous crime of witnessing the rape happening in their locality and acting merely as witness without taking any action shows the horrible alarming and distressing mentality of the people. The natural morality of people has degraded to such low level that they tend to save their skins without thinking of others welfare and benefit. The play starts with a grave and pressing discussion on the crime going on for a week without any hindrance between the young middle-class couple Leela and Bhasker in their neighbourhood. Leela was terrified and unsettled with the constant scream and crying of the victim which Rai described "destroy the fabric of domesticity of a middleclass couple." (Rai, 2006: 22). Manjula Padmanabhan skilfully employs the aesthetic nuances of Karuna rasa to navigate the complexities of compassion within the context of her narrative. This exploration unfolds against the backdrop of societal and personal challenges, offering a canvas upon which the characters grapple with their own emotions and the emotions of others. Set within the framework of a family facing unforeseen trials, Padmanabhan's play delves into the intricacies of human relationships, societal expectations, and the innate capacity for compassion amid adversity. The narrative unfolds with a dynamic interplay of characters, each navigating their own emotional landscapes as they confront the

shadows cast by societal norms and personal dilemmas.

Literature Review

In their article “Women and Violence in the Selected Works of Manjula Padmanabhan, Poile Sengupta And Kishwar Desai” (2024), the authors Khanna and Singh investigate how the works of notable Indian writers Manjula Padmanabhan, Poile Sengupta, and Kishwar Desai depict and treat women. Their stories primarily centre on women and question traditional gender roles and biases in Indian society. A detailed examination of the chosen plays reveals a repeating theme: women's empowerment and opposition to traditional male restrictions.

In her 2023 article titled “Manjula Padmanabhan’s Writing: A ‘Machete Strike’ At the Realities of Indian Society,” Padma examines Padmanabhan’s play *Lights Out*, which delves into the activities linked to the darkness of both physical and mental realms. She also explores the renowned play *Harvest*, which addresses the trade of human body organs by external and internal forces, highlighting the exploitation of impoverished individuals in India through the lens of advanced medical technology.

In the article titled “Dystopia in Disguise: Disintegrated Societies in Manjula Padmanabhan’s *Harvest* and *Lights Out*” (2023), Sunghee Pak explores the profound significance of Padmanabhan’s portrayal of two distinct worlds, illustrating how dystopia is intricately woven into reality, making it highly relevant to contemporary audiences. *Harvest* presents itself as a fictional dystopia, yet it retains enough realistic elements for modern viewers to identify with. Conversely, *Lights Out* begins as a realistic drawing-room drama rooted in a social incident but gradually incorporates dystopian qualities so extreme that the audience is inclined to deny their plausibility. Despite utopias and dystopias being theoretically non-existent, Padmanabhan's depiction of dystopian societies that intersect with everyday life amplifies her critique and commentary on the existing world.

The article “Beyond Reproduction: An epistemological Search for a “Woman” in Manjula Padmanabhan’s *Escape* and *The Island of Lost Girls*” (2023), Amrita Basu and Priyanka Tripathi have tried to find out the answer of the question, who is a woman. In today's gender-fluid world, answering this question requires exploring non-linear and nonbinary possibilities. As the biological definition expands to include fragmented identities, it becomes increasingly complex to pinpoint a simple, clear truth of recognition. Within third-world dynamics, this question gains additional layers. Against the backdrop of mass female genocide due to the pursuit of perfect cloning, Manjula Padmanabhan addresses the issue of womanhood in her futuristic dystopian works. She navigates this topic to reach a philosophical space that allows the modernist concept of ‘woman’ as a well-defined category to be reborn within a postmodern framework. This shift moves women beyond their traditional roles of reproduction and menial labour. By analyzing these selected works, this research article aims to define a model of womanhood and illustrate the plurality of truth in action.

In their article “Rereading the Impressions of Violence on Women's Identity and Mental Health – A Literary Analysis through the Works of Manjula Padmanabhan, Poile Sengupta, and Kishwar Desai (2023), Khanna and Singh aim to explore the impact of violence on women's identity and mental health through a literary analysis of the works of Manjula Padmanabhan, Poile Sengupta, and Kishwar Desai. The purpose is to investigate how these prominent female authors depict the effects of violence on the lives, self-perception, and psychological well-being of women in their writings. By examining these authors' works, this study seeks to contribute to a deeper understanding of the complex relationship between violence, gender, and mental health. The authors aim to highlight the specific challenges women face in societies affected by violence. The findings of this research could provide policymakers, mental health practitioners, and social

advocates with valuable insights to inform the development of tailored interventions to support women who have been affected by violence.

Azmat Ara in 2022 in the article “Gender Studies in Works of Manjula Padmanabhan” investigates gender studies and concerns in Padmanabhan’s works, including *Harvest*, *Lights Out*, and *Escape*, which are mostly about women. Padmanabhan addressed gender issues in them. Jaya, Leela, and Mejie are popular female characters who confront discrimination, gender issues, and subjection.

In the article “Echoes of Phallocentrism in Manjula Padmanabhan’s *Lights Out*” (2021), Nandini Vyas sheds light on the phallogentric, voyeuristic, sexist, and exhibitionist mindset that is impotent to solve the enigma, reflecting on Padmanabhan’s *Lights Out* as a cruel contradiction on the postmodern, urban, and educated, hypocritical phallogentric men and women.

In their 2021 article titled “The Theme of Gender Violence in Manjula Padmanabhan’s Play *Lights Out*,” Shivam Singh and Gunjan Sushil analyse Padmanabhan’s play *Lights Out* through the lens of gender insensitivity and violence.

In their article “*Lights Out*: The darkness in the lives of Women” (2020), Mane and Deshmukh examine the play in the light of gender insensitivity and its implications on women in Indian society.

In the year 2020, Mona Gupta and Abha Panday in their article “Gender and Communication in Manjula Padmanabhan’s *Lights Out*” have highlighted prominent ideas pertaining to language in relation to gender disparity and explore their relevance within the context of the play.

Bathina Rajesh Kumar in his article “Feminine Sensibility in the Select Play of Manjula Padmanabhan’s *Lights Out*,” (2017) aims to explore and examine the feminine psyche and sensibility of the protagonists in Padmanabhan’s work, while also addressing the broader challenges faced by Indian women within the urban and patriarchal

context of Mumbai. The play primarily centres on the marginalization and sensibility of women, highlighting the harsh realities of their lives in India. Manjula Padmanabhan aims to shed light on the suffering and hardships that women endure. The play captures the emotional and psychological turmoil that women experience.

In the article “Surrendering to the Menacing Darkness of Silence, Fear and Inaction in Manjula Padmanabhan’s *Lights Out*”, Anindita Ganguly (2014) examines the consumption of violence and the resulting fear, subtly depicted through the elements of silence and fear in Manjula Padmanabhan’s play *Lights Out*. Inspired by a real event, the play portrays a gang rape that could have potentially been prevented by the terrified, silent onlookers who passively witnessed the crime, thereby consuming violence with indifference. The play emphasizes the dramatization of helpless spectators, who, instead of intervening, engage in debates and speculate about the events happening in the neighbouring building. This scenario reveals the 'near-schizophrenic' existence within a metropolitan setting, where the "lights out" symbolizes not just something hidden behind closed doors, but a profound darkness in human compassion and humanity itself.

The Purpose of the Paper

The primary objective of this study is to dissect and interpret the multifaceted implementation of Karuna rasa in *Lights Out* by Manjula Padmanabhan, focusing specifically on how compassion is portrayed, embodied, and expressed by the characters and for this purpose, the paper examines and assesses numerous occurrences in the text to identify and point out instances and expressions of compassion through the perspective of Karuna rasa. Through a meticulous examination of the characters’ emotional journeys, the study aims to contribute to a deeper understanding of the interplay between artistic representation, human experiences, and the transformative power of compassion.

Novelty of the Study

The importance of this research lies in its potential to enrich our comprehension of compassion as a central theme in literature, particularly within the realm of dramatic expression. By unravelling the artistic nuances of Karuna rasa in *Lights Out*, the research contributes to the broader discourse on emotions in literature and theatre. Moreover, the study's findings will have implications for readers, audiences, and scholars alike, fostering a deeper appreciation for the portrayal and interpretation of compassion in the realm of dramatic arts.

Materials and Methods of Research

The primary source for this exploration is the text *Lights Out* by Padmanabhan. By employing a descriptive qualitative methodology, this research seeks to unravel the nuances of Karuna rasa, shedding light on its role in evoking and managing compassionate sentiments and this is complemented by the collection and analysis of secondary data. The author gathered information by reading books and article reviews, and conducting an Internet exploration to gain a deeper understanding of Karuna rasa. The procedure involved taking down notes and highlighting key points. The collected data was then organized and refined to extract essential information. Consequently, the study is specifically aimed at observing how Karuna rasa is depicted in the chosen work.

The Rasa theory, originated by Bharata Muni in primeval India in his renowned work *Natyashastra* stands as a fundamental notion in Indian aesthetics and the performing and theatrical arts. Rasa, denoting “essence” or “juice”, signifies the aesthetic encounter or spiritual and emotional resonance induced in the spectator through artistic expression. It establishes a structure for comprehending and assessing the emotive effect of artistic and creative performances. As per the Rasa theory, number of primary rasas exist is nine: love (sringara), laughter (hasya), compassion (karuna), anger (raudra), valor (vira), fear (bhayanaka), disgust (bibhatsa), wonder

(adbhuta), and tranquillity (shanta). All the rasas are linked to a particular emotional state (bhava), conveyed by characters or performers. The depiction of rasas encompasses diverse components like vibhavas (stimulants or determinants), anubhavas (consequents), sthayibhava (permanent emotional state), and alamkaras (ornamental enhancements). Vibhavas present the setting and stimulation for experiencing emotions, whereas anubhavas include artists' appearances, physical movements, vocal variations, and other physical aspects. The sthayibhava serves as the foundation for rasa manifestation, while alamkaras (ornamentation or adornment) enhance the aesthetic attractiveness. (Chakravarty, Thenmozhi, 2023: 387-393).

Rasa, according to Bharata, is a synthesis of bhavas (sentiments or emotions transmitted by a performer). It is the depiction of an individual's psycho-physiological state in a certain scenario. Rasa can also be found in music, dancing, painting, poetry, or any other artistic work capable of eliciting emotions or a psycho-physiological state. Bharata notes in his *Natyashastra* that rasa is formed by the mixture or combining of the vibhavas (stimulants or determinants), anubhavas (observable manifestations), and vyabharibhavas (transitory emotion) respectively. Bharata's famous statement regarding Rasa theory is *Vibhava-nubhava-vyabharicari-samyogaad-rasa-nispattih* (*Natyashastra*, VI). This is the key sentence for comprehending the rasa notion. Many other interpretations have been proposed, but the simple meaning of this sloka is that, the maturing/ripeness/appearance/production of rasa takes place/happens as a result of the combining, co-appearance – combination – of vibhava (determinates), anubhava (consequences), and sancaribhavas (transitory mental states). He also provides one example to further illustrate this concept. “When various condiments and sauces and herbs and other materials are mixed, a taste is felt, which is different from the individual tastes of the components and when the mixing of

materials like molasses with other materials produces many kinds of tastes, so also along with the different bhavas the Sthayi bhava (prevalent or permanent emotional state) becomes a taste, which is rasa, flavour or a feeling” (Rangacharya, 1996: 54-55). It implies that when we study a poem, a play, or a novel, there are certain reasons, certain events which bring us happiness, a few that produce gloom and sadness, and others that may even make us mourn because they trigger and stimulate our psychological and physical modes of body and we begin experiencing things this way.

The Rasa theory acknowledges the vital role of the spectator in accomplishing the aesthetic experience. The spectator’s emotive openness, compassion, and connection with the performances are pivotal for the realization and enjoyment of rasas. The full actualization of rasas occurs through the interplay among performers, characters, artistic elements, and the audience. This interaction is fundamental to the comprehensive experience of the emotional and aesthetic aspects. The theory maintains its influence in contemporary performing arts, spanning theatre, dance, music, and literature. It furnishes a comprehensive terminology and structural framework for dissecting and enjoying the emotive and aesthetic facets of artistic articulation. Delving into and mastering rasas not only enrich the artistic encounter but also offer profound insights into human emotions and the capacity of art to elicit and convey those emotions. This study examines the manifestation of Karuna rasa, with the primary focus on the play *Lights Out*.

Findings and Discussions

Lights Out by Manjula Padmanabhan deals with domestic and sexual assault on a woman and how it disrupts the neighbourhood’s complacent equilibrium. The play opens with conversation between Leela and Bhasker, the middle – class couple in whose locality the crime (the gang-rape) is taking place. The indifference of the male characters (Bhasker and Mohan) towards the crime is indeed an abominable act which is

horrifying. The pleading urge of calling the police by Leela to put an end to the crime is always overheard by both her husband and the others. The unbearable screaming of the innocent victim becomes intolerable for Leela but for Bhasker and Mohan, it becomes the source of never-ending discussion. First, Bhasker was in the phase of avoidance and deny Leela’s call to inform the police. Second, the visit of Mohan starts the denial phase where they both unaccepted the fact that the rape is happening and start discussing the crime in the light of religion, exorcism and others. Third, the denying happens with Bhasker and Mohan reluctant to accept the fact that the rape is happening daily in front of their apartment and trying their best to tag it with ridiculous explanation of their own. Male arrogance and prejudice are at its height when Bhasker’s friend Mohan stands with the former and mockingly refers to Sushila, Leela’s friend and neighbour as an intellectual while laughing at her assessment of their involvement in the crime. Although Leela is on the edge of losing her mind because the image of the abused woman lingers in her memory both consciously and subconsciously, the masculine world is unaware of the seriousness of the crime.

Leela patiently waits for them to conclude which never happens and ultimately, she loses her sanity, cries out in hopelessness and wiggle in pain as the perpetrators leave the crime scene. She accepts her fate and registers to the fact that she has to bear and continue with the pain of witnessing the rape happening outside her apartment daily. She wanted remedy for the problem whereas her husband is totally indifferent to the situation and tagged the crime as “minor little offenses” (BB, 5) which is not even worth for mentioning to the police. Bhasker ignores and overlooks Leela’s emotional degeneration and left her alone to face the problem all by herself. Her repeated cry to call the police was ignored again and again

LEELA: The police. Call the police.

BHASKER (makes a face): Oh, for god’s sake! (going back to his paper).

You're still worrying about that thing?

LEELA: All the time!

BHASKER: Look – what do you want me to do?

LEELA: Call them – you said you would!

BHASKER: But there's no point!

LEELA: Have you tried?

BHASKER: No. (Tries to concentrate on the paper)

LEELA (suddenly): You don't care what I feel, what I go through every day!" (BB, 2000: 4).

Through Leela's character, the sense of loss, fear and the feeling of hopelessness run through the play from the very beginning till to the end, giving rise to the Karuna rasa. The rape of the hapless unknown woman acts as the vibhava (cause) and the consequent or anubhava is Leela's pathetic mental state, her tears and fear, her despair and her utter state of despondency. The sthayibhava or the permanent mental state of her character is hopelessness, despair and pity. This pathetic state of Leela, the complaint, sadness and the despondency can be felt from the beginning of the play in her and is very clearly stated by her conversation with her husband

LEELA: I feel frightened. All through the day, I feel tense – ... My fear, it's – as if my insides are knotted up. ... I carry it around all day. ... it's like a shawl, ... wraps itself around my shoulders I start to shiver The sounds are bad enough! (Pause) And... you're sure we can't call the police?" (BB, 2000: 5-6).

Bhasker's indifference to interfere to other's business and lives makes Leela more and more helpless and hysterical. She suffered a lot due to negligence from her husband and amplified Karuna rasa throughout the play. Both Leela and Bhasker's way of interpretation of the crime is poles apart. Leela is consumed with fear because of the shrill screaming of the victim whereas Bhasker is casual and composed. Leela is sympathetic towards the women being raped, feels pity for them for their suffering, them being the victim of men's libidinous desire.

This may be because of the fact that she being a woman feels guilty and responsible towards the sufferer which in the words of Miceli and Castelfranchi is "The feeling of guilt is social where victim and bearer both suffer for the injury inflicted on victim and emphatically share the latter's suffering" (Miceli & Castelfranchi, 1998: 291). She urges her husband for help but all in vain. She strongly feels that "that by watching it, we're making ourselves responsible -" (BB, 2000: 6). For Bhasker, the screaming of the victim holds no place of pity and sympathy for him and he excused himself by saying that "If the police had to worry about things like that they'd be psychiatrists, not policemen I don't want to stick my neck out, that's all ..." (BB, 2000: 7). Thus, both Leela and Bhasker's outlook for the crime is different. Bhasker is more casual and callous minded towards the crime whereas Leela is sympathetic, more humanistic and she suffers at the pain and distress of the victim. She, thus, evokes karuna rasa. Also, as the conversation between the two proceeds, we came to know that Bhasker is secretly watching the crime taking place and this revelation makes Leela more hysteric, sick and vulnerable. The screaming is enough for her to make her sleepless and restless all through day and night. She is anxious and perturbed as the sounds are loud and taint the idea of her clean home. The horrible sound outside her home is so scary and disturbing that it paralyses Leela to such an extent that her daily life has come to an abrupt stop. She loses the capacity to act and behave like a normal human being.

LEELA (struggling in his half embrace): But their sounds come inside, inside my nice clean house, ... Why do they have to do it here? Why can't they go somewhere else? ... I feel awful, I can barely eat, I feel so sick" (BB, 2000: 8 – 9).

Leela's suffering from the unwanted disturbances created by the perpetrators in her life is manifested not only by her mental degradation but also by her physical gesture. Weissbecker & Clark has stated in their research article that "pains in the heart or

chest, nausea or upset stomach, dizziness” (Weissbecker & Clark, 2007: 914-919) are the best examples of physical and mental maladjustment whereas Pill on the same subject has described “chronic pain, ..., cardio-pulmonary, sexual, or conversion symptoms, ... signs of physical disorder that result from psychic or emotional disturbance” (Pill et al., 2017: 180).

The mental stress and paranoia of being constantly nagged by the hoarse and loud cry of the victim and the perpetrators add more intense physical pain for Leela. Desperate to stop the sound made by the raped woman and her perpetrators from reaching her, she “holds her head with both hands, covering her ears” (BB, 2000: 35). Because of her high level of compassion and empathy and pain for the raped woman, her body language and gesture indicate how she herself is suffering and traumatized from the crime and exhibiting the same through her bodily and psychological signs. She is so paranoid about the crime that even when the incident is not happening, she imagines the same and undergoes the same horrifying agony and experience. She starts to hallucinate and gets so hooked on the bellow that prior to the rape commences she conveys how awful it makes her feel. It is also noticed that the screams become so realistic to Leela that even when it is not happening, she imagines them and becomes obsessive which is gradually taking a toll on her, consuming her and disturbing her whole day routine. She openly admits to Bhasker that the screams of the victim in real or in her imagination affected her body extremely, wrapping her like a shawl. Thus, the atrocity committed outside her home made Leela so pathetic and grief stricken that she started showing bodily symptoms of nausea and loss of appetite, issue with mobility, sleep deprivation and “off balance” (BB, 2000: 17) resulting in her unsuccessful attempt to cope up with the physical and psychological distress and agitation. In the words of Lieder “unlike Bhasker, Leela experiences the daily screaming outbursts profoundly through expressing bodily agony ... describes as

affective response, or the physical embodiment of another’s pain” (Lieder, 2015: 525). Leela feels and understand the pain of the raped woman as she herself is going through the pain and trauma caused and inflicted by others. Caruth’s analysis and study in “Unclaimed Experience: Trauma, Narrative, and History” the witness hears the cry of anguish and pain in her own thoughts since she has personally felt it. Although the two aches may not specifically relate to one another, somehow one pain is able to or can communicate with the other (Caruth, 1995). Throughout the play, Leela diligently waits for the other members of her circle to act, to stop the crime and the ugly horrific sound but to her utter dismay, no one came forward to act, they continue with their discussion how to handle the situation which ultimately break the barrier of Leela’s patience and she suddenly breaks down amidst all the chaos.

LEELA: AAAAAAAHHHHH

BHASKER: Frieda, water!

NAINA: Leela, Leela!

LEELA: AAAAAAAHHHHH

MOHAN (at a safe distance): Come on, now, come on –

NAINA: Don’t take it so hard, Leela, no-one’s going to hurt you

LEELA. ... The sounds torture me. Tell the police I can’t sleep at nights . . .

tell the police the goondas must go away and take their dirty whores somewhere else! (She is losing control again.) . . . out of my hearing . . . out of my life . . .

(The hysterical note has returned.)” (BB, 2000: 43–44).

Leela’s continuous crying, yelling and refusing to drink or socially interacting with others are the signal she displayed of her inner turmoil and the pain she is going through. By bursting out, she releases all her pent-up emotion and fear. She tries her uttermost to keep herself calm, cool and patience both physically and mentally by practising yoga as suggested by her yoga guru. On the same note, to avoid the noise from interfering her daily routine and her

relationship with others, she picks up conversation on any topic with her friends. Like with Naina, she started conversing on random topic to keep her sanity intact and also to distract Naina's attention towards the commotion happening outside the house. For example, when the screaming started, Leela started talking with Naina, constantly changing the subject of discussion to divert her attention. Also, Leela's behaviour display moments of hysteria and anxiety and disconcertment as her husband informed her of an impending arrival of a guest to their home for dinner. Her way of dealing and handling the whole situation, the way she took breaks and pauses in her way of speaking all indicates the immense pain she is undergoing and her inability to deal with it. She was stiff and unfriendly towards Mohan, the guest, not because she wants to but the painful scenario she is in, made her act unfriendly and unresponsive. Her debacle to smile and sit comfortably or held a conversation and her looking out for the watch (the time when the animalistic shouts of the perpetrators and the hapless cry of the raped victim began) at regular interval reveals her inner turmoil and her inability to adjust with the painful situation. Her current situation is too painful for her to handle giving rise to her emotional breakdown and her character strongly evokes Karuna rasa.

Throughout the play, Leela embodies compassion, yet she remains powerless to stop the violence or even to convince others to care. This helplessness in the face of overwhelming apathy is a key element in evoking Karuna rasa. Leela's sorrow is deepened by her recognition that her compassion has no outlet, no way to effect change. As Leela's compassion is repeatedly thwarted, the audience is led to feel a profound sense of pity for her. They see her as a tragic figure, someone who feels deeply but is rendered ineffective by the social structures around her. This evokes a deep emotional response, fulfilling the purpose of Karuna rasa. In the climactic moments of the play, Leela's emotional state reaches its peak.

Her realization that no one will intervene, and that the violence will continue unchecked, leads to an emotional breakdown. The tragedy of her situation is fully realized, as the audience is left to contemplate the full weight of her sorrow. By applying Karuna rasa to Leela's character, her compassion and the resultant pity are brought to the forefront of the narrative. Leela becomes a tragic figure whose deep empathy and concern for others isolate her and lead to her emotional downfall. The audience is moved to pity her, not only for her compassion but also for the societal indifference that makes such compassion a source of personal pain. In this way, Leela's character serves as a powerful vehicle for evoking Karuna rasa, making the audience feel deep sorrow and empathy for her plight.

During the creation and emergence of Karuna rasa, the gang-rape of the innocent woman, and the perpetrators served as vishayaalambanavibhava (the cause and the reason) whereas the cry of the victim herself and Leela act as ashrayaalambanavibhava (the result of the cause and reason). The atmosphere of the whole neighbourhood, Leela's apartment and the animalistic cry serve as uddipanavibhava (the external stimulus). The anubhava (consequents) of the whole situation is Leela's conversation with the other family members, tears in her eyes, her pathetic condition and pale face. The transitory feeling or the Vyabhicharibhavas is weakness, anxiety, distress, terror, trembling and tears. Grief or shoka is the permanent emotion or sthayibhava in this condition. Leela begged repeatedly to her husband and the others to call the police and put an end to the crime happening outside her apartment daily and the criminals should be punished. The ongoing crime has disturbed and destroyed the peace of the neighbourhood and Leela's both, and had made her worried beyond words. This dreadful and devastating condition evokes the feeling of pity into the minds of the readers (shadharanikarana) and they realize the sentiment of pity, thus evoking karuna rasa.

Another character in the drama who strongly arouses *karuna rasa* and compassion is the unknown and unseen raped woman. The physical and mental harassment of the victim started and continued for a long time. The victim's cry for help is unheard and ignored by the people of the neighbourhood. Because of the fear and threats of the perpetrators, the people of the neighbourhood from the evening closed their windows and put the lights off. While Leela pleaded with her husband and his friend Mohan to call the police, they simply avoided the situation by saying that they have no right to interfere in others personal life. The victim's cry of help was unanswered only because people are opinionated about the woman's status, that she doesn't belong to the upper-class. They are sceptical about the rape. They even started the discussion as whether the rape is actually happening or it is some kind of religious rituals of the lower-class masses. Even they dare not go to the police as the role of police is doubtful in their mind. Even they went to the extent of tagging the woman as a prostitute who doesn't need to be saved from the clutches of her perpetrators stating that "Why should we get involved with some filthy woman and her paramours? A decent woman would never be with four men at once." (BB, 2000: 40-41). They very clearly sort out the difference between a decent woman and a prostitute, a prostitute has no sense of decency and can't be respected in the society and can be abused and taken advantage of at any time.... "You see, if she were a decent woman, we people would go to her rescue! (Pause). She is not, and so she's being left to her fate!" (BB, 2000: 41).

Amidst their heated debate about the ferocity of the crime, they do forget the fact that even for a whore, rape is both a psychological and physical torture inflicted on a woman and in the whole process, her self-dignity and her existence as a human being is denied and violated. As Paul Gilbert observed: "Shame, stigmatization, and ostracism are part of the process by which a group of others decide whom to associate and

cooperate with, and whom to exclude, reject, and avoid." (Gilbert, 2003: 1216). It has been an established fact revealed by various study that rape terrorize and threaten the moral dignity and awareness of a woman. Both her mental and physical state is shattered by this heinous crime and left her in a state of utter shock and disintegration. It has been explained by the famous psychologist Griffin that "rape victim's entire inner self is disintegrated by this physical and emotional exploitation, ... annihilation of woman by men as a human being" (Griffin, 2015: 129). Thus, rape is a deeply traumatic experience that can have profound and long-lasting effects on a woman's physical, emotional, psychological, and social well-being. The consequences of rape extend far beyond the immediate act itself, encompassing a complex array of physical and psychological challenges that can reshape the trajectory of a woman's life. The same happened with the innocent victim. She was in a state of shock and was shattered by the torture inflicted on her both mentally and physically. Her constant plea and cry for help was ignored and overheard by everyone present. Instead of helping her, they were busy discussing the intensity of the crime and whether to report the crime to the police or not. In their animated and intense discussion, they forget the naked truth that it is pointless and fruitless to point out and tag the identification of a woman who has been harassed and tortured to the point of death and whose "life is over. She will commit suicide, if she lives at all" (BB, 2000: 46).

At the beginning, the voice of the woman crying and screaming for help was heard clearly and distinctly. The voice was truly ragged and hoarse. At first, it was intense and fresh with determination and robust, the woman fighting against her perpetrators "let me go", (BB, 2000: 29), seeking and crying for help "help me" (BB, 2000: 29). But as the time passes, her voice gradually reduces to "general screaming, sobbing, with a jagged, tired edge to it." (BB, 2000: 29). With the passage of time she grew

tired and exhausted, screaming replaced with hiccups, and ultimately by the end of the evening, it stopped totally “must be over for tonight.” (BB, 2000: 53). The goons left after they are done with her, leaving her shattered, broken and scarred for life. The whole scenario is sombre and serene. The physical and mental condition of the victim, the helpless innocent woman cannot be determined and gauged. Her feelings and thoughts were far more complicated to be analysed and understood from the mere understanding of Bhasker and Mohan and others. Within the timespan of an evening, she went through a tough phase of pain and misery inflicted on her both psychologically and physically.

The raped woman’s helpless situation, her constant plea for help, to save her from the criminals, to save her from being dishonoured garners the sympathy of the readers. She is a victim full of empathy and pity, thus evoking Karuna rasa. Through the character of the rape woman, Karuna rasa is powerfully evoked in *the play*. Her unseen suffering, helplessness, and the injustice she endures all serve to draw out deep feelings of compassion, pity, and sorrow in the audience. Her character becomes a symbol of broader societal issues, making the emotional impact of her plight even more profound. The rape woman’s tragic fate, combined with the indifference of those around her, creates a compelling and heart-wrenching experience of Karuna rasa, leaving the audience to grapple with the intense emotions her character evokes. The whole scenario poignantly reflected Karuna rasa. The reason or the vishayaalambanavibhava in this case is the rape taking place, the evil deeds of the perpetrators and the ignorance of the people. And the result or the ashrayaalambanavibhava of this evil deed is the helpless cry and effort of the raped victim to free herself from the clutches of her abuser. The grim atmosphere and the animalistic cry of the goons act as the external stimuli or uddipanavibhava. The anubhava or the consequents of the situation consists of victim’s helpless cry for help, and

the horrible pathetic situation she is in. The transitory feeling or the Vyabhicharibhavas consists of fear, anxiety, distress, and terror. The permanent emotion or sthayibhava is pity and grief (shoka). This horrible and catastrophic circumstance creates compassion in the minds of the readers (shadharanikarana), and they recognize the pity, invoking karuna rasa.

Conclusion

Thus, it is revealed that the characters in Manjula Padmanabhan’s *Lights Out* suffers throughout their time span. The pain, suffering and the pitiful state of the characters stir up our sentiment of Karuna rasa. The characters, Leela and the unknown rape victim present pathetic illustration of humanity’s unending helplessness and inherent shortcomings. The drama depicts the harrowing and terrible journey of Leela and the raped woman in a few hours. They evoke and symbolize universal human emotions such as empathy, sympathy, and compassion. Because of their struggle and the suffering, they experienced at the hands of patriarchal society, the audience’s sympathy grows for them. The main (or the angī) rasa produced in the play *Lights Out* is karuna rasa. Padmanabhan has masterfully evokes Karuna rasa through her characters, creating a profound emotional experience for the audience. Leela's deep empathy and moral struggle embody Karuna rasa, as her compassion becomes both her strength and her source of profound sorrow. Her visible distress, isolation, and emotional breakdown evoke pity and sorrow in the audience, who are moved by her futile attempts to incite action in an indifferent society. Through Leela, the audience experiences the tragic consequences of unheeded compassion, leading to a powerful sense of Karuna rasa.

The rape woman, though never seen, serves as the emotional core of the play. Her cries of suffering, combined with the indifference of the surrounding characters, create a potent atmosphere of pathos. The audience is compelled to feel deep compassion for her plight, as she symbolizes

the broader issue of societal apathy towards gender-based violence. Her character elicits a profound sense of sorrow and pity, reinforcing the play's overarching theme of helplessness in the face of unchecked brutality. Together, these characters and their experiences weave a narrative that consistently draws the audience into an emotional journey filled with compassion and sorrow. *Lights Out* thus becomes a powerful exploration of Karuna rasa, leaving the audience with a lingering sense of pity, empathy, and contemplation about the moral failures of society. The play's ability to evoke such deep emotional responses underscores its significance as a work of contemporary Indian drama, with Karuna rasa at its heart. By applying Karuna rasa, the analysis uncovers the deep layers of emotional and ethical complexity within the play. The characters, their responses, and the tragic atmosphere they inhabit are all crucial to understanding how Padmanabhan uses compassion to critique society and evoke a powerful emotional response in the audience.

In the grand tapestry of human emotions, karuna rasa stands as a testament to our capacity for understanding, empathy, and compassion. It is a force that transcends cultural, geographical, and temporal boundaries, uniting us in our shared humanity. Whether found in the pages of literature, the science of psychology, or our everyday lives, karuna rasa serves as a guiding light, inspiring us to alleviate suffering, foster connection, and build a more compassionate world.

It's crucial to recognize that Karuna rasa, as a creative and artistic notion, predominantly resides within the domain of aesthetics and theatrical and performance studies. Although there might be research endeavours delving into associated aspects like the emotional or physical impacts of emotions like compassion or pity, it is imperative to distinguish between the artistic notion and scientific investigations into sentimental feelings and emotions. Additional research is encouraged to delve into how Karuna rasa can be employed in various

artistic expressions and its possible therapeutic significance and emotional and psychological well-being.

Corpus Material

BB – Lakshmi, C. S. (2000). *Body Blows: Women, Violence, and Survival: Three Plays*. Seagull Books Pvt Ltd.

References

Ara, A. (2022). Gender Studies in Works of Manjula Padmanabhan, *The Criterion: An International Journal in English*, 13 (I), 167–175. (In English)

Basu, A. and Tripathi, P. (2023). Beyond Reproduction: An epistemological search for a “Woman” in Manjula Padmanabhan’s *Escape and The Island of Lost Girls*, *Acta Universitatis Sapientiae. Philologica*, 15 (1), 37–53. (In English) <https://doi.org/10.2478/ausp-2023-0003>

Caruth, C. (1995). *Trauma: Explorations in Memory*, JHU Press. (In English)

Chakravarty, K. and Thenmozhi, M. (2023). Understanding Human Emotion: An Intervention of Anger through Raudra Rasa in Dina Mehta’s Drama *Brides Are Not for Burning*, *World Journal of English Language*, 13 (7), 387-387. (In English) <https://doi.org/10.5430/wjel.v13n7p387>

Gilbert, P. (2003). Evolution, social roles, and the differences in shame and guilt, *Social Research an International Quarterly*, 70 (4), 1205–1230. (In English) <https://doi.org/10.1353/sor.2003.0013>

Griffin, S. (2015). *Rape: The Politics of Consciousness*, Open Road Media. (In English)

Ganguly, A. (2014). Surrendering to the Menacing Darkness of Silence, Fear and Inaction in Manjula Padmanabhan’s *Lights Out*, *Galaxy: International Multidisciplinary Research Journal*. 3 (III). 1–6. (In English)

Khanna, A. and Singh, R. (2024). Women And Violence In Selected Works Of Manjula Padmanabhan, Poile Sengupta And Kishwar Desai, *Educational Administration: Theory and Practice*, 30 (4), 6491–6498. (In English)

Kumar, B. R. (2017). Feminine Sensibility in the Select Play of Manjula Padmanabhan’s *Lights Out*”, *Research Journal of English language and Literature*, 5 (3), 807–810. (In English)

Khanna, A. and Singh, R. (2023). Rereading the Impressions of Violence on Women's Identity and Mental Health – A Literary

Analysis through the Works of Manjula Padmanabhan, Poile Sengupta, and Kishwar Desai, *Journal of Namibian Studies: History Politics Culture*, 34, 1312–1326. (In English)

Lieder, K. F. (2015). Lights out and an ethics of spectatorship, or can the subaltern scream?, *Peace & Change*, 40 (4), 517–538. (In English) <https://doi.org/10.1111/pech.12146>

Muni, B. (1986). *Nāṭyaśāstra: (English Translation with Critical Notes)*. (In English)

Miceli, M. and Castelfranchi, C. (1998). How to Silence One's Conscience: Cognitive Defenses Against the Feeling of Guilt, *Journal for the Theory of Social Behaviour*, 28 (3), 287–318. (In English) <https://doi.org/10.1111/1468-5914.00076>

Masson, J. M. and Patwardhan, M. V. (1970). *Aesthetic rapture: the Rasādhyāya of the Nāṭyaśāstra. I. Text*. Deccan College Postgraduate and Research Institute. (In English)

Mane, M. N. and Deshmukh, A. S. (2020). Lights out: The Darkness in the lives of Women, *Studies in Indian Place Names*, 40 (71), 1746–1753. (In English)

Pill, N., Day, A. and Mildred, H. (2017). Trauma responses to intimate partner violence: A review of current knowledge, *Aggression and Violent Behavior*, 34, 178–184. (In English) <https://doi.org/10.1016/j.avb.2017.01.014>

Patnaik, P. (1996). *Rasa in aesthetics: An Application of Rasa Theory to Modern Western Literature*, D.K. Printworld Pvt. Ltd. (In English)

Padma, N. D. R. (2023). Manjula Padmanabhan's writing: A 'Machete Strike' at the realities of Indian Society, *Journal of Namibian Studies*, 37, 1094–1102. (In English) <https://doi.org/10.59670/1pkfs366>

Pak, S. (2023). Dystopia in disguise: Disintegrated societies in Manjula Padmanabhan's *Harvest and Lights out*, *the CEA Critic/CEA Critic*, 85 (2), 153–170. (In English) <https://doi.org/10.1353/cea.2023.a901809>

Pandey, A. and Gupta, M. (2020). Gender and Communication in Manjula Padmanabhan's *Lights Out*, *Journal of Teaching and Research in English Literature*, 11 (3), 23–27. (In English)

Rai, R. N. (2006). Perspectives and Challenges in Indian-English Drama, in Tandon, N. (ed.) *Perspectives and Challenges in Indian English Drama*, Nice Printing Press, Delhi, India, 11–23. (In English)

Rangacharya, A. (1996). *Introduction to Bharata's Nāṭyaśāstra*. Munshirm Manoharlal Pub Pvt Limited. (In English)

Singh, N. S. and Sushil, N. P. G. (2021). The theme of gender violence in Manjula Padmanabhan's play *Lights Out*, *The Creative Launcher*, 5(6), 34–38. (In English) <https://doi.org/10.53032/tcl.2021.5.6.06>

Sirkar, B. (1974). *Evam Indrajit*, Oxford University Press, Calcutta. (In English)

Tagore, R. (1954). *Three Plays: Mukta-dhara, Natir Puja, Chandalika*. (In English)

Vyas, M. N. A. (2021). Echoes of Phallocentrism in Manjula Padmanabhan's *Lights Out*, in *Modern Narratives in Literature, Cinema, Culture and Society Select Conference Papers of LUMOS, English Conference-2021*, 21. (In English)

Weissbecker, I. and Clark, C. (2007). "The impact of violence and abuse on women's physical health: Can trauma-informed treatment make a difference?", *Journal of Community Psychology*, 35 (7), 909–923. (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.

Kheya Chakravarty, M. A. in English Literature, Research Scholar, Department of English, School of Social Sciences and Languages Vellore Institute of Technology, Vellore, India.

Thenmozhi M, M. A., M. Phil and Ph. D in English Literature, Assistant Professor Senior, Department of English, School of Social Sciences and Languages Vellore Institute of Technology, Vellore, India.

УДК 82.0

DOI: 10.18413/2313-8912-2024-10-3-1-0

Шулятьева Д. В.¹ 

**Опыт погони: нарративная скорость и ритм
в современных разветвленных нарративах**

¹ Национальный исследовательский университет «Высшая школа экономики»
ул. Мясницкая, 20, Москва, 110000, Россия
E-mail: dshulyatyeva@hse.ru
ORCID: 0000-0001-7498-7395

*Статья поступила 1 апреля 2024 г.; принята 15 сентября 2024 г.;
опубликована 30 сентября 2024 г.*

Аннотация. В статье рассматривается проблематика скорости и ритма в современных разветвленных нарративах и способы создания высокоскоростного переживания для читателя. Разветвленные повествования включают в себя события, которые только могли произойти с героем, но не произошли в пределах нарративного мира. Разветвленные повествования получают распространение в современной прозе (П. Остер, К. Аткинсон, Ф. Рот, Б. Жиро и др.) и, при всех сюжетных отличиях, специфичны тем, что моделируют для читателя переживание высокоскоростного движения. Такое переживание может создаваться за счет репрезентации физического движения в повествовании, за счет преобладания «резюме» и «эллипсисов» над «паузами» и «сценами» (в терминах Ж. Женетта), за счет чередования «показа» и «рассказа», за счет увеличения количества «поворотных моментов» и их преобладания над «филлерами» (по Фр. Моретти), за счет сокращения «виртуальных» событий и увеличения «актуальных» (по М.-Л. Райан), а также при помощи режимов «умножения» и «вычитания» (по К. Хьюм). Принципиальным становится чередование повествовательных элементов, отвечающих за ускорение, и иных, создающих замедление. Но нарративная скорость может быть рассмотрена и в энактивистской перспективе – такая попытка (в дополнение к предшествующей традиции) тоже предпринимается в статье. Высокоскоростное переживание в таком случае формируется за счет создания менее опосредованного контакта читателя с происходящими в повествовании событиями: это переживание только усиливается и дополняется ритмическими приемами, отличными от скоростных. Если создание скорости опирается на принцип различия (смены), то создание ритма – на принцип повторения. Их одновременное воздействие позволяет читателю устанавливать телесно-миметические отношения с повествованием, которые и характеризуют его опыт при взаимодействии с таким типом нарративов. При этом спецификой такого типа нарративов становится преобладание виртуального над актуальным: и в таком случае виртуальное, замедляя нарративную скорость, парадоксальным образом формирует высокоскоростное переживание для читателя.

Ключевые слова: Нарративная скорость; Ритм; Телесный мимесис; Разветвленное повествование; Замедление; Ультракинетический монтаж; Энактивизм

Информация для цитирования: Шулятьева Д. В. Опыт погони: нарративная скорость и ритм в современных разветвленных нарративах // Научный результат.

Вопросы теоретической и прикладной лингвистики. 2024. Т. 10. № 3. С. 194-215.
DOI: 10.18413/2313-8912-2024-10-3-1-0

UDC 82.0

DOI: 10.18413/2313-8912-2024-10-3-1-0

Dina V. Shulyatyeva¹ 

**The chase experience: narrative speed
and rhythm in contemporary forking-path narratives**

¹HSE University,
20 Myasnitskaya St., Moscow, 110000, Russia
E-mail: dshulyatyeva@hse.ru
ORCID: 0000-0001-7498-7395

Received 1 April 2024; accepted 15 September 2024; published 30 September 2024

Abstract: The article examines the issues of narrative speed and rhythm in contemporary forking-path narratives and narrative devices that enable to create a fast-paced experience for the reader. Forking-path narratives represent events that could only have happened to the hero, but did not happen within the narrative world. They are becoming increasingly common in contemporary fiction (P. Auster, P. Roth, K. Atkinson, B. Giraud). Despite the differences in plot, these narratives are unique in that they offer the reader a sense of high-speed movement through the story. Such an experience can be shaped by the representation of physical movement in the narrative, by the predominance of “summaries” and “ellipses” over “pauses” and “scenes” (in terms of G. Genette), by the alternation of “showing” and “telling”, by an increase in the number of “turning points” and their predominance over “fillers” (according to Fr. Moretti), by reducing the “virtual” events and increasing the “actual” ones (according to M.-L. Ryan), as well as by using the “multiplication” and “subtraction” modes (according to K. Hume). But narrative speed can also be considered in an enactivist perspective – such an attempt (in addition to the previous tradition) is also made in the article. In this case, a fast-paced experience is formed by creating an indirect contact of the reader with the events taking place in the narrative: this experience is only enhanced and complemented by rhythmic techniques. If the creation of speed is based on the principle of difference (change), then the creation of rhythm is based on the principle of repetition. Their simultaneous impact enables the reader to establish bodily-mimetic relationships with the narrative, which characterize his/her experience when interacting with this type of narratives. At the same time, the specificity of this type of narratives becomes the predominance of the virtual over the actual: and in this case, the virtual, slowing down the narrative speed, paradoxically forms a high-speed experience for the reader.

Keywords: Narrative speed; Rhythm; Bodily mimesis; Forking-path narrative; Retardation; Ultrakinetic editing; Enactivism

How to cite: Shulyatyeva, D. V. (2024). The chase experience: narrative speed and rhythm in contemporary forking-path narratives, *Research Result. Theoretical and Applied Linguistics*, 10 (3), 194-215. DOI: 10.18413/2313-8912-2024-10-3-1-0

**Нарративная скорость в теории: от
формалистов к энактивистам**

Повествование как в литературе, так

и в кино моделирует для читателя/
зрителя опыт темпорального движения и

переживания. В создании такого опыта участвуют и нарративная скорость, и ритм повествования: нередко эти понятия используются как взаимозаменяемые (Caracciolo, Kukkonen, 2021: 134), но они все-таки требуют аккуратной дифференциации и сопоставления – со структурной точки зрения, но и с прагматической: они отвечают за разные измерения читательского переживания. В англоязычных исследованиях встречается понятие “narrative pace”, “narrative speed”, “narrative velocity”, “narrative acceleration”, “narrative rhythm”. В чем их отличия? Скорость репрезентации отвечает за продвижение событийности, характеризует нарративную прогрессию, а потому определяется *различием*: переходом от одного события к другому, сменой точек зрения, ракурсов, объектов изображения. Ритм же, в свою очередь, определяется *повторением*: возвращением к тому или иному типу события, временному промежутку, способу репрезентации.

Проблема нарративной скорости – одна из самых недостаточно изученных в современной нарратологии, на что нередко указывают сами исследователи (Hoffman, 2024; Caracciolo, 2020), хотя внимание этому аспекту функционирования повествования уделялось еще формалистами, затем – представителями классической нарратологии, после – более поздней постклассической нарратологической традицией, включая знактивистскую. Мы последовательно рассмотрим эти традиции в попытке определить, каким образом нарративная скорость функционирует в повествовании, на каких уровнях его организации задается, как участвует в создании того читательского и зрительского переживания, которое можно назвать «высокоскоростным». Значимыми для нас окажутся и те подходы, которые рассматривают нарративную скорость со структурной точки зрения, но и подходы, ориентированные на прагматическую традицию. Не меньшее значение имеет и

соотношение понятий «нарративная скорость» и «ритм повествования» – мы постараемся выяснить, как они соотносятся и как могут быть дифференцированы.

В статье «Связь приемов сюжетосложения с общими приемами стиля» В. Б. Шкловский говорил о замедлении как о значимом свойстве любого повествования – ведь без него, без этой «отложенной» событийности, любая история могла бы быть рассказана всего за несколько предложений. Такое «замедление» («задержание» в его терминах) может создаваться разными способами, но в том числе, на что указывает Шкловский в статье, при помощи ритма, создаваемого разными типами повторов: они-то и позволяют «задержать» действие, отложить момент наступления развязки (Шкловский, 1929: 33). «Замедление», о котором говорил Шкловский, позднее получило название «нарративной ретардации» (Hume, 2005: 105): это понятие связывает два уровня событийности в повествовании, позволяет описать их взаимодействие, указать на то, как модифицируется темпоральность фабульной событийности при ее «переводе» в сюжет. Оно, однако, не дает нам четкого представления о том, как функционирует читательский опыт при взаимодействии с повествованием: «замедление», о котором пишет Шкловский, не предполагает непосредственной оглядки на читателя, не пытается охарактеризовать его переживание. Нарративная ретардация потому (в формалистской традиции) – не эффект, который повествование способно производить на читателя, а структурное свойство повествования, характеризующее способ рассказывания, но не intersubjective взаимодействие читателя с ним.

Метафора замедления возникает и в другой работе Шкловского – уже в применении к искусству. В статье «Искусство как прием» он пишет о

замедлении как об особой операции, которой искусство подвергает привычные вещи и процессы: в результате они оказываются показаны иначе, но – что, наверное, важнее – и восприняты иначе, и иначе пережиты. Этот опыт иного (замедленного) восприятия транслируется и читателю, он тоже подключается к нему. Опыт замедленного восприятия, по Шкловскому, позволяет читателю пережить «каменность» камня (Шкловский, 1929: 13). И это чувственное переживание, как следует из его размышления, по-видимому, связано и со скоростью, и с ритмом повествования (повторение способствует автоматизации, в его терминах, а обновление, т.е. смена, создает деавтоматизацию), но может быть рассмотрено не только со структурной точки зрения, но и в перспективе соучастия читателя в этом переживании.

Еще Генри Джеймс указывал на два способа репрезентации событий в нарративе – их он называл «показ» и «рассказ», и уже в XX веке его наблюдение было подхвачено нарратологами, получив обоснование и осмысление. В самом общем смысле «показ» предполагает такое сокращение дистанции между читателем и повествовательным миром, что присутствие нарратора (или иного посредника) в этом взаимодействии не ощущается. Рассказ предполагает обратное: участие нарратора в читательском взаимодействии с повествованием ощутимо, а читатель вовлечен не только в изображаемые события, но и в процесс рассказывания о них. Эффект иммедиации, возникающий в случае с «показом», создается во многом за счет того, что «оценка» ситуации (инференция) перекладывается на читателя, тогда как в случае с «рассказом» эта «оценка» уже была произведена нарратором и теперь сообщается читателю. Сравним два примера, которые приводят Т. Клаук и Т. Кёппе (Klauk, Köppe, 2014: 846): «Джон рассердился на жену» и «Джон посмотрел на жену, нахмурил брови, сжал

губы и сжал кулаки. Потом он встал, хлопнул дверью и вышел из дома». В первом случае читателю предложено описание ситуации, предполагающее, что кто-то (нарратор) уже сделал собственное заключение о ней (Джон рассердился), потому его присутствие и должно являться ощутимым (хотя бы имплицитно). Во втором случае это «заключение» может сделать только читатель, опираясь на косвенные указатели, представленные в повествовательной ситуации: описанные действия героя и его состояние лишь направляют читательское воображение в сторону «вывода», но не сообщает его напрямую. Показ потому требует большего внимания читателя, но и большей временной «отдачи» со стороны повествования: он продвигает события медленнее, чем это делает рассказ (что заметно даже на этом примере). Показ предполагает, что читатель становится в позицию свидетеля происходящего, тогда как в случае с «рассказом» он доверяется иному свидетелю – нарратору. Показ потому в большей степени направлен на воссоздание опыта восприятия, в процессе которого наблюдатель, вероятно, не сразу делает заключение о том, что уже произошло, но «включается» в еще происходящее и разворачивающееся на его глазах. Показ, по-видимому, участвует в создании зрительно-миметических отношений между читателем и повествовательным миром, он направлен на постепенное «включение» читателя в происходящие события. Эта «постепенность» проявляет себя, конечно, и на уровне нарративной скорости. Показ замедляет скорость повествования, рассказ – эту скорость увеличивает (Klauk, Köppe, 2014: 850), но переживание этой скорости читателем все-таки зависит не от первого и второго по отдельности, но от их чередования: так, получается, в создании высокоскоростного переживания участвуют не только сами способы репрезентации событий, но и их ритмическое сочетание и чередование.

По-своему в дискуссию о показе и рассказе включается и нарратолог-классик Ж. Женетт. Показ в его представлении о повествовании включен в структуру рассказывания и отличается от «рассказа» тем, что сам факт рассказывания в таком случае предельно скрыт или редуцирован. «Показа» в терминах Джеймса для него как будто бы не существует: и в первом, и во втором случае речь идет о рассказывании, и без него, по Женетту, невозможна репрезентация событий в повествовании. О нарративной скорости Женетт потому говорит в ином контексте, когда размышляет о «длительности» событий. Позднее он признавал, что в этой части его знаменитой работы речь шла скорее именно о нарративной скорости: «...и теперь я думаю, что я должен был назвать эту главу не «Длительность», но «Скорость»» (Genette, 1983: 23).

Для понимания нарративной скорости в терминах Женетта принципиальным является различие между историей и дискурсом. Нарративная скорость, по Женетту, зависит от соотношения времени истории и времени повествования (Женетт, 1998: 117-140): так, «сцена» предполагает их сближение; «пауза» – остановку времени истории, при которой время повествования (наоборот) прогрессирует; «резюме» увеличивает скорость повествования за счет преобладания времени истории над временем повествования; а «эллипсис» и вовсе создает нарративную лауну, в которой события истории (фабулы) оказываются не представленными.

Женетт описывает функционирование нарративной скорости со структурной точки зрения, но проблема читательского переживания этой скорости им не обсуждается. Он, вместе с этим, указывает на проблему читательского внимания (в «резюме» события подаются быстрее и потому чаще оказываются незаметными для читателя), но читательское переживание он в целом не проблематизирует.

Поэтому позднее С. Паккард, развивая мысль Женетта, предлагает различать понятие скорости повествования и его ускорения (acceleration): последнее, подчеркивает он, имеет большее значение для читательского переживания, поскольку оказывается им в большей степени осязательно, в большей степени заметно для него (Packard, 2008). Нарративная скорость в таком случае – если продолжать размышление Паккарда – определяется не только соотношением времени истории и времени повествования, но и соотношением ускорения повествования и его замедления, оно-то и оказывает влияние на моделирование читательского опыта, связанного с переживанием высокоскоростного движения при взаимодействии с нарративом. Так – с опорой на Паккарда – можно говорить о том понимании нарративной скорости, которое включает в себя сразу два параметра: идею скорости как смены событий и идею ритма как частотного чередования ускорения и замедления.

Такое понимание читательского переживания, конечно, связывает нарративную скорость с проблемой читательского внимания, на что косвенно указывал и Женетт. Замедление повествовательного темпа позволяет читателю концентрироваться на представленных событиях, его ускорение – больше упускать, «не замечать», терять из виду. Это не значит, однако, что ускорение повествовательного темпа обязательно ведет к «утрате» той информации, которая в таких эпизодах представлена: скорее она формирует лауну, «слепую зону», к которой читатель может вернуться ретроспективно (и так нередко происходит – о чем пишут исследователи-когнитивисты (см. Vernaerts et al., 2013)). Проблема читательского переживания нарративной скорости, таким образом, связана с управлением его вниманием и тем движением по повествованию, которое этой скоростью моделируется: замедление захватывает внимание читателя «здесь и

сейчас» (в настоящий момент чтения), ускорение требует от него ретроспективного движения – возвращения к прежде уже представленным событиям.

Традицию Женетта в исследовании нарративной скорости продолжает и Б. Гингрич – в диахронической перспективе. Он (вслед за Женетом) определяет нарративную скорость как соотношение между «сценами» и «резюме» и исследует динамику этих способов репрезентации событий в романе XIX-XX века (Gingrich, 2021). Нарративная скорость – в таком определении – сближается с понятием ритма, ведь предполагает не только смену (событий, режимов репрезентации), их чередование, но и их частотное повторение – обеспечивающее, по Гингричу, создание скорости в повествовании. Работа Гингрича примечательна тем, как она позволяет по-новому пересмотреть историю романа как динамической литературной формы – теперь с точки зрения скорости. Но не избегает она и некоторых ограничений, связанных прежде всего с самим определением нарративной скорости. Несмотря на то что Гингрич (в начале книги) предлагает весьма широкое определение нарративной скорости («large-forward-rhythmic-shifting-dynamic-temporal narrative movement») (Gingrich, 2021: 10), на практике оно оказывается «суженным» до соотношения двух типов репрезентации, выделенных еще Женеттом, которые, к тому же, соотносимы с иной теоретической традицией, описывающей те же способы репрезентации как «показ» и «рассказ». В действительности работа Гингрича и посвящена исследованию соотношения этих двух способов репрезентации, тогда как концептуализация скорости в более объемном ее понимании скорее остается в стороне.

Однако нарративную скорость можно описывать не только через соотношение «показа» и «рассказа», «сцен» и «резюме».

Иной взгляд на создание нарративной скорости предлагает Фр. Моретти (Моретти, 2014) в исследовании классического реалистического романа, опираясь на структуру повествования, предложенную Р. Бартом.

Р. Барт писал о соотношении разных типов функций в структуре нарратива: среди них он выделял кардинальные (или ядерные) функции и функции-катализаторы (Барт, 1987: 397–398). Первые специфичны тем, что предполагают открытие альтернативной возможности развития действия, вторые – лишь сопровождают первые, являются дополнительными по отношению к ним. Такое различие типов событий в более поздней теории было воспринято Фр. Моретти для описания функционирования реалистического романа: кардинальные функции он назвал «поворотными моментами», а функции-катализаторы – «филлерами» (в русском переводе «fillers» переведены как «балласт») (Моретти, 2014: 102–103). Его подход отличен тем, что он рассматривает эти структурные элементы повествования с точки зрения их участия в создании ритма романа; фактически – не проговаривая это напрямую – Моретти рассматривает это соотношение, уделяя больше внимания тому, какое переживание для читателя создают разные типы событий и их чередование. Его интересует «ритм романа», и именно он позволяет проблематизировать миметические отношения между читателем и нарративом.

Говоря о классическом реалистическом романе, Моретти подчеркивает: филлеры в нем «наступают», тогда как «поворотные моменты» убывают, да и к тому же теряют свое значение, поскольку являются в значительной степени предсказуемыми для читателя. Преобладание филлеров замедляет продвижение действия, но и упорядочивает его, по-своему «успокаивая» читателя, – в этом ему

помогает и организованный в повествовании за счет повторений ритм.

Можно предположить, что иное соотношение поворотных моментов и филлеров – с преобладанием первых – позволяет, напротив, создавать ускорение в продвижении действия. Такое соотношение тоже может участвовать в создании и скоростного, и ритмического переживания для читателя.

Функции, о которых писал Барт, описывают действия героев, но повествование состоит не только из них. В создании скорости повествования потому участвуют и иные события, которые представляют собой теперь не действия героев, а смену их ментальных состояний. Об участии в нарративной прогрессии таких повествовательных элементов писала еще М.-Л. Райан (Ryan, 1986). Репрезентацию или реконструкцию намерений, желаний, фантазий, предположений и пр. героев она называла «встроенными» нарративами, подчеркивая, что в сюжетной динамике они играют роль не меньшую, чем те действия, которые в действительности совершаются героями. Не меньшее значение они имеют и для читателя – с точки зрения его переживания и соотнесения себя с персонажами. Читатель соотносит себя не только с тем, что в действительности происходит с героем, но и с тем, что только могло бы с ним произойти, но осталось нереализованным (об этом пишет, например, Р. Герриг (Gerrig, 1993)). События, относящиеся к действительности повествовательного мира, Райан предлагает называть «актуальными»; события, являющиеся возможными, но нереализованными, она описывает как «виртуальные» (Ryan, 1991: 148–175).

Но какое значение соотношение этих событий имеет для создания нарративной скорости? На эту особенность событийности в нарративе указывает Дж. Принс, продолжающий размышление Райан о «несвершившемся» в повествовательном мире. Такие

нарративные элементы он называет «дизнаррацией» и подчеркивает их участие в производстве эффекта реалистичности происходящего: указание на то, что могло бы произойти, но все-таки не свершилось, говорит он, создает контраст для действительно произошедшего, чем усиливает правдоподобность событий (Prince, 1988: 4). Но он же отмечает и роль таких элементов в создании нарративной скорости: дизнаррация, обращает внимание Принс, замедляет ход повествования, поскольку, можно предположить, уводит читателя из зоны «актуального» в зону «виртуального», смещает акцент со смены действительных событий на то, что в действительности повествовательного мира вовсе не происходит. Такой «расклад», по-видимому, справедлив для реалистической прозы, в которой дизнаррация занимает положение служебное и дополнительное – фактически подчиненное по отношению к актуальным событиям, находящимся в реалистической прозе на первом плане. В применении к более поздним нарративам (в том числе некоторым типам современных) участие виртуальных событий в создании ускорения и замедления повествования может быть пересмотрено: хотя бы потому, что в прозе постмодернизма и постпостмодернизма виртуальные события нередко начинают преобладать над актуальными, чем нарушается необходимый для создания скорости контраст. На это отчасти указывал и Принс, говоря об изменении функций дизнаррации в более поздней литературной традиции, которые еще предстоит (исследователям) проследить. Но и в реалистической прозе соотношение актуального и виртуального, если рассматривать его в читательской перспективе, не всегда поддается четкому разграничению, ведущему к «простому» определению механизмов создания ускорения и замедления. Так, например, в романах Дж. Остин (как показывает исследовательница Х. Данненберг

(Dannenberg, 2005: 159)) виртуальные события (предположения, иллюзии, ложные представления) нередко подаются читателю сначала как в действительности происходящие и разоблачаются позже – по ходу развертывания сюжета. Это значит, что с точки зрения читателя граница между актуальным и виртуальным в повествовании оказывается куда более зыбкой (виртуальное «притворяется» актуальным и выявляется как таковое лишь позднее), а роль виртуального в создании ускорения и замедления – требующей более комплексного описания.

Теоретические размышления о нарративной скорости, представленные ранее, опирались прежде всего на работу с классическим реалистическим романом; к современным же литературным формам исследователи этого феномена по-прежнему обращаются редко. Но такие попытки все-таки предпринимаются, среди них – наиболее заметная и влиятельная работа К. Хьюм, сначала вышедшая в виде отдельной статьи (Hume, 2005), затем дополненная (в соавторстве с Дж. Бейтенсом) (Baetens, Hume, 2006), после чего она же была включена в ее более объемный труд, посвященный современному американскому роману, где исследования нарративной скорости оказались вписаны в общую систему концептуализации так называемой (в ее терминах) «агрессивной литературы» (Hume, 2012). Хьюм выделяет несколько подходов к описанию нарративной скорости: она может быть рассмотрена с точки зрения репрезентации физической скорости; с точки зрения нарративной ретардации (традиция формалистов); в соотношении количества событий, представленных в повествовании, и того текстового объема, которые они занимают (традиция Женетта); и, наконец, скорость может рассматриваться в ее непосредственном обсуждении (метарефлексии) в литературных текстах.

Репрезентация физической скорости при этом, подчеркивает Хьюм, совсем

необязательно моделирует высокоскоростное переживание для читателя, за создание которого отвечает нарративная скорость. Нарративная скорость потому отлична от изображения движения или перемещения героя, хотя это переживание (героя и, напротив, читателя) могут и совпадать.

Изображение высокоскоростного движения героя, вместе с этим, тоже имеет значение, но уже не с точки зрения переживания скоростного, сколько с точки зрения ритмического. Изображение такого движения способно устанавливать телесно-миметические отношения между героем и читателем/зрителем, интенсифицировать непосредственность такого переживания. Так нередко происходит и в других литературных и кинематографических примерах, которые Хьюм не рассматривает.

В дополнение к предшественникам Хьюм выделяет и анализирует иные средства создания нарративной скорости. Среди них – «умножение» (multiplication) и «вычитание» (subtraction). Первый способ создания скоростного переживания опирается на быструю смену героев, событий, локаций, точек зрения и пр. Такой способ репрезентации, по сути, имитирует монтажный принцип, позволяющий соединять разнородные элементы, при этом подчеркивая их разность и зазор между ними. Одного нагромождения предметов, героев, событий, однако, оказывается недостаточно – подчеркивает Хьюм (Hume, 2005: 111). В создании скоростного переживания задействован потому и второй принцип, который позволяет «вычесть» из повествования те элементы, которые бы отвечали за связность перечисляемых героев и событий. Само по себе перечисление – настаивает Хьюм – может сделать повествование лишь объемнее, но не «быстрее». Скорость ему задают отсутствующие «связки», выполняющие – если продолжать ее размышление – функцию монтажных

швов. Подчеркнутая разнородность повествования, его переключение от одного режима к другому (на самых разных уровнях) потому способствует увеличению нарративной скорости, но к тому же и управляет вниманием читателя: более «медленные» нарративные эпизоды привлекают этого внимания гораздо больше, тогда как «скоростные» производят эффект беспорядка, озадачивают читателя, вынуждают его переживать бессвязность предлагаемого ему повествовательного мира.

Подобные нарративные приемы, по Хьюм, способны ввести читателя в переживание дезориентации, потерянности, уязвимости, способны создавать ощущение утраты контроля, заставляют его пасовать перед наступающим нарративным хаосом.

Резюмируя существующие подходы к определению нарративной скорости, исследовательница К. Кукконен отмечает: в ее создании участвуют приемы, относящиеся к разным уровням организации повествования. Она предлагает потому выделять несколько «скоростей» в нарративе, среди которых – «storyworld speed» и «discourse speed», первая создается движением героя по сюжету, вторая – движением читателя по тексту (она определяется «скоростью, с которой текст предположительно должен быть прочитан, и включает в себя стилистическую плавность <...>, а также особенности публикации» (Kukkonen, 2020: 8)).

Развивая наблюдение Кукконен, можно заключить, что в создании высокоскоростного переживания для читателя задействованы, по-видимому, разные уровни организации повествования, каждый из которых вносит свой вклад в моделирование такого переживания.

Нарративная скорость, таким образом, представляется зонтичным понятием, которое объединяет сразу несколько феноменов: объем фабульных

событий; скорость их репрезентации на сюжетном уровне; дискурсивную скорость; репрезентацию физического движения (героя); переживание читателем высокоскоростного движения.

На структурном уровне скорость повествования, таким образом, определяется сразу несколькими способами: в ее создании участвует, например, чередование «показа» и «рассказа»; кроме того – эллипсисы и резюме ускоряют продвижение читателя по сюжету, паузы и сцены – его тормозят. Вместе с этим, поворотные моменты способствуют наращиванию темпа, а филлеры – его замедлению. К тому же, дизнаррация, по Принсу, тоже снижает скорость репрезентации. Встроенные нарративы – в их понимании М.-Л. Райан – указывают читателю на виртуальные события в повествовании, и своим отличием от актуальных тоже замедляют продвижение действия.

В прагматической перспективе принципиальным в понимании нарративной скорости оказывается ее переживание читателем: и потому с этой точки зрения нарративная скорость понимается как обязательное чередование ускорения и замедления, которое может создаваться прежде отмеченными приемами, но также и сочетанием «умножения» и «вычитания», рассмотренными прежде Хьюм.

Но можно ли описывать повествовательную скорость иначе? Ответ на этот вопрос предлагают исследователи энактивистского направления постклассической нарратологии, вновь возвращающиеся к этой проблематике и переосмысляющие ее: они исследуют связь между конкретными структурными элементами, способствующими созданию скорости и замедления в повествовании, и теми эффектами, что эти элементы могут производить на читателя.

Энактивистская нарратологическая традиция рассматривает нарративную скорость иначе, чем это делали ее

предшественники: теперь – в контексте воплощенного читательского взаимодействия с повествовательным миром. Это взаимодействие понимается исследователями как моделирование телесного движения (внутри повествовательного мира), сопровождаемое телесными и аффективными реакциями читателя. Такое интерсубъективное взаимодействие («воплощенное действие», «воплощенная фабула» (Caracciolo, 2014b; Caracciolo, Kukkonen, 2021; Hven, 2017; Hven, 2022)), согласно исследователям, может создаваться при помощи репрезентации в тексте опыта восприятия героя: в таком случае задействуется внутренняя фокализация, текстовые указатели (cues), позволяющие читателю реконструировать физическое движение героя и «включиться» в него в собственном воображении. Репрезентацией опыта восприятия и движения, однако, в такой перспективе все не может ограничиваться, поэтому не меньшее внимание исследователи уделяют тем измерениям повествования, которые способны создавать менее опосредованный контакт читателя/зрителя с происходящим. К ним относится аффективное измерение нарратива, скоростной и ритмический аспект повествования. Нарративная скорость в таком случае имеет значение тогда, когда способна ощущаться и переживаться читателем, когда создает (в его воображении) скоростное переживание, когда предполагает соответствующий телесный ответ. Нарратологи-энактивисты нередко не различают понятие скорости и ритма, фактически их уравнивая (Caracciolo, 2020), хотя это различие может быть продуктивным и для описания читательского взаимодействия с повествованием в их перспективе. Обращаясь к конкретным примерам (фильм «Беги, Лола, беги» Т. Тыквера или новелла «Сердце-обличитель» Э. По) (Caracciolo, 2014a; Hven, 2017), они

показывают, как создаваемый на уровне репрезентации ритм устанавливает телесно-миметические отношения читателя с повествовательным миром, как сокращает дистанцию по отношению к нему, как способствует имитации непосредственного контакта и переживания.

Нарративная скорость, рассматриваемая теперь с точки зрения ее способности быть переживаемой читателем, ставит перед исследователями и новые вопросы: например, каким образом (в этой перспективе) может быть охарактеризовано читательское взаимодействие с героем (и его движением), а каким – со скоростью и ритмом самого повествования? В чем отличие между этими двумя типами воплощенного взаимодействия и как они используются в конкретных примерах?

Ритм в повествовании, как подчеркивают М. Карачоло и К. Кукконен (Caracciolo, Kukkonen, 2021: 142-143), создается разного рода повторениями: эти повторения в свою очередь организуют (в их терминологии) «поток» (flow), в который включается читатель. Такие повторения связаны с проблематикой нарративных ожиданий и предсказуемости, создаваемой отдельными нарративными элементами. Предсказуемость, как следует из размышления, один из значимых эффектов ритмической организации нарратива; она обеспечивает его связность, она же способствует погружению и концентрации (необходимой для погружения) внимания. Ритм в энактивистской перспективе рассматривается не столько как феномен нарративной формы, сколько как феномен, характеризующий процесс взаимодействия читателя с текстом. Он значим не со структурной точки зрения, но с прагматической: в том, какое переживание моделирует для читателя, как управляет его вниманием, какие телесные реакции предполагает и как позволяет «включить» его в повествовательный мир,

способствует его «вдействию» и «вчувствованию». Потому Я. Поповой, например, ритм определяется через взаимодействие между нарратором и читателем (Ророва, 2015: 83): он выходит за пределы только формальной организации повествования, пересекает границу между нарратором и читателем и участвует теперь в организации воплощенного взаимодействия читателя с повествовательным миром. Как – при этом – он соотносится с нарративной скоростью, как участвует в создании зрительно-миметических и, напротив, телесно-миметических отношений, исследователями пока не проясняется.

Развивая идеи, эксплицированные в работах Карачоло, Кукконен, Хвена и Поповой, можно противопоставить понятия нарративной скорости и нарративного ритма со следующих точек зрения: если ритм создает в повествовании «поток», вовлекает в него читателя за счет повторения на разных уровнях, «внушает» тем самым ему предсказуемость движения и развития событий, то скорость отвечает за создание иного и «противовесного» переживания – дестабилизирующего, заставляющего что-то в происходящем неизбежно упускать, а потому связана с переживанием непредсказуемости и дезориентации. При этом функционирование нарративной скорости опирается на принцип чередования (ускорение/замедление, показ/рассказ, сцена/резюме, актуальное/виртуальное) и потому предполагает взаимодействие с ритмической организацией повествования. Высокоскоростное переживание, создаваемое в нарративе, потому может интенсифицироваться задаваемым в нем ритмом.

Современные разветвленные нарративы: проблематизация нарративной скорости

Исследователи нарративной скорости чаще обращаются в ее рассмотрении к классической литературе, но эта проблематика оказывается более чем

актуальной для изменений, происходящих в современных нарративах.

Они нередко характеризуются как «усложненные», предполагая такую организацию повествования, которая озадачивает, дестабилизирует, дезориентирует читателя: в них несколько сюжетных линий смешиваются между собой, не позволяя читателю отличать одно время от другого, одну модальность – от иной. Такие современные нарративы потому получают название и нарративов-пазлов, и лабиринтов, и сетей, и даже – хаотических. Среди них развитие получают и разветвленные нарративы, которые предлагают читателю теперь следить не за одной жизнью героя, а за ее несколькими версиями, каждая из которых предъясняется ему последовательно или поочередно.

Такие нарративы получают распространение в литературе и кино последних десятилетий, и вариации в них проблематизируют соотношение актуального и виртуального в нарративе. Они ставят акцент на том, что только могло бы произойти с героем, избегая репрезентации того, что с ним в действительности произошло. Таким образом устроено повествование в фильмах К. Кесьлевского, Т. Тыквера, Д. Джонса; эта же нарративная модель задействована в романах последнего десятилетия – П. Остера, Ф. Рота, К. Аткинсон, Б. Жиро. Разветвленные нарративы оказываются востребованы в разных нарративных медиа и в разных национальных традициях (американской, британской, французской), но едины в том переживании, которое предлагают читателю и зрителю. И в литературе, и в кино, заодно с проблематизацией актуального и виртуального, наррации и диннаррации, они моделируют переживание высокой скорости происходящих событий, но добиваются этого эффекта разными способами.

В результате читатель оказывается вовлечен в переживание

высокоскоростного движения: разветвленные нарративы предлагают ему опыт погони, и указание на него мы нередко замечаем уже в названиях произведений. Так называется роман Б. Жиро – «Жить быстро», так назван и фильм Т. Тыквера – «Беги, Лола, беги», который в русском переводе обретает дополнительный ритмический импульс (за счет повторения императива «беги», который в оригинале – *Lola rennt* – отсутствует).

Как структурные особенности таких нарративов позволяют им моделировать опыт погони для читателя и зрителя? Как разные аспекты повествовательной организации участвуют в создании такого опыта? И как он может быть описан – какие эффекты для него характерны?

Мы рассмотрим проблематизацию нарративной скорости и ритма на отдельных примерах из нескольких романов и фильмов, относящихся к разветвленному повествованию: среди них – «Случай» (1981) К. Кесьлевского, «Беги, Лола, беги» (1998) Т. Тыквера, «4321» (2017) П. Остера, «Жить быстро» (2022) Б. Жиро.

Создание такого переживания происходит при помощи нескольких приемов. Так, в моделировании переживания высокоскоростного движения задействована репрезентация физического движения героя; преобладание поворотных моментов над филлерами в структуре повествования поддерживает такое переживание; оно усилено резюме и эллипсисами, превалирующими над сценами и резюме; и дополнено сокращением дикшнации. Кроме того, в создании такого переживания участвуют и ритмическая организация повествования – в совокупности все эти приемы позволяют читателю устанавливать телесно-миметические отношения с повествовательным миром.

Blow up: принцип «умножения»

Существенное преобладание времени истории над временем повествования

ощутимо уже с первых страниц романа «4321» Пола Остера: перед нами – рассказ о нескольких поколениях семьи героя, который, занимая немного времени повествования, включает в себя сразу несколько десятилетий на уровне времени истории. Такая разность двух повествовательных времен с самого начала задает высокую скорость разворачивающимся событиям, но только этой разности оказывается недостаточно для создания высокоскоростного переживания у читателя, поэтому она дополняется другими способами создания нарративной скорости, усиливая заданный вначале эффект.

Высокая скорость в данном эпизоде создается и множественным перечислением – тем, о котором писала Хьюм, называя такой прием «умножением». Рассказ о семье главного героя специфичен и большим количеством «участников» этих событий (всех родственников в таком объеме едва ли можно запомнить!), и резким переходом от одной семейной линии к другой, от одного персонажа – к иному. Эффект ускорения, создаваемый таким обильным и динамичным перечислением, только усиливается благодаря тому, что переход от одного персонажа к иному не кажется мотивированным: не прослеживаются связи между главным героем (а ведь речь в романе будет идти именно о нем) и его родственниками, для читателя не моделируются ожидания того, какую роль эти родственники будут играть в жизни главного героя в дальнейшем развитии действия – это «умножение» потому характеризуется и тем «вычитанием», о котором писала Хьюм и которое мы обнаруживаем у Остера, имитирующим монтажный принцип репрезентации событий при помощи резких переходов от одного «кадра» (или изображения) к другому. Эффект ускорения интенсифицируется и благодаря явной избыточности такого перечисления и рассказа: всего на нескольких страницах

романа располагается больше десятка персонажей и рассказов о них.

Высокоскоростное переживание в разветвленном нарративе усилено и тем преобладанием времени истории над временем повествования, которое задается самой нарративной конструкцией: разветвленный нарратив всегда содержит в себе не одну фабулу, а несколько (Richardson, 2019: 8); он состоит из нескольких версий жизни героя, а потому «переполнен» событиями не только на уровне наррации, но и – по умолчанию – на фабульном уровне. В разветвленном нарративе сочетается сразу несколько приемов, создающих «избыточность» в повествовании: множество фабульных линий, обильное перечисление, прием «умножения». Они не только участвуют в создании ускорения повествования, но и могут быть описаны при помощи метафоры, которую предложил М. Антониони в фильме «Фотоувеличение». Его фильм в оригинале назывался «Blow up», что обозначает «вздутие», «разбухание», «взрыв». Подобный эффект переполнения и вздутия производится и в разветвленном нарративе – хоть и остается дополнительным по отношению к высокоскоростному переживанию.

Наперегонки: репрезентация физического движения

Такая работа с повествовательным временем и со способом репрезентации событий дополняется репрезентацией физического действия, что особенно заметно в фильмах, построенных по разветвленной модели, которые, по собственному признанию Остера, являются предтечей романа. В фильме Кесьлевского, например, каждая новая вариация начинается бегом героя по перрону – камера сближается с ним, попутно приближая к его движению и зрителя. Но приближение это не только физическое: бег героя по перрону – решающий, он во многом определяет его дальнейшую судьбу, и для зрителя это

становится совершенно ясно, поскольку этот бег повторяется несколько раз за фильм. Такое повторение имеет не только интенсифицирующий ритмический эффект, оно моделирует ожидания зрителя, указывает ему на скорое наступление решающего момента, и потому приближает его к герою еще сильнее – теперь на уровне нарративной интриги, напряжения, созданного прогнозированием дальнейшего развития событий. В фильме Тыквера «Бег» и вовсе перестает быть метафорой: героиня бежит весь фильм, и камера упрямо следует за ней, не давая зрителю увернуться от такого «странного сближения» с героиней. Кажется, она и вовсе не способна на остановку, а зрителю почти не дают возможности «посмотреть по сторонам»: выбраться из скоростного потока, заданного репрезентацией ее движения. В фильме Кесьлевского скорость задается и за счет присутствия поездов в фильме: они, как показывал прежде исследователь нарративной скорости Д. Белл (Bell, 2004), и тематизируют скорость, но и подчеркивают стремительное перемещение героя в пространстве, которое к тому же будет дополнено и перемещением во времени – герой «без труда» с каждой новой вариацией возвращается во времени в исходную точку, перенося в этот исходный момент и зрителя, только усиливая это скоростное переживание, охватывающее и воображаемое движение зрителя в пространстве, но и во времени. То же происходит и в «Лоле»: она бежит, пока все вокруг движется в привычном (не скоростном) темпе, и этот контраст и становится тем «ускорением», значимым на фоне «замедления». Ее бег визуально дополняется мелькающими на экране поездами, вездесущими автомобилями, но и аудиовизуальный монтаж не уступает по скорости движению героини: резкая смена углов съемки в изображении ее действия и пространства, дополненная быстрой музыкой, не оставляют зрителю шанса вырваться из этого временного потока,

отвлечься, уклониться от миметического взаимодействия с ней.

Роман Остера тоже наполнен репрезентацией активного физического действия: нередко читатель оказывается «вместе» с героем в поездах, автомобилях; «участвует» с ним в спортивных матчах – герой увлечен бейсболом; он вовлечен и в шествия-митинги, и в перелеты через Атлантику. Значимо в данном случае не только упоминание таких действий героя, но и то, как они на глазах читателя становятся полноценными событиями, как с репрезентацией физического действия у Остера нередко связаны «поворотные моменты» в судьбе героя. Так, герой попадает в аварию, мчась на автомобиле, – она вынуждает его провести немало времени в больнице, но и спасает его жизнь: благодаря полученным увечьям его не призывают в армию, а он не оказывается во Вьетнаме. Увлечение бейсболом и изображение активного скоростного движения героя на соревнованиях связано с другим «поворотным моментом»: однажды на одном из матчей гибнет его друг, внезапно и скоростно, никем не предсказуемо. Такое событие – не только резкая остановка действия в романе, не только «контраст» для прежде изображаемого высокоскоростного движения. Оно – и трагический (то есть оркестрированный аффективно), и поворотный момент для самого героя: после этого события он навсегда оставит бейсбол, и его дальнейшая профессиональная жизнь сложится иначе. Так происходит и в других случаях: шествия-митинги – не «фон» в повествовании, но полноценная «фигура». Герой участвует в них, потому что тяжело переживает гибель Дж. Ф. Кеннеди, а его заочное влияние на него становится во многом определяющим для становления самого героя.

Но бежит не только герой и не только событийное время в романе, но и время календарное: скоростное движение, моделируемое для читателя, усиливается

датировками – эксплицитным указанием на время и его ход в повествовании. Несколько глав романа Остера содержат такие указатели: они делают эксплицитной временную прогрессию в повествовании, «напоминают» читателю о ней, производят потому эффект «листов календаря». Преобладание фабульной событийности потому становится дополнительно эксплицитно для читателя – при помощи перечисляемых в романе дат. Но подобные датировки, регулярно встречающиеся читателя в тексте, наделяются и ритмическим эффектом: регулярно повторяясь, чередуясь с нарративным ускорением, они позволяют читателю соотнести физическое (скоростное) движение героя с ускоренным бегом времени – таким образом и самому включиться в это высокоскоростное переживание.

«Поворотные моменты» обгоняют «филлеры»

Соотношение поворотных моментов и филлеров, которое, по словам Моретти, позволяет «логике рационализации» проникнуть в сам ритм классического реалистического романа (Моретти, 2014: 118), оказывается значимым и для разветвленного нарратива. Только само это соотношение меняется принципиально: поскольку, как мы видели прежде, время истории в нем значительно превосходит время повествования, а сама фабула множится, превращаясь в многолинейную, кардинальная событийность в ней тоже преобладает – и существенно теснит «филлеры». Роман Остера переполнен событиями, которые, как писал Барт о кардинальных функциях, могли бы произойти иначе. Большинство событий, происходящих с ним, имеет либо мгновенный, либо отложенный эффект – и такое переполнение повествования кардинальной событийностью тоже участвует в создании высокоскоростного переживания, но производит и другие эффекты. Если множество событий могут изменить судьбу героя, стать поворотными

моментами, то такое их скопление требует от читателя предельной концентрации – внимания к происходящему. Оно погружает его в состояние, в котором отвлечение и расслабление невозможно, ведь каждый момент может стать поворотным, принципиально изменить фабулу. Так происходит и на уровне общей структуры романа: никогда не знаешь, когда одна линия прервется и начнется следующая. Так происходит и в отдельных эпизодах романа: повторное замужество матери главного героя кажется лишь фоном по отношению к его жизни, оно – на первый взгляд – мало что меняет в его собственном пути; но вскоре выясняется, что эта встреча (с отчимом) станет для героя определяющей: он сам, вслед за ним, изберет путь музыканта, что без этой встречи было бы невозможно. Так происходит и со встречами, которыми полнится роман, и с совпадениями (неожиданными пересечениями), и с событиями, которые только на первый взгляд кажутся дополнительными, сопровождающими, но очень быстро открывают свою истинную, «поворотную» природу. Мир разветвленного повествования становится потому мало предсказуемым для читателя: и в создании этого переживания участвует и сама событийность, но и та скорость репрезентации событий, которая ее характеризует. Хьюм, размышляя о современном американском романе, писала, что подобная акселерация производит скорее дезориентирующий эффект на читателя: нельзя ни предсказать, когда одна линия прервется, ни угадать, какое из событий станет поворотным, а какое – останется филлером. В романе Остера такое устройство событийности требует от читателя быть все время «на чеку», но оно же и смеется над ним, не позволяя строить хоть сколько-нибудь оправдывающиеся прогнозы. Эта непредсказуемость, конечно, только усиливается за счет вариативности пути героя и за счет той особенности

скоростной репрезентации, которая предполагает, что удержать во внимание все обилие изображенного оказывается невозможно.

Резюме и эллипсисы преобладают над паузами и сценами

Б. Гингрич убедительно показал, как ритмичное соотношение между «резюме» и «сценами» управляет нарративной скоростью классического реалистического романа. Но что происходит в современных разветвленных нарративных формах?

Если рассматривать эту повествовательную форму с точки зрения тех режимов репрезентации событий, которые выделял и описывал Женетт, станет явным системное преобладание в них не только резюме над сценами, но и эллипсисов – над паузами. Во всех рассматриваемых нами примерах действие в репрезентации героев оказывается важнее разного рода «бездействия» (диалогов, описаний, размышлений). Не вызывает сомнений потому высокая скорость репрезентации событий: время фабулы существенно превосходит время повествования, сцены и паузы редуцированы. Такой способ репрезентации событий, конечно, влияет на взаимодействие читателя с повествованием: он соотносит себя с постоянно действующим субъектом (героем), он не отвлекается ни на описания, ни на обильные диалоги, он стремительно продвигается «по времени прочь».

Но вместе с этим привычные читателю по реалистическому роману сцены и резюме меняют свои функции. Они не только перераспределяют свое соотношение с точки зрения объема в повествовании; они – теперь – участвуют в создании иного читательского переживания, наделяются новыми с этой точки зрения функциями. Прежде, как отмечал еще Женетт, сцены позволяли создать эффект драматизации, привлекали и концентрировали внимание читателя, тогда как резюме, как скороговоркой,

представляли читателю все то, что, по мысли Женетта, вероятнее всего останется для него в «слепой зоне», то есть будет пропущено (Женетт, 1998: 139). Разветвленные нарративы, наполняясь резюме и эллипсисами, переопределяют их функции: теперь внимания и концентрации требуют резюме, эллипсисы (как нарративные лакуны) тоже требуют внимания, поскольку становятся непостоянными; сцены же – напротив – убывая, создают контраст для резюме, выступают в роли небольшой передышки для читателя, прежде чем он вновь сможет погрузиться в высокоскоростное переживание представленных в резюме событий.

Своеобразное обнажение этого приема мы наблюдаем в фильме Тыквера, когда нам несколько раз за фильм показывают, как сложится судьба второстепенных персонажей. Их судьба уместается в несколько фотоснимков, показанных зрителю один за одним. Перед нами – концентрированное резюме, лишенное всякой «примеси»; оно занимает экстремально краткое время повествования в его сравнении со временем истории. Эффект резюме усиливается и тем, что события показываются при помощи фотографии: это не поток, а выхваченные моменты из жизни героев, связь между которыми намеренно опускается, создавая, разумеется, и ритмичное, и скоростное движение, но – одновременно с ними – и подчеркнутую дискретность повествования, демонстрирующую (фактически) его монтажные швы.

«Виртуальное» теснит «актуальное»

Ускорение в разветвленных нарративах всегда соседствует с замедлением, которое к тому же принимает форму резкого торможения, непредсказуемой и немотивированной остановки движения и действия. Внезапно свой бег прерывает Лола, неожиданно (для зрителя) она стремительным жестом подбрасывает в воздух предметы – и они

останавливаются, зависают в воздухе, как по мановению волшебной палочки. Остановки физического действия дополняются и сюжетными «обрывами»: читателя переключают от одной сюжетной линии к другой, и когда произойдет это переключение, ему неизвестно, предсказать это нельзя. Набранная сюжетная скорость сменяется резким торможением, а затем – таким же резким переключением на уже другую жизнь героя, к тому же начинающуюся заново в исходной точке (в прошлом по отношению к моменту настоящего, в котором действие останавливается). Разветвленная форма сама по себе структурно предполагает такие «обрывы», как предполагает и скачки во времени: но в романе Остера эта структурная особенность только усиливается за счет большего дробления сюжетных эпизодов. Читателя не только переключают от одной жизни к другой (как происходит у Кесьлевского, например), от одной попытки к иной (как в «Беги, Лола, беги»), его заставляют «перепрыгивать» от эпизода из одной жизни к эпизоду из другой, и так – множество раз подряд в романе. Такое неизбежное движение читателя по предлагаемой ему событийности – постоянные скачки во времени, резкие остановки, невозможность продолжать движение в потоке одной сюжетной линии – сопровождается и другими способами замедлить, а иногда и прервать повествование, создав необходимый для ускорения контраст.

Так происходит с использованием дизнаррации: в отличие от классического реалистического романа, о котором писал в этом контексте Принс, дизнаррация в романе Остера встречается гораздо реже. Но указания на то, что только могло бы произойти, но все-таки не свершилось, в романе присутствуют. Их существенно меньше, чем в реалистическом романе, но дело не только в их объеме, но и в измененных функциях. Убывая, дизнаррация в высокоскоростном повествовании производит эффект

«резкого торможения»; контраст, который она создает, только усиливает чередование ускорения и замедления, тем самым интенсифицируя переживание, предложенное читателю.

Сокращение дизнаррации в таком типе нарратива связано и с тем, что отчасти ее функции на себя берет повествовательная форма в целом: дизнаррация открывала читателю доступ к «виртуальным» событиям, в их отличии от «актуальных». Разветвленные нарративы больше не нуждаются в ней, поскольку в целом опираются прежде всего на «виртуальную» событийность: они предъявляют читателю только возможные версии произошедшего, не проговаривая то, что с героем произошло на самом деле. Конструкция разветвленного нарратива потому предполагает выдвижение на первый план сразу нескольких событийных линий, каждая из которых открывает читателю альтернативный возможный мир, состоящий из виртуальных, но не актуальных событий.

Виртуальная событийность в разветвленном повествовании преобладает, но от этого – парадоксальным образом – не утрачивает способность создавать высокоскоростное переживание: в противопоставление тому, как Райан, размышляя о соотношении виртуального и актуального и о значении этого соотношения для создания нарративной скорости, описывала реалистический роман.

В разветвленном нарративе «виртуальное» существенно теснит «актуальное», но – за счет иных повествовательных приемов – позволяет создавать для читателя переживание высокоскоростного движения. Так, получается, прием, позволявший создавать ускорение и замедление в реалистической прозе, переосмысливается и «переворачивается» в своем эффекте в современном разветвленном нарративе.

Ритм интенсифицирует высокоскоростное переживание

В моделировании читательского переживания в разветвленных нарративах приемы, обеспечивающие ускорение повествования, соседствуют с теми, что отвечают за производство ритма, дополняют и усиливают друг друга.

При всем их отличии – о котором шла речь прежде – без их комплиментарного взаимодействия обеспечение телесно-миметического контакта читателя с таким повествовательным миром представляется невозможным.

Ритмические приемы наполняют разветвленные повествования, начиная уже с ее общей нарративной структуры. Разветвленные нарративы предлагают читателю постоянное возвращение к исходной точке, которое в какой-то момент развития действия становится для него и предсказуемым, и ожидаемым: он движется от одной версии к другой, от одного «детства» героя к другому, от одного «если» – к иному. Этот ритмичный ход разветвленного повествования создает одновременно и предсказуемость его структуры, но и ее дискретность, производящую эффект «разрыва», постоянного «прерывания», нестабильности. Ритмичное чередование сюжетных линий (версий) поддерживается в романе «Жить быстро» Б. Жиро иным (графическим) повторением: каждая новая глава начинается с нового «если», и эти «если» – и есть тот счет, который главная героиня ведет в попытке проработать травматическое переживание, связанное с гибелью мужа, и есть та попытка упорядочивания и рационализации, спасающая ее, по-видимому, от неконтролируемого хаоса переживания. То же происходит и в романе Остера. Названия глав сменяются: 1.1., 1.2., 1.3., 1.4., затем – 2.1., 2.2., 2.3., 2.4. и т.д., сочетая таким образом и повторения, и различия (все главы в его романе обозначены только цифрами). Этот ритмический эффект и в одном, и в другом романе поддержан синтаксическими периодами, а в романе Остера – еще и

имитацией дневникового жанра, который содержит последовательно предъявляемые читателю даты событий (они возникают через равный промежуток «текстуального» времени). Аналогичный способ создания ритма встречается и в «Беги, Лола, беги»: в фильме он только усиливается за счет взаимодействия нарративной канвы и ее визуальной и аудиальной репрезентации. Ритм создается битами музыки, шумами (биение сердца, звук часов), визуальной репрезентацией бегущих по циферблату стрелок, соотношением эпизодов – каждому из них (как и каждой попытке Лолы) отведено лишь 20 минут, и эти 20 минут повторяются вновь и вновь. В «Лоле» ритмическая организация всего повествования производит потому утроенное воздействие: изображение подкрепляется звуком, их работа усиливается ритмом самого повествования.

«Буйные», «агрессивные», «экстремальные»: нарративная скорость в современном культурном контексте

Современные разветвленные повествования включены в более объемный комплекс сложных нарративов, отличающихся проблематизацией временных отношений в повествовании и предлагающих потому «усложненный» опыт читателю и зрителю. При их описании исследователи щедро сыплют метафорами: буйные, непокорные, экстремальные, агрессивные, озадачивающие, стремительные, хаотические... (см., например: Bordwell, 2023; Richardson, 2006). Скорость и ритм в таких нарративах, как мы видели, участвуют в создании такого переживания, но и по-своему пересекаются с теми характеристиками, которыми наделяется современная цифровая культура.

Так, С. Шавиро подчеркивает: для цифрового кино характерна эстетика постконтинуальности, с акцентом на дискретности, нелинейности, пространственно-временных разрывах, с эффектами дезориентации, «хаоса»

(Shaviro, 2016). На создание таких эффектов направлены аудиовизуальные приемы: резкая смена углов съемки, стремительное движение камеры, переключение от одной к другой, звуковые и визуальные шоки. Все это в результате включается в понятие ультракинетического монтажа, характеризующего цифровое кино, который выдвигает на первый план, как следует из его именованья, особую, экстремальную подвижность изображения и звука, их подчас усложненные отношения. Ультракинетический монтаж характерен для такого кино, но такая особенность цифрового фильма не может не влиять и на другие нарративные медиа: и с этой задачей, кажется, успешно справляется, если судить об этом влиянии по современным разветвленным нарративам.

Понятие скорости в его применении к современной цифровой культуре обсуждается и в работах П. Вирильо (Virillio, 1989), и Б. Стиглера (Стиглер, 2019). Скорость у Вирильо и Стиглера в том числе обеспечивает ту эстетику исчезновения и забывания, которую они описывают в качестве побочного, но неизбежного эффекта от нагнетания разного рода образности. Так происходит, как мы видели, и в разветвленных нарративах, в которых «избыточность» неизбежно соседствует с «лакунарностью», «ускорение» – с «резким торможением», читательская концентрация на происходящем – с неизбежным пропуском той или иной информации, остающейся для него (благодаря повествовательным приемам) в «слепой зоне».

Но скорость – понятие, описывающее не только режим предъявления этих образов (в кино ли или в литературе), но и режим их восприятия. Потому Стиглер в своем размышлении приходит к описанию темпоральности чтения – в попытке представить, как разные режимы чтения (в современной культуре) соотносятся между собой, как они меняются и даже

взаимобогащаются. Среди них – чтение как восприятие информации (новостной и повседневной) и чтение художественного произведения. Первое предполагает высокую скорость, мгновенное схватывание, центрирование на одном событии, но и стремительное отбрасывание этого события, обеспечивающее исчезание, в пользу другого, нового. Стиглер написал статью «Время чтения» еще в 1990 году, но из 2024 года мы без труда назовем описываемый им режим чтения скроллингом – знакомым всякому читателю. Отчасти такое чтение, как мы помним по Р. Барту, подготавливалось и литературой, в основном популярной, с ее попыткой моделировать для читателя «свободный ритм чтения», «чтения для раскрытия сюжетной разгадки» (Барт, 1989: 468-469). Это чтение «на бегу» – поскорее преодолеть все страницы, чтобы дойти до конца и до искомого ответа – поддерживалось и теми практиками, которые все больше и больше ориентировали читателя на восприятие текста на буквальном «ходу»: в поезде, в самолете, в вагоне метро. Такие практики чтения как будто бы не лишали его возможности основательно погрузиться в предлагаемый воображаемый мир, но и не гарантировали такого погружения, скорее предполагали восприятие темпорально ограниченное, со множеством отвлекающих факторов, требовали от читателя концентрации и цепкости в выявлении того, что же в этом выдуманном мире происходит. Это чтение «на бегу», конечно, отличается от традиционного литературного чтения: с возможностью вдуматься, когда нужно – остановиться, жадно впиваться в каждое слово, обнаруживая его связность с другими, с возможностью неспешного осмысления и размышления, с возможностью «полного погружения» – так, как будто никакого другого мира вокруг не существует. Такое чтение, конечно, можно назвать «медленным» (неслучайно эта практика,

закрепившаяся впоследствии в филологии, в русскоязычном контексте получила название «медленного чтения», а в англоязычном – «close reading», в котором слово «close» взято из семантического поля, связанного с дистанцией; так, в русскоязычном варианте пространственная семантика заменяется темпоральной, указывающей уже на скорость восприятия).

Режим медленного чтения задается и создается не только ситуацией восприятия (зависимой от самого читателя и очень индивидуальной), но и самой художественной формой: тем, насколько последовательно и связно она раскрывает нарративный мир, насколько неспешно и «объяснительно» предъявляет этот мир читателю, как поэтапно – шаг за шагом, прием за приемом – втягивает читателя (перцептивно, эмоционально, ментально) в происходящее в этом мире. Такое медленное чтение на нарративном уровне обеспечивается приближением события и времени фабульного и события и времени дискурсивного – зазор чаще есть, но небольшой, к тому же он сюжетно мотивирован и тоже обеспечивает вовлечение (например, создает интригу). Разветвленное повествование на уровне скорости, как мы видели, функционирует совершенно иначе, имитируя для читателя режим «быстрого», а не «медленного» чтения.

Переживание высокоскоростного движения в разветвленных нарративах, таким образом, моделируется при помощи сразу нескольких повествовательных средств: приемом «умножения» событий и персонажей, преобладанием «поворотных моментов», перевесом «резюме» над «сценами» и «паузами», репрезентацией физического движения героя, ритмической интенсификацией. Оно позволяет устанавливать телесно-миметические отношения между читателем и повествовательным миром, позволяет потому в теле читателя воплотить переживание, характерное для

современной цифровой культуры. Нарративные средства, задействованные в моделировании такого переживания, сближаются с аудиовизуальными элементами репрезентации, характерными для посткинематографа. Изменения, наблюдаемые в аудиовизуальной репрезентации в кино, находят аналогичное воплощение на нарративном уровне, прежде всего на уровне организации нарративной скорости и ритма повествования.

Список источников

Auster P. 4321. London: Faber & Faber, 2017. 1070 p.

Giraud B. *Vivre vite*. Paris: Flammarion, 2022. 205 p.

Kieslowski K. *Przypadek*. 1981.

Туквер Т. *Lola rennt*. 1998.

Список литературы

Барт Р. Введение в структурный анализ повествовательных текстов / Пер. Г. К. Косикова // *Зарубежная эстетика и теория литературы XIX–XX вв.: трактаты, статьи, эссе*. М.: МГУ, 1987. С. 387–422.

Барт Р. *Избранные работы*. Семиотика. Поэтика. М.: Прогресс, 1989. 616 с.

Женетт Ж. *Фигуры*. Работы по поэтике. В 2 т. М.: Издательство имени Сабашниковых, 1998. 944 с.

Моретти Ф. *Буржуа: между историей и литературой*. М.: Издательство Института Гайдара, 2014. 264 с.

Стиглер Б. *Время чтения и новые инструменты памяти // Лаканалия*. 2019. № 32. С. 142–168.

Шкловский В. Б. *О теории прозы*. М.: Федерация, 1929. 265 с.

Baetens J., Hume K. *Speed, Rhythm, Movement: A Dialogue on K. Hume's Article "Narrative Speed"* // *Narrative*. 2006. № 14 (3). P. 349–355. DOI: [10.1353/nar.2006.0008](https://doi.org/10.1353/nar.2006.0008)

Bell D. F. *Real Time: Accelerating Narrative from Balzac to Zola*. Urbana: University of Illinois Press, 2004. 157 p.

Bernaerts L., De Geest D., Herman L., Vervaeck B. *Stories and Minds: Cognitive Approaches to Literary Narrative*. Lincoln: University of Nebraska Press, 2013. 223 p.

Bordwell D. *Perplexing Plots: Popular Storytelling and the Poetics of Murder*. New York

Chichester, West Sussex: Columbia University Press, 2023. 512 p.

Caracciolo M. *Slow Narrative and the Perception of Material Forms // How Literature Comes to Matter: Post-Anthropocentric Approaches to Fiction*. Edinburgh: Edinburgh University Press, 2020. P. 49–70.

Caracciolo M. *Tell-Tale Rhythms: Embodiment and Narrative Discourse // Storyworlds: A Journal of Narrative Studies*. 2014a. № 6 (2). P. 49–73. DOI: [10.5250/storyworlds.6.2.0049](https://doi.org/10.5250/storyworlds.6.2.0049)

Caracciolo M. *The Experientiality of Narrative: An Enactivist Approach*. Berlin, Boston: De Gruyter, 2014b. 244 p.

Caracciolo M., Kukkonen K. *With Bodies: Narrative Theory and Embodied Cognition*. Columbus: Ohio State University Press. 2021. 230 p.

Dannenberg H. *Ontological Plotting: Narrative as a Multiplicity of Temporal Dimensions // The Dynamics of Narrative Form: Studies in Anglo-American Narratology*. Berlin, New York: De Gruyter, 2005. P. 159–190.

Genette G. *Nouveau discours du récit*. Paris: Seuil, 1983. 128 p.

Gerrig R.J. *Experiencing Narrative Worlds: On the Psychological Activities of Reading*. New Haven, London: Yale University Press, 1993. 273 p.

Gingrich B. *The Pace of Fiction: Narrative Movement and the Novel*. Oxford: Oxford University Press, 2021. 214 p.

Hoffman Y. *The Pace of Fiction: Narrative Movement and the Novel*. Brian Gingrich // *Modern Philology*. 2024. Vol. 122. № 2. E000.

Hume K. *Aggressive Fictions: Reading the Contemporary American Novel*. Ithaca: Cornell University Press, 2012. 200 p.

Hume K. *Narrative Speed in Contemporary Fiction // Narrative*. 2005. Vol. 13. № 2. P. 105–124. DOI: [10.1353/nar.2005.0010](https://doi.org/10.1353/nar.2005.0010)

Hven S. *Cinema and Narrative Complexity: Embodying the Fabula*. Amsterdam: Amsterdam University Press, 2017. 260 p.

Hven S. *Enacting the Worlds of Cinema*. New York: Oxford University Press, 2022. 208 p.

Klauk T., Köppe T. *Telling vs. Showing: Handbook of Narratology / P. Hühn, J. Ch. Meister, J. Pier, W. Schmid (eds.)*. Berlin, München, Boston: De Gruyter, 2014. P. 846–853. <https://doi.org/10.1515/9783110316469.846>

Kukkonen K. The Speed of Plot // *Orbis Litterarum*. 2020. № 75. P. 73–85. <https://doi.org/10.1111/oli.12251>

Packard S. Two-dimensional Narrative Celocity // *Journal of Literary Semantics*. 2008. № 37 (1). P. 55–67. DOI: [10.1515/jlse.2008.004](https://doi.org/10.1515/jlse.2008.004)

Popova Y. Stories, meaning, and experience: narrativity and enaction. New York: Routledge, 2015. 210 p.

Prince G. The Disnarrated // *Style*. 1988. Vol. 22. 1. P. 1-8.

Richardson B. A Poetics of Plot for the Twenty-first Century: Theorizing Unruly Narratives. Columbus: Ohio State University Press, 2019. 218 p.

Richardson B. Unnatural Voices: Extreme Narration in Modern and Contemporary Fiction. Columbus: Ohio State University Press, 2006. 166 p.

Ryan M.-L. Embedded Narratives and Tellability // *Style*. 1986. № 20 (3). P. 319–340.

Ryan M.-L. Possible Worlds, Artificial Intelligence and Narrative Theory. Bloomington: University of Indiana Press, 1991. 291 p.

Shapiro S. Post-Continuity: An Introduction // *Post-Cinema: Theorizing 21st-Century Film*. S. Denson, J. Leyda (eds). Falmer: ReFrame Books, 2016. 990 p.

Virilio P. Esthétique de la disparition. Paris: Éditions Galilée, 1989. 125 p.

Corpus Material

Auster, P. (2017). *4321*. London: Faber & Faber.

Giraud, B. (2022). *Vivre vite*, Flammarion, Paris.

Kieslowski, K. (1981) *Przypadek*.

Tykwer, T. (1998). *Lola rennt*.

References

Barthes, R. (1987). Introduction à l'analyse structurale des récits. *Zarubezhnaya estetika i teoriya literatury XIX—XX vv.: traktaty, stat'i, esse* [Foreign aesthetics and theory of literature of the XIX—XX centuries: treatises, articles, essays], MGU Publ., Moscow, Russia, 387–422. (*In Russian*)

Barthes, R. (1989). *Izbrannye raboty. Semiotika. Poetika* [Selected writings. Semiotics. Poetics], Progress Publ., Moscow, Russia. (*In Russian*)

Genette, G. (1998). *Figures*, Izdatel'stvo imeni Sabashnikovykh, Moscow, Russia. (*In Russian*)

Moretti, F. (2014). *The Bourgeois: Between History and Literature*, Izdatel'stvo Instituta Gaydara, Moscow, Russia. (*In Russian*)

Stiegler, B. (2019). Les temps de la lecture et les nouveaux instruments de la mémoire, *Lakanaliya*, 32, 142–168. (*In Russian*)

Shklovskiy, V. B. (1929). *O teorii prozy* [On the Theory of Prose], Federatsiya Publ., Moscow, Russia. (*In Russian*)

Baetens, J., Hume, K. (2006). Speed, Rhythm, Movement: A Dialogue on K. Hume's Article "Narrative Speed", *Narrative*, 14(3), 349–355. DOI: [10.1353/nar.2006.0008](https://doi.org/10.1353/nar.2006.0008) (*In English*)

Bell, D. F. (2004). *Real Time: Accelerating Narrative from Balzac to Zola*, University of Illinois Press, Urbana. (*In English*)

Bernaerts, L., De Geest, D., Herman, L., and Vervaeck, B. (2013). *Stories and Minds: Cognitive Approaches to Literary Narrative*, University of Nebraska Press, Lincoln. (*In English*)

Bordwell, D. (2023). *Perplexing Plots: Popular Storytelling and the Poetics of Murder*, Columbia University Press, New York Chichester, West Sussex. (*In English*)

Caracciolo, M. (2014a). Tell-Tale Rhythms: Embodiment and Narrative Discourse, *Storyworlds: A Journal of Narrative Studies*, 6 (2), 49–73. (*In English*) DOI: [10.5250/storyworlds.6.2.0049](https://doi.org/10.5250/storyworlds.6.2.0049)

Caracciolo, M. (2014b). *The Experientiality of Narrative: An Enactivist Approach*, De Gruyter, Berlin, Boston. (*In English*)

Caracciolo, M. (2020). Slow Narrative and the Perception of Material Forms, in *How Literature Comes to Matter: Post-Anthropocentric Approaches to Fiction*, Edinburgh University Press, Edinburgh, 49–70. (*In English*)

Caracciolo, M., Kukkonen, K. (2021). *With Bodies: Narrative Theory and Embodied Cognition*, Ohio State University Press, Columbus. (*In English*)

Dannenberg, H. (2005). Ontological Plotting: Narrative as a Multiplicity of Temporal Dimensions, in J. Pier (Ed.), *The Dynamics of Narrative Form: Studies in Anglo-American Narratology*, De Gruyter, Berlin, New York, 159–190. (*In English*)

Genette, G. (1983). *Nouveau discours du récit*, Seuil, Paris. (*In French*)

- Gerrig, R.J. (1993). *Experiencing Narrative Worlds: On the Psychological Activities of Reading*, Yale University Press, New Haven, London. (In English)
- Gingrich, B. (2021). *The Pace of Fiction: Narrative Movement and the Novel*, Oxford University Press, Oxford. (In English)
- Hoffman, Y. (2024). The Pace of Fiction: Narrative Movement and the Novel. Brian Gingrich, *Modern Philology*, 122, 2, E000. (In English)
- Hume, K. (2012). *Aggressive Fictions: Reading the Contemporary American Novel*, Cornell University Press, Ithaca. (In English)
- Hume, K. 2005. Narrative Speed in Contemporary Fiction, *Narrative*, 13 (2), 105–124. (In English) DOI: [10.1353/nar.2005.0010](https://doi.org/10.1353/nar.2005.0010)
- Hven, S. (2017). *Cinema and Narrative Complexity: Embodying the Fabula*, Amsterdam University Press, Amsterdam. (In English)
- Hven, S. (2022). *Enacting the Worlds of Cinema*, Oxford University Press, New York. (In English)
- Klauk, T., Köppe, T. (2014). Telling vs. Showing, in P. Hühn, J. Ch. Meister, J. Pier, W. Schmid (eds.). *Handbook of Narratology*, De Gruyter, Berlin, München, Boston, 846-853. (In English) <https://doi.org/10.1515/9783110316469.846>
- Kukkonen, K. (2020). The Speed of Plot, *Orbis Litterarum*, 75, 73–85. (In English) <https://doi.org/10.1111/oli.12251>
- Packard, S. (2008). Two-dimensional Narrative Velocity, *Journal of Literary Semantics*, 37(1), 55-67. (In English) DOI: [10.1515/jlse.2008.004](https://doi.org/10.1515/jlse.2008.004)
- Popova, Y. (2015). *Stories, meaning, and experience: narrativity and enaction*, Routledge, New York. (In English)
- Prince, G. (1988). The Disnarrated, *Style*, 22 (1), 1–8. (In English)
- Richardson, B. (2006). *Unnatural Voices: Extreme Narration in Modern and Contemporary Fiction*, Ohio State University Press, Columbus. (In English)
- Richardson, B. (2019). *A Poetics of Plot for the Twenty-first Century: Theorizing Unruly Narratives*, Ohio State University Press, Columbus. (In English)
- Ryan, M.-L. (1986). Embedded Narratives and Tellability, *Style*, 20 (3), 319–340. (In English)
- Ryan, M.-L. (1991). *Possible Worlds, Artificial Intelligence and Narrative Theory*, University of Indiana Press, Bloomington. (In English)
- Shaviro, S. (2016). Post-Continuity: An Introduction, in S. Denson, J. Leyda (eds), *Post-Cinema: Theorizing 21st-Century Film*, Reframe Books, Falmer. (In English)
- Virilio, P. (1989). *Esthétique de la disparition*, Éditions Galilée, Paris. (In French)
- 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.
- Шулятьева Дина Владимировна**, кандидат филологических наук, доцент, заместитель руководителя Школы философии и культурологии, Факультет гуманитарных наук, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия.
- Dina V. Shulyatyeva**, Candidate of Sciences in Philology, Associate Professor, Deputy Head of School of Philosophy and Cultural Studies, Faculty of Humanities, HSE University, Moscow, Russia.