ОТЧЕТ О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

ЯЗЫКОВОЕ И КОГНИТИВНОЕ МНОГООБРАЗИЕ В ФОРМАЛЬНЫХ МОДЕЛЯХ, КОМПЬЮТЕРНЫХ ИНСТРУМЕНТАХ И ОБРАЗОВАТЕЛЬНЫХ РЕСУРСАХ (заключительный)

2025

СОДЕРЖАНИЕ

# СПИСОК ИСПОЛНИТЕЛЕЙ

Руководители темы:

|  |  |  |
| --- | --- | --- |
| науч. рук. | \_\_\_\_\_\_\_\_\_\_\_\_ | Д. Николс |
| НИУ ВШЭ, PhD | подпись, дата | (введение и заключение) |
| науч. сотр., и. о. зав. лаб. | \_\_\_\_\_\_\_\_\_\_\_\_ | Г. А. Мороз |
| канд. филол. наук | подпись, дата | (введение и заключение) |

Исполнители:

|  |  |  |
| --- | --- | --- |
| ведущий науч. сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Ландер Ю.А. |
| кандидат филол. наук | подпись, дата | (раздел ) |
| ведущий науч. сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Майсак Т.А. |
| доктор филол. наук | подпись, дата | (раздел ) |
| ведущий науч. сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Плунгян В.А. |
| доктор филол. наук | подпись, дата | (раздел ) |
| старший науч. сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Коряков Ю.Б. |
| кандидат филол. наук | подпись, дата | (раздел ) |
| научный сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Ермолова М.В. |
| кандидат филол. наук | подпись, дата | (раздел ) |
| научный сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Земичева С.С. |
| кандидат филол. наук | подпись, дата | (раздел ) |
| научный сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Кошелюк Н.А. |
| кандидат филол. наук | подпись, дата | (раздел ) |
| научный сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Наккарато М.К. |
| phd | подпись, дата | (раздел ) |
| научный сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Оскольская С.А. |
| кандидат филол. наук | подпись, дата | (раздел ) |
| научный сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Рыжова Д.А. |
| кандидат филол. наук | подпись, дата | (раздел ) |
| научный сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Халилова З.М. |
| phd | подпись, дата | (раздел ) |
| научный сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Шушурин Ф.Г. |
| phd | подпись, дата | (раздел ) |
| научный сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Яковлева А.В. |
| кандидат филол. наук | подпись, дата | (раздел ) |
| младший науч. сотр. | \_\_\_\_\_\_\_\_\_\_\_\_ | Холодилова М.А. |
| кандидат филол. наук | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Алексеева А.П. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Балаханов Т.Д. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Власов Е.Э. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Головина А.А. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Григорьева Е.А. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Дедов Т.Г. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Дунаева К.О. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Иванова А.Д. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Кузнецова С.В. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Логвинова Н.Н. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Макарчук И.В. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Манусов А.В. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Мартыненкова А.Е. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Меленченко М.Г. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Наследскова П.Л. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Осоргин И.Г. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Падалка П.В. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Русских А.А. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Филатов К.В. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Финкельберг Л.Д.Н. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Чиркин А.Д. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Шведова Е.Е. |
|  | подпись, дата | (раздел ) |
| стажер-исследователь | \_\_\_\_\_\_\_\_\_\_\_\_ | Якубсон Б.М. |
|  | подпись, дата | (раздел ) |
| менеджер | \_\_\_\_\_\_\_\_\_\_\_\_ | Стогова О.И. |
|  | подпись, дата | (раздел ) |

# РЕФЕРАТ

Отчет … стр., рис. …, табл. …, … источник.

ЛИНГВИСТИЧЕСКАЯ ТИПОЛОГИЯ, ЯЗЫКИ РОССИИ, ИРАНСКИЕ ЯЗЫКИ, ЛИНГВИСТИЧЕСКОЕ РАЗНООБРАЗИЕ, ФОРМАЛЬНО-ЛИНГВИСТИЧЕСКИЕ МОДЕЛИ, КОМПЬЮТЕРНЫЕ МЕТОДЫ ОБРАБОТКИ ЯЗЫКОВЫХ ДАННЫХ, КВАНТИТАТИВНАЯ ЛИНГВИСТИКА, ЛИНГВИСТИЧЕСКИЕ КОРПУСА

Исследование “Языковое и когнитивное многообразие в формальных моделях, компьютерных инструментах и образовательных ресурсах” посвящено изучению феномена лингво-когнитивной вариативности в языках Российской Федерации и сопредельных государств. Языки в рамках проекта исследуются в синхронной, диахронической и типологической перспективе. Целью является построение и эмпирическая верификация методами логики, лингвистики и философии формальных и функциональных моделей, лежащих в основе лингво-когнитивной вариативности.

Методологическая база проекта включает в себя: методы типологической и ареальной лингвистики, экспериментальной и квантитативной лингвистики, когнитивного и формально-семантического моделирования, методы корпусной и вычислительной лингвистики, векторной семантики и автоматического извлечения информации.

По проекту получены следующие новые результаты: получено комплексное типологические описание, моделирующее дискурсивные, морфосинтаксические, лексические и фонетические аспекты языкового многообразия в пределах трех лингво-географических ареалов; описаны и статистически валидированы факторные модели усвоения и повседневного употребления русского языка в условиях контактирования с другими языками Российской Федерации; созданы формальные модели, описывающие комплекс лингвистических явлений в ряде языков России; с точки зрения лингво-семиотического подхода предложены способы разрешения затруднений, препятствующих эффективному информационному обмену в условиях языкового и когнитивного многообразия; разработаны нейросетевые методы и технологии, оптимизирующие процесс преподавания русского языка как иностранного, а также совершенствующие методы автоматической обработки и генерации больших языковых данных на русском языке, в том числе - для государства и общества. Полученные результаты предназначены для применения в образовании, в сфере языковой политики, в социальной сфере для повышения качества жизни населения Российской Федерации. Ряд технологий находятся на стадии внедрения в культурных и образовательных институциях России, а также используются для мониторинга отечественного рынка автоматических лингвистических инструментов. В долгосрочной перспективе полученные разработки будут иметь значимый положительный эффект для российского общества. Дальнейшее их развитие предполагается на последующих этапах проекта.

# ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ (переведут на русский)

В настоящем отчете о НИР применяют следующие сокращения и обозначения.

2 — второе лицо;

2/3 — второе или третье лицо;

2sg — второе лицо единственное число;

3 — третье лицо;

3pl — третье лицо множественное число;

3sg — третье лицо единственное число;

abl — аблатив;

abs — абсолютив;

acc — аккузатив;

add — аддитив;

adv — наречие;

agr — согласование;

an — одушевлённый;

aug — наращение;

ben — бенефактив;

caus — каузатив;

cng — коннегатив;

comp — показатель подчинения;

comp — компаратив;

coord — сочинительный союз;

cvb — конверб;

d1 — демонстративы разных серий;

d3 — демонстративы разных серий;

dat — датив;

def — определённый;

dem — указательное местоимение;

detr — детранзитив;

dim — диминутив;

du — двойственное число;

edge — граница;

el — элатив;

emp — эмфатический;

emph — эмфатический;

eq — экватив;

erg — эргатив;

ess — эссив;

f — женский род;

freq — фреквентатив;

fut — будущее время;

gen — генитив;

i — класс I;

imp — императив;

imp — императив;

impf — имперфектив;

in — в контейнере;

indef — неопределённый;

indef — неопределенное местоимение;

inf — инфинитив;

ins — творительный;

int — вопросительный;

intj — междометие;

io — непрямой объект;

ipfv — имперфектив;

it — итив;

loc — местный;

m — мужской род;

msd — масдар;

n — средний род;

neg — отрицание;

nfin — нефинитный;

nfut — не будущее время;

nom — именительный;

npfv — имперфектив;

npst — непрошедшее время;

nsg — средний род единственное число;

o — косвенная форма местоимения;

obl — косвенный;

p — пациентивный аргумент переходного глагола;

pass — пассив;

pf — перфектив;

pfv — перфектив;

pl — множественное число;

plz — множественность;

poss — посессив;

pot — потенциалис;

prf — перфект;

prn — местоимение;

proh — прохибитив;

prs — настоящее время;

pst — прошедшее время;

ptcl — частица;

ptcp — причастие;

purp — целевой;

ref — референциальный;

refl — рефлексив;

s — единственный аргумент непереходного глагола;

sa — агентивный аргумент;

sg — единственное число;

КМА — комплексная моторная афазия;

ЭфА — эфферентная афазия;

Сем — семантическая афазия

сенс — сенсорная афазия;

АМ — акустико-мнестическая афазия;

Т — тотальная афазия (типология афазий дана по [1–3];

Дидакт — дидактические тексты;

Худож — художественные тексты;

Упрощ — упрощенные из датасета RuSimplAphasia

# 1. Введение

Данный отчет посвящен построению формальных и функциональных моделей, лежащих в основе языковых компетенций, рассуждений и информационных взаимодействий когнитивных агентов, а также эмпирической верификация этих моделей: (а) унификация формальных методов анализа когнитивного и языкового многообразия в логике, лингвистике и философии; (б) исследования малых языков с использованием полученных моделей. Эмпирическая верификация включает в себя исследование и документация языков России и сопредельных государств, включая диалекты и различные регистры и пограничные варианты русского языка в синхронной, диахронической и типологической перспективе. Это предполагает создание цифровых инструментов и ресурсов, теоретическое и квантитативное моделирование языковых явлений и естественных рассуждений.

Наши задачи:

— разработка формальных методов логического и лингвистического анализа факторов, влияющих на когнитивную обработку информации в условиях языкового многообразия. — сбор и анализ полевых материалов (элицитация, тексты и переводы текстов); создание новых инструментов и ресурсов по языкам России и сопредельных государств (корпуса, словари, базы данных по отдельным языковым явлениям, в том числе наборы данных, которые могут использованы для тестирования компьютерных моделей на способность осваивать различные грамматические категории естественного языка). — создание новых цифровых ресурсов и инструментов для изучения русского языка и культуры в различных перспективах (базы данных, корпуса, архивы, словарные платформы и др.). Разработка комплексной цифровой среды для хранения и анализа разноплановых языковых данных на материале малых иранских языков. — квантитативное моделирование языковых систем на основе корпусов и словарей (в том числе произведенных в ходе реализации проекта), а также данные языковых экспериментов. — фундаментальные исследования: моделирование языковых явлений на основе данных языков России (включая русский) и сопредельных государств; апробация компьютерных моделей (включая современные методы искусственного интеллекта) на языковом материале. — синхронное изучение пограничных языковых явлений, интерференции, прагматики, дискурса, а также синтаксическое моделирование языковых единиц. — типологическое изучение русского языка: разработка и редактирование типологических анкет; описание фрагментов лексического профиля русского языка на типологическом фоне; выявление ключевых особенностей диахронических процессов в русской признаковой и глагольной лексике в широкой типологической перспективе и со специальным акцентом на близкородственных славянских, а также иранских языках. — разработка нейросетевых методов 1) для коррекции текстов, написанных носителями русского языка, в т.ч. на иностранных языках; 2) для морфологического и синтаксического анализа исторических текстов. Компьютерное моделирование дискурса, в частности академического. — исследование закономерностей, проявляющихся в академических текстах на английском языке, написанных носителями русского языка, в области прагматических функций многословных последовательностей в различных регистрах академической речи.

В разделе освещаются типологические исследования в пределах трех ареалов: Кавказского, Памирского и Северной Евразии (включая Арктику). Для языков Кавказа исследуется ареальная и генетическая вариативность дискурсивных единиц. Малые иранские (языки Памира, главным образом шугнанский) изучаются типологически в отношении целого ряда категорий и свойств, которые позволяют моделировать соответствующие языковые системы для обеспечения их цифровыми ресурсами. Ареал Северной Евразии и Арктики типологически описывается в двух аспектах: морфосинтаксис пространственных отношений и фонетика-фонология-просодия. Отдельный раздел составляют лексико-типологические исследования, развивающие оригинальную научную школу. Они опираются на широкий круг языков, не связанных друг с другом географически или генетически.

Представьте, что здесь введение от общеязыкового.

По мере накопления эмпирических знаний о языках России все более актуальным становится построение новых и уточнение существующих теоретических моделей с учетом этих знаний. В то же время разрабатываемые формальные модели служат более мощным и современным инструментом анализа эмпирических языковых данных. Одной из задач настоящего исследования является представление данных языков России с использованием формальных моделей (формальной семантики, синтаксиса, прагматики). Эта задача достаточно объемная и многоаспектная. В 2025 г. для моделирования были выбраны следующие классы явлений: кодирование временных отношений в языке, анафора и связывание, передвижение в полипредикации, согласование с нетривиальными контролерами, полярность (логические операторы в языке). Помимо этого, разрабатывались ресурсы, служащие эмпирической базой для разрабатываемых моделей. Языки России, которые попали в сферу интереса, — это русский, хантыйский, абазинский, кильдинский саамский, мишарский татарский и др.

Для сохранения языкового богатства должны быть отработаны технологии преподавания языка. В числе прочего должны быть созданы инструменты для отслеживания сходства учебных текстов и текстов, написанных носителями. Одним из общепринятых вариантов обучения является написание эссе на заданную тему, которое содержало бы элементы обоснования точки зрения обучающегося. Удобным является и создание инструментов для автоматического обнаружения ошибок, заключающихся не только в нарушениях грамматики, но и синтаксиса, логической структуры предложения или порядка следования фраз для обоснования точки зрения обучаемого. Помимо этого, необходимо поместить изучаемый язык в общий языковой контекст. В связи с этим в данном разделе даются результаты исследования риторической структуры учебных текстов, её сравнение со структурой текстов, написанных профессионалами, работающими на выбранном языке. Также предлагается метод сравнения структуры предложений разных языков, потенциально, позволяющий отличать влияние родного языка обучаемого и вносить исправления в процесс обучения.

Исследование коммуникативных компетенций ограниченно рациональных агентов было направлено на поиск новых подходов к разрешению затруднений, препятствующих эффективному информационному обмену в условиях языкового и когнитивного многообразия. В центре внимания находились такие классические для логики и формальной эпистемологии проблемы как концептуальная неопределенность и референциальная непрозрачность, скептицизм в отношении значения, логические и семантические парадоксы, а также новые теоретические и практические вызовы, обусловленные развитием неклассических логических систем, систем искусственного интеллекта и социальных сетей. Методологически исследование характеризовалось укорененностью в существующих логико-философских традициях и обращением к их истокам не в сугубо исторической, но в реактуализирующей эти традиции перспективе, которое сопровождалось их переосмыслением в контексте современных проблем семантики, семиотики и когнитивных наук.

В данном разделе рассматриваются различные характеристики синтетических, т.е. генерируемых автоматически, текстов, возможности использования больших языковых моделей для решения ряда задач автоматической обработки больших данных на естественном языке, а также собранные подразделением наборы данных, примененные для дообучения моделей в целях реализации прикладных проектов для государства и общества.

В данном разделе будут представлены результаты внедренческой деятельности подразделения — технологии и ресурсы, которые разрабатываются для публичного использования как лингвистами, так и широкими слоями потребителей.

# 2. Типологические исследования

## 2.1 Ареальная и генетическая вариативность в распределении императивного междометия ‘на, возьми!’ в языках Кавказа

В некоторых языках мира, включая многие языки Кавказа, существуют специальные слова, которые используются в повелительных высказываниях, когда говорящий просит адресата взять что-либо из его рук, как в (1; 2). В ситуации непосредственной коммуникации такие высказывания часто сочетаются с жестом говорящего, протягивающего адресату предмет, который он держит в руках. Такие слова, как агульское *ma* или русское *на* (используемое в переводе примеров ниже), не являются глагольными формами императива и вообще глагольными формами, а представляют собой самостоятельные лексемы.

1. Агульский < лезгинские < нахско-дагестанские (корпус текстов)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| «ma» | pu-naa, | «me | ik' | we | čanta.ji» |
| TAKE.IT | говорить.PFV-PRF | этот | IN.класть(IMP) | твой | сумка(IN) |
| '{Дала бутылку с ядом и сказала:} ‘На, положи к себе в сумку.’ ' | | | | | |



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| «ma | čʷa-s | xibu | guni» | pu-na... |
| TAKE.IT | вы-DAT | три | хлеб.ABS | говорить.PFV-CVB |
| '‘Нате вам три хлеба, – сказал он.’ {С этим хлебом пейте молоко.}' | | | | |

Специальные лексемы типа агульского *ma* или русского *на* мы будем называть императивными междометиями со значением ‘на, возьми!’ (соответственно, в подстрочных глоссах они будут обозначаться ярлыком «TAKE.IT»). Эти слова являются **императивными** по значению, поскольку встречаются в повелениях, адресованных второму лицу. Они относятся именно к классу **междометий**, а не к глаголам или особым повелительным частицам, которые используются в некоторых языках для придания дополнительной прагматической окраски форме повелительного глагола (типа русской частицы *-ка*).

Императивные междометия ‘на, возьми!’ можно рассматривать как подтип конативных междометий, функцией которых является получение реакции от адресата [4: 479]. Призывы к тишине вроде *шш!* *цыц!* или обращения к животным типа *брысь!* *кыш!* *тпру!* также являются примерами этого типа междометий. Согласно распространенному определению, прототипические междометия — это монолексемные лексические единицы, которые обычно составляют высказывание сами по себе. Повелительные междометия ‘на, возьми!’ не полностью соответствуют этому определению, поскольку они часто встречаются в более полных предложениях, включающих также другие составляющие (см. ниже). Кроме того, в некоторых языках междометие ‘на, возьми!’ может иметь формы словоизменения (например, показатель множественного числа адресата), что делает его отчасти похожим на глагол.

Императивные междометия ‘на, возьми!’ никогда не были предметом специального типологического исследования, хотя более широкий класс императивных междометий привлекал внимание в литературе. Так, А. Ю. Айхенвальд упоминает существование «особых императивных лексем, которые стоят отдельно от других классов слов» и «могут сопровождаться особыми жестами или другими паралингвистическими сигналами с функцией повеления» [5: 317–325]. Одним из распространённых типов таких лексем являются императивные слова, обращенные к животным и побуждающие их удалиться из определённого места, приблизиться к говорящему, остановиться или начать движение. Аналогично, Гусев [6: 99–102] обсуждает «неглагольные императивы», включая междометия, а также подчеркивает, что в морфосинтаксических терминах мы имеем дело с континуумом между императивными формами глаголов и императивными междометиями, причем между этими двумя крайностями находится довольно много промежуточных форм.

Целью нашего исcледования было установление того, в каких **языках Кавказа** имеются специальные императивные междометия со значением ‘на, возьми!’, классификация таких междометий по типу фонологической формы, а также установление морфосинтаксических особенностей междометий с данным значением. В кратком изложении результаты исследования легли в основу одной из глав Типологического атласа языков Дагестана [7], расширенная статья по материалам исследования подана в журнал «Transactions of the Philological Society».

Основными **источниками** данных для этого исследования были словари (в основном двуязычные, в большинстве из которых русский язык является либо первым, либо вторым их пары). В некоторых случаях информация была найдена в грамматиках. Учитывая, что в русском языке есть специальное междометие *на!* ‘вот, возьми!’ (хотя и омонимичное предлогу *на*), в словарях или указателях проводился поиск выражения *на*, которое часто встречалось в качестве отдельного лексического входа русской части. В словарях с доступным текстовым слоем поиск шел по единицам *на*, (с запятой), *нате* (форма с суффиксом множественного числа адресата) и по глагольным формам императива 2-го лица единственного числа глаголов, а именно *возьми* (совершенный вид) и *бери* (несовершенный вид). В словарях или грамматиках, написанных на английском языке, поиск шел по выражениям *take it* и *take this*, в источниках на французском языке — по *tiens* ‘возьми (это)!’. При этом учитывались только специальные слова для выражения ‘на, возьми!’. Если единственным эквивалентом, предоставленным в источнике, была обычная глагольная форма повелительного наклонения глагола ‘брать, взять’ (например, *al* ‘возьми!’ в азербайджанском), считалось, что данный язык лишен особого императивного междометия ‘на, возьми!’.

Рассмотрим некоторые типичные примеры употребления императивных междометий ‘на, возьми!’, представленные в источниках. Как правило, такие примеры содержат помимо междометия и другой материал. Практически во всех имеющихся примерах междометие стоит в начале предложения.

Одной из распространенных конструкций, в которых встречаются императивные междометия ‘на, возьми!’, является простое повелительное предложение с **прямым дополнением** (3; 4). В эргативных языках нахско-дагестанской семьи дополнение выражено абсолютивным (номинативным) падежом.

1. Хиналугский < нахско-дагестанские [8: 214]

|  |  |
| --- | --- |
| ma | pšä |
| TAKE.IT | bread.ABS |
| '‘Возьми хлеб (на хлеб)!’' | |

1. Гунзибский < цезские < нахско-дагестанские [9: 146]

|  |  |
| --- | --- |
| hũʔũ | ẽš |
| TAKE.IT | apple.ABS |
| '‘Возьми яблоко (на яблоко)!’' | |

Помимо прямого дополнения предложения с повелительными междометиями могут включать **реципиентный** аргумент (обычно в дательном падеже), который в подавляющем большинстве случаев указывает непосредственно на адресата (5).

1. Даргинский литературный < даргинские < нахско-дагестанские [10: 662]

|  |  |  |
| --- | --- | --- |
| ma | ħed | žuz |
| TAKE.IT | you.SG.DAT | book.ABS |
| '‘Возьми книгу (на тебе книгу)!’' | | |

Кроме того, нередки примеры, в которых за ‘на, возьми!’ следует повелительная форма обычного глагола, чаще всего собственно ‘брать, взять’, но также и других (6). Предложение, следующее за междометием, может быть и полным повелительным предложением с выраженными аргументами (7). В таких примерах междометие ‘на, возьми!’ может быть отделено запятой от следующей клаузы предложения, хотя это так не во всех примерах.

1. Кумыкский < тюркские [11: 20]

|  |  |
| --- | --- |
| ma, | aša! |
| TAKE.IT | eat.IMP |
| '‘На, ешь!’' | |

1. Лезгинский < лезгинские < нахско-дагестанские [12: 226]

|  |  |  |  |
| --- | --- | --- | --- |
| gʷa, | qːaču | i | ktab! |
| TAKE.IT | take.IMP | this | book.ABS |
| '‘На, возьми эту книгу!’' | | | |

Интересной особенностью императивных междометий ‘на, возьми!’ является возможность присоединения суффиксальных показателей **множественного числа адресата** (см. о них [13]) в тех языках, где такие показатели есть (8). Это свойство императивных междометий свойственно и русскому *на!* (2 л. ед.ч.) ~ *на-те!* (2 л. мн.ч.), см. обсуждение в работе [6: 99–102].

1. Мегебский < даргинские < нахско-дагестанские [14: 129]

|  |  |
| --- | --- |
| ma! | ma-na! |
| TAKE.IT | take.it-IMP.PL |
| '‘На, возьми!’ ‘Нате, возьмите (к нескольким людям)!’' | |

Обратимся теперь к результатам по варьированию **фонологической формы** императивных междометий ‘на, возьми!’ в языках Кавказа. Прежде всего отметим, что в большинстве языков нашей выборки специальное императивное междометие ‘на, возьми!’ было выявлено. В удинском, азербайджанском и татском языках такого специального элемента нет. В грузинском и армянском языках близкий функциональный эквивалент ‘на, возьми!’, имеющий одну и ту же форму *aha*, выполняет функцию более общего сигнала внимания (типа ‘вот, обрати внимание!’) и, следовательно, не подпадает под наше понимание специального междометия ‘на, возьми!’. (На карте ниже это междометие, тем не менее отмечено, чтобы показать своеобразие его ареального распределения.)

В языках, где есть особое императивное междометие ‘на, возьми!’, оно, как правило, имеет односложную форму CV (в некоторых языках также с конечным ларингальным /h/, дающим вариант CVh). В меньшинстве языков, а именно в цезских и нахских группах, ‘на, возьми!’ — двусложное слово со структурой CVCV.

Сегментный состав императивных междометий весьма ограничен. Так, начальными согласными могут быть /m/, /g, gʷ/ или /h/. Гласный после /m/ — /a/ или /ä/. Гласный после /g/ — /o/, а после /gʷ/ — /a/. Таким образом, эти два варианта, скорее всего, связаны звуковым переходом от /gʷa/ к /go/. Наконец, гласный после /h/ может быть как простым, так и назализованным, а именно /a, ã/, /oː/, /ĩː/ и /ũ/.

На карте ниже все междометия типа ‘на, возьми!’ сгруппированы в три основных типа в соответствии с их фонологической формой ([таблица 2.1](#tbl-1.1)). Учитывая, что все междометия CVCV в цезском и нахском языках имеют начальный /h/ и носовой гласный, их можно отнести к типу *HA(n)*. Грузинское и армянское междометие *aha* показано на карте как представляющее отдельный тип *AHA*.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 2.1 – Формальные типы повелительных междометий «на, возьми!»   | Тип | Примеры | | --- | --- | | **MA-type** | ma, mah, mä, mäh | | **GwA-type** | go, gʷa | | **HA(n)-type** | ha, hoː, hĩː, hũ, hã; hãʔã, hũʔũ, hãħ(iː) / haħ(iː), hane | |

Распределение фонологических типов императивных междометий ‘на, возьми!’ демонстрирует как **генеалогический**, так и **ареальный** сигнал.

Генеалогически обусловленное распределение проявляется в наличии одного и того же типа фонологической формы междометия в пределах одной семьи или одной ветви. Так, тип *GwA* встречается только в андийских языках (во всех языках этой ветви), а тип *HA(n)* – только в цезских и нахских (опять же, во всех языках этих ветвей). В свою очередь, тип *MA* встречается во всех даргинских языках нашей выборки, преобладает в лезгинской ветви (за исключением удинского), а также среди тюркских языков выборки (за исключением азербайджанского). Таким образом, формальный тип побудительных междометий устойчив в отдельных нахско-дагестанских ветвях и, в меньшей степени, среди тюркских языков Дагестана (в двух из трёх языков).

|  |
| --- |
| Рисунок 2.1 – Ареальное распределение фонологических типов междометий ‘на, возьми!’ в языках Кавказа (автор карты – А. А. Иванова) |

С другой стороны, в более широкой перспективе можно увидеть чёткий ареальный сигнал в распределении типов. Прежде всего, тип *MA* преобладает на обширной территории и встречается в центральных и южных ветвях нахско-дагестанских языков (лакской, даргинской, лезгинской, хиналугской), в аварском, который в этом отношении соответствует центральным ветвям, а не андийским или цезским языкам, и в двух тюркских языках (кумыкском и ногайском). Интересно, что ареал *MA* может охватывать и гораздо большую территорию, поскольку побудительные междометия ‘на, возьми!’ с формой *ma* также зафиксированы, по крайней мере, в некоторых западнокавказских языках и в карачаево-балкарском (тюркском).

Во-вторых, единственные две ветви с императивными междометиями типа *HA(n)*, а именно цезские и нахские, географически близки (хотя их разделяет андийская ветвь, имеющая свой собственный уникальный тип *GwA*).

Наконец, отсутствие императивного междометия ‘на, возьми!’ наблюдается на южной/юго-восточной периферии ареала, а именно в азербайджанском, удинском и татском языках. Эти три языка относятся к трём разным языковым семьям, но распространены на одной территории.

Кроме того, можно отметить, что междометия с формой *aha* встречаются в двух генеалогически неродственных, но географически смежных языках – грузинском и армянском (ср. также абхазское *aħǝ*). Помимо общей формы, эти два междометия объединяет также их функция, которая носит более общий характер привлечения внимания, а не собственно повеление ‘возьми!’.

Учитывая выявленное распределение фонологических типов междометий, можно задаться вопросом о том, может ли распространение междометий типа *MA* на Кавказе быть обусловлено влиянием тюркских языков, которые контактируют как с западнокавказскими (карачаево-балкарский), так и с нахско-дагестанскими языками (кумыкский, азербайджанский, в некоторой степени также ногайский). В качестве альтернативы, тюркские языки Кавказа могли заимствовать междометия типа *MA* от собственно кавказских языковых семей. Второй сценарий был бы маловероятен, если бы оказалось, что другие тюркские языки, не распространенные на Кавказе, также обладают императивными междометиями ‘на, возьми!’ типа *MA*. Мы проверили эту возможность, чтобы уменьшить количество гипотетических сценариев распространения междометия.

Тюркские языки образуют большую языковую семью, включающую десятки языков, распространенных на обширной территории Евразии. Для дополнительной выборки, необходимой для проверки наличия в тюркских языках, распространенных за пределами Кавказа, междометия ‘на, возьми!’ типа *MA*, мы выбрали еще шесть языков в дополнение к четырем названным выше: это туркменский, татарский, узбекский, якутский, хакасский и чувашский, представляющие различные ветви семьи. Как и в случае с нашей основной выборкой, основным источником для этого дополнительного микроисследования были словари. Результаты исследования были однозначными: все шесть тюркских языков дополнительной выборки обладают специальными междометиями ‘на, возьми!’ типа *MA*: ср. *ma* в узбекском и хакасском языках, *mä* в татарском и якутском, *me* в туркменском и чувашском. Это однозначно свидетельствует о том, что существование междометий типа *MA* в тюркских языках Кавказа никоим образом не объясняется влиянием автохтонных кавказских языков. Напротив, весьма вероятно, что тюркские языки Кавказа принесли междометия типа *MA* на Кавказ и могли послужить источником последующего ареального распространения этого типа междометий как в западной (западнокавказские языки), так и в восточной частях Северного Кавказа (нахско-дагестанские языки центральной и южной ветвей). При этом окончательное происхождение междометий типа *MA* со значением ‘на, возьми!’ в тюркских, а также в монгольских и тунгусских языках остается неясным и требует более широкого межъязыкового исследования.

## 2.2 Исследования иранских языков

В 2025 году исследования по иранским языкам были посвящены таджикским заимствованиям в шугнанский; системам указательных местоимений и дейксиса в восточноиранских языках: хотаносакском, бартангском; категории рода в языках шугнано-рушанской группы; базисной лексике языков шугнано-рушанской группы; различению именных частей речи в шугнанском; созданию первой языковой модели машинного перевода для шугнанского языка; реляционным свойствам некоторых локативных имен; акустическому анализу лексической и фразовой просодии, вокализма и морфосинтаксиса вопросительных частиц шугнанского языка.

Летом 2025 года студенты и сотрудники проекта ездили в экспедицию на Памир, где продолжили изучение шугнанского языка, так что большинство исследований основаны на полевых данных. Шугнанский язык — восточноиранский язык памирского ареала, на котором говорят в Горно-Бадахшанской автономной области в Таджикистане и в провинции Бадахшан в Афганистане.

### 2.2.1 Таджикские заимствования в шугнанском языке

Таджикский язык, являясь официальным языком в Таджикистане, оказывает серьезное влияние на шугнанский и другие бесписьменные языки, поскольку это язык средств массовой информации, документооборота, а главное — язык преподавания. В настоящее время мы наблюдаем лексические и грамматические изменения, связанные с постоянным языковым взаимодействием с таджикским языком. В шугнанский язык проникли многие таджикские слова и выражения, так, заимствуются существительные, прилагательные, наречия, местоимения, числительные, глаголы, модальные слова, предлоги и целые синтагмы (если заимствуется именная синтагма целиком, то сохраняется свойственный таджикскому языку грамматический показатель — изафет).

Заимствуясь, лексемы подвергаются фонетической адаптации, а также развивают собственные значения, которых нет в языке-источнике. По степени фонетической адаптации мы можем судить об относительном времени заимствования. К примеру, если слово фонетически совпадает со своим источником, это скорее всего новое заимствование, а если слово адаптировано и без специальных знаний о фонетических изменениях в восточно- и западноиранских языках трудно сопоставимо со своим источником, то это, очевидно, старое заимствование. Существует и промежуточный класс слов — с минимальной фонетической адаптацией. Здесь также следует сказать, что арабские заимствования, как правило, проникли в шугнанский через таджикский язык.

К наиболее частым и очевидным фонетическим переходам относятся следующие (перечислим лишь некоторые из них):

1. Поскольку в шугнанском языке нет фонемы *h*, все таджикские слова, содержащие эту фонему, отражаются без неё, например, существительное *amro* 1. ʽпопутчик, спутникʼ, 2. ʽдруг; сослуживец, товарищ по работе, по (какому-л.) занятиюʼ, 3. ʽвместе, совместноʼ из тадж. *ҳамроҳ* 1. ʽспутник, попутчикʼ; 2. ʽвместе, совместноʼ, кроме фонетических изменений, развило в шугнанском свое собственное лексическое значение ʽдруг, сослуживецʼ. Если же фонема *h* оказывается в таджикском в интервокальной позиции, между двумя огубленными гласными, она отражается в шугнанском как *w*: *mowūt* ʼсукно, суконныйʼ из тадж. *моҳут* ʽсукноʼ, если же гласные неогубленные, то как *y*: *mayīn* прил. ʽтонкий, нежныйʼ из тадж. *маҳин* ʽтонкий, нежныйʼ. Вполне вероятно, что именно в таком виде эти слова были заимствованы не из литературного таджикского, а из ближайших бадахшанских говоров, где фонема *h* также выпадает.
2. Довольно регулярно таджикская фонема *š* отражается в шугнанском как *x̌*: *nix̌tār* ʽострыйʼ из тадж. *нештар* ʽланцет, скальпельʼ, перен. ʽострыйʼ; *pex̌* ʽперед, прежнийʼ из тадж. *пеш* 1. ʽперед, передняя частьʼ; 2. ʽпередний, предшествующийʼ. При этом в недавних заимствованиях такой переход не происходит: *ošiq* ʽлюбовник, влюбленныйʼ из араб./тадж. *ошиқ* ʽвлюблённыйʼ.
3. Таджикское *z* переходит в шугнанском в *ӡ*: *awʒ* ʽсущ. водоем, бассейн; прил. глубокийʼ из тадж./араб. *ҳавз* ʽхауз, бассейн, водоемʼ (в шугнанском это слово может быть существительным и прилагательным, тогда как в таджикском только существительным).
4. Таджикское *v* регулярно отражается в шугнанском как *w*: *kawok* ʽдуплистый; пустой; рыхлыйʼ из тадж. *ковок* 1. ʽпустой, полый; дуплистыйʼ; 2. ʽрыхлый, пористыйʼ; *lawand* ʽлодырьʼ из тадж. *лаванд* ʽбездельник, лодырьʼ.
5. Таджикское сочетание *on* отражается как *ůn* в относительно новых заимствованиях: *mêrabůn* ʽмилостивыйʼ из тадж. *меҳрубон* 1. ʽласковый, нежный, приветливый; добрыйʼ; 2. ʽлюбезный; обходительный, милосердныйʼ; *důnix̌* / *důniš* ʽзнание, образованиеʼ из тадж. *дониш* 1. ʽзнание, знанияʼ; 2. ʽмудростьʼ; тогда как в старых заимствованиях как *en*: *miyénd* / *mayénd* ʽпоясʼ из тадж. *миён* кн. ʽремень, поясʼ, сюда же можно отнести случаи с глагольными суффиксами каузатива: в старых заимствованиях используется суффикс -*en-*, тогда как в недавних — суффикс -*ůn-*: шугн. *garδentow* ʽвращать, вертеть, кружить; поворачивать и др.ʼ из тадж. *гардон(и)дан* ʽвращать, вертеть, кружить; крутить и др.ʼ, тогда как шугн. *čakůntow* ʽзакапывать, капать и др.ʼ из тадж. *чакон(и)дан* ʽкапать, наливать каплямиʼ. Также как *ů* отражается таджикская *ū*: шугн. *kůr* ʽслепец, слепойʼ из тадж. *кӯр* ʽслепой; слепец; незрячийʼ.

Для приблизительного анализа количества заимствованной лексики мы рассмотрели три группы лексем:

1. простые глаголы (это замкнутый класс слов, в котором чуть больше 600 лексем);
2. имена, которые в зависимости от своей синтаксической позиции могут быть существительными или прилагательными (некоторые из них могут также использоваться в качестве наречий), в эту группу попало около 450 пар;
3. список базисной лексики, или стословный список М. Сводеша.

По материалам исследования подготовлен доклад на международной конференции «Индоиранские языки памиро-гиндукушского ареала и за его пределами» (4–6 декабря 2025 г., Санкт-Петербург, ИЛИ РАН).

### 2.2.2 Базисная лексика, или стословный список Сводеша для шугнанского и бартангского

Стословные списки базисной лексики М. Сводеша были собраны у носителей шугнанского и бартангского языков, проводились параллели между двумя языками, анализ заимствований из таджикского и других языков, был подробно проанализирован ряд нетривиальных этимологий. Расхождения между двумя языками немногочисленны, и только два расхождения удается зафиксировать между исконными лексемами двух языков. Вероятно, их число могло быть и больше, если бы не значительный пласт таджикских заимствований, имеющий место не только в языках шугнано-рушанской группы, но и других иранских идиомах Памира. Заимствованные относительно недавно, они отразились преимущественно сходным образом в обоих языках. Случаев расхождения, где один таджикизм противопоставлялся бы другому, не наблюдается. В базисной лексике из списка Сводеша число заимствованных существительных (25 из 53) и прилагательных (7 из 15) в процентном отношении выше, чем глаголов (3 из 15); личные местоимения и числительные не заимствованы. Этот список интересен исследователям лексики иранских языков своей ареальной лексикой, не имеющей параллелей в других иранских языках и иногда тяжело поддающейся этимологизации, или сохранившимися иранскими словами, которых нет в таджикском и персидском. В результате исследования подготовлена публикация статьи в журнале «Родной язык» (готовится к печати; предполагаемая дата публикации — декабрь 2025 г.).

### 2.2.3 Проблема различения именных частей речи в шугнанском

Современный шугнанский отличается бедной именной морфологией, что приводит, в частности, к размыванию границ между именными классами: многие имена могут употребляться в синтаксических позициях, характерных для существительных и прилагательных, без каких-либо морфологических маркеров, см. (9) и (10).



|  |  |
| --- | --- |
| mu | kaxoy |
| PRN.1SG.O | женщина |
| 'моя жена' | |



|  |  |
| --- | --- |
| kaxoy | gāp |
| женщина | разговор |
| 'женский разговор [элицитация]' | |

Эта особенность характерна для многих иранских языков, однако, насколько нам известно, нет работ, в которых бы подробно анализировалась эта проблематика и выделялись бы критерии, позволяющие разграничить классы прилагательных и существительных. В нашем исследовании мы ставим перед собой задачу поиска таких критериев для шугнанского языка.

Согласно [15: 23], [16: 293], задача разграничения классов существительных и прилагательных легко решается в языках, где есть грамматическая категория именного рода, поскольку для существительных она словоклассифицирующая, а для прилагательных — словоизменительная (этот же критерий предлагается в [17] для авестийского). В шугнанском эта категория устроена нетривиально. Выделяется мужской и женский род, и многие глаголы в форме перфекта и / или претерита согласуются по этому признаку с субъектом. Однако некоторые существительные имеют две формы (ср. *buc* ‘мальчик’ — *bic* ‘девочка’), ряд существительных относится в одном значении к мужскому роду, в другом — к женскому (*ziv*, m ‘язык, речь’ — *ziv*, f ‘язык (часть тела)’), и еще целый класс имен вызывает согласование то по мужскому, то по женскому роду в зависимости от референта, ср. *duxtur* ‘доктор’ [18: 33–41]. С другой стороны, большинство слов, употребляющихся в атрибутивной позиции и попадающих в семантический класс признаков (см. типично признаковые семантические классы в [19], [15]), не согласуются с определяемым существительным по роду, см. (11)–(12). Кроме того, многие признаковые имена, для которых зафиксированы две родовые формы в словаре Д. Карамшоева [20], в современном языке одну из этих форм утратили.



|  |  |  |  |
| --- | --- | --- | --- |
| ［yi | bašānd | sůg］ | lův! |
| INDEF | хороший | сказка(F) | сказать.IMP |
| 'расскажи хорошую сказку! [Buor-at Nuqrā]' | | | |



|  |  |  |  |
| --- | --- | --- | --- |
| ［mi | rāng | bašānd | x̌ār］=at… |
| D1.M.SG.O | цвет | хороший | город(M)=COORD |
| 'такой хороший город, а… [Safed Xitur]' | | | |

Другая типологически распространенная диагностика — употребление маркеров сравнительной степени, которые сочетаются, прежде всего, с прилагательными. Этот критерий для шугнанского также не очень показателен. В шугнанском есть морфологический компаратив *-di*, однако присоединяется он не только к признаковым основам (13), но и к именам с семантикой, больше характерной для существительных, см. (14).



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Yam | piš | as | wam | piš | ʒulik-di. |
| D1.SG | кошка | EL | D3.F.SG.O | кошка | маленький-COMP |
| 'Эта кошка меньше той кошки [элицитация]' | | | | | |



|  |  |  |  |
| --- | --- | --- | --- |
| Dod | az | puc | pazanda-di |
| отец | EL | сын | повар-COMP |
| 'Отец больше повар, чем сын [элицитация]' | | | |

Тем не менее можно выделить ряд конструкций, в которых предполагаемые прилагательные и существительные ведут себя по-разному. Это, во-первых, предикативная конструкция с субъектом во множественном числе, где некоторые именные слова могут опционально присоединять маркер множественного числа (→ сущ.), а некоторые не могут (→ прил.), см. (15)–(16). Во-вторых, конструкции с глаголом, где в позицию глагольного модификатора легко встраиваются прилагательные и хуже встраиваются существительные (17)–(18).



|  |  |  |
| --- | --- | --- |
| Māδ | čīd-en=en | ɣullā(\*-yen). |
| D1-PL | дом-PL=3PL | большой |
| 'Эти дома --- большие [элицитация]' | | |



|  |  |  |
| --- | --- | --- |
| Karamšo=yat | Safīnā=yen | raīs(ᴼᴷ-en). |
| Карамшо=COORD | Сафина=3PL | начальник-PL |
| 'Карамшо и Сафина --- начальники [элицитация]' | | |



|  |  |
| --- | --- |
| tez | tīdow |
| быстрый | идти.INF |
| 'быстро идти [элицитация]' | |



|  |  |
| --- | --- |
| \*kaxoy | nīwdow |
| женщина | плакать.INF |
| 'Ожидаемое значение: ‘по-женски плакать’ [элицитация]' | |

### 2.2.4 Системы указательных местоимений и дейксиса в восточноиранских языках: хотаносакском и бартангском

Эти системы отличаются умеренной сложностью, наличием как минимум трех элементов-серий, выделением в отдельную серию ближнего, среднего, дальнего либо анафорического компонентов. Хотаносакский является вымершим языком, не оставившим после себя потомков. Таким образом, в качестве основного метода исследования был выбран анализ корпуса сохранившихся текстов, а также историческая реконструкция. В результате исследования было установлено, что система указательных местоимений хотаносакского языка демонстрирует значительные инновации, отсутствующие в других иранских языках. Во-первых, это нетипичная композиция системы: различаются анафорическая, ближняя и дальняя серии местоимений. Во-вторых, все три серии образуются от супплетивных основ *ṣ*- [*ʐ*-] / *tt*- [*t*-] с разницей в аффиксальной части. В силу типологической обособленности хотаносакских указательных местоимений от общеиранского контекста была предпринята попытка сравнить их с указательными местоимениями ареально близких тохарских языков. Среди параллелей в пратохарском можно выделить то же, что и в хотаносакском функциональное распределение указательных местоимений в трехчастной системе (анафорический, ближний, дальний), образование всех трех серий от супплетивных основ \**s*- / \**t*-, маркирование дейктического контраста при помощи суффиксальных показателей, а также аналогии в происхождении этих показателей. Таким образом были сформулированы аргументы в пользу того, что хотаносакская и тохарская системы указательных местоимений структурно идентичны в результате конвергенции, вызванной языковыми контактами. По материалам работы был выполнен доклад на конференции по сравнительно-историческому языкознанию «XX традиционные чтения памяти С. А. Старостина» (24–25 марта 2025 г., Москва, НИУ ВШЭ), подготовлен текст статьи к публикации.

Система указательных местоимений бартангского языка изучалась посредством экспериментальных методов в условиях полевой работы (в экспедиции в области распространения бартангского языка в июле 2025 г. (Рушанский район, ГБАО Таджикистана)). За основу для исследования были взяты анкеты, разработанные в [21], [22] и основывающиеся на данных современной лингвистической типологии. В результате серии экспериментов, проведенных на открытой местности в естественных для консультантов условиях, удалось выяснить, что три серии указательных местоимений бартангского языка ранжируются по двадцати пяти контекстам, см. [рисунок 2.2](#fig-deic-pic). На основании экспериментов были сделаны выводы о доменах каждой из трех серий указательных местоимений; установлены актуальные факторы вариативности, такие как ±видимость, ±досягаемость, ±контакт с объектом, ±область взаимодействия локуторов, ±принадлежность и др.; определен «немаркированный элемент» в системе указательных местоимений, то есть тот, который доминирует в подавляющем большинстве контекстов. В бартангском языке «немаркированным элементом» является указательное местоимение второй серии YID, отвечающее за прототипические контексты среднего дейксиса, то есть условную близость объекта к адресату. Однако местоимение YID не является абсолютно допустимым для всех контекстов, см. (1), (3), (12) и (25) на [рисунок 2.2](#fig-deic-pic). Это наблюдение разнится с (распространенным в иранистической литературе) представлением об указательном местоимении второй серии как «общеуказательном», то есть, лишенном своего дейктического значения. В действительности, YID не теряет своего значения, но распространяется на смежные значения двух других серий, что примечательно, не доходя в этом до логического конца. По материалам исследования подготовлен доклад на международной конференции «Индоиранские языки памиро-гиндукушского ареала и за его пределами» (4–6 декабря 2025 г., Санкт-Петербург, ИЛИ РАН).

|  |
| --- |
| Рисунок 2.2 – Распределение указательных местоимений бартангского языка по основным дейктическим контекстам. |

### 2.2.5 Категория рода в языках шугнано-рушанской группы

В результате исследования значительно пересмотрены существующие в иранистических работах представления о категории рода в бартангском, учитывается значение референциального / нереференциального употребления той или иной именной группы, в силу чего род имени может меняться на противоположный. Сформулирована классификация таких изменений в соответствии с расширенной иерархией одушевленности. Также тема рода изучалась в близкородственном шугнанском языке. В результате исследования, проведенного методом элицитации, установлены особенности показателей категории рода, связанные с выражением значений собирательности или массовости (также встречается определение «массовое число»). Были выработаны аргументы в пользу того, чтобы считать собирательность отдельной грамматической категорией в шугнанском языке, выражаемой преимущественно при помощи грамматических показателей мужского рода, также были обнаружены новые значения категории собирательности, которые прежде не отмечались для шугнанского языка, к примеру, неконтейнируемость объектов, т. е., невозможность четко определить их объем. Доклад, посвященный смежным в бартангском языке значениям рода и референциальности, основанный на материалах, собранных ранее в результате элицитации, был представлен на международной конференции «Tenth International Conference on Iranian Linguistics (ICIL 10)» (13–15 января 2025, University of Tuscia, Viterbo, Италия), подготовлен текст статьи к публикации.

### 2.2.6 Сложные глаголы в шугнанском языке

Сложные глаголы (СГ), то есть многокомпонентные глагольные лексемы, являются характерной чертой иранских языков. Сложные глаголы состоят из неглагольного компонента (НК), который обычно несет основное лексическое значение, и легкого глагола, который выражает грамматические значения. В шугнанском сложные глаголы широко используются, однако до сих пор не получили должного внимания лингвистов. Грамматики [23], [24] приводят только краткие описания шугнанской системы сложных глаголов.

Термин сложные глаголы применяется к разным конструкциям, от лексикализованных до композициональных и продуктивных. В нашем исследовании мы применили к шугнанским СГ различные диагностики, используемые в литературе о синтаксических свойствах СГ ([25]; [26]; [27]; [28], среди прочих). Цель нашего исследования заключалась в том, чтобы исследовать синтаксические свойства шугнанских СГ, а также границу между лексикализованными и композициональными глагольными конструкциями.

Во время полевой работы, проведенной в г. Хороге (Таджикистан) летом 2025 года, мы рассмотрели около 40 шугнанских глаголов с именными и адъективными неглагольными компонентами. В общем и целом все исследованные единицы показали похожее синтаксическое поведение, независимо от типа НГ.

Первый набор диагностик, проведенных нами, сосредоточен на разделимости шугнанских СГ, то есть могут ли части сложного глагола быть разделены другими синтаксическими единицами. Исследование показало, что шугнанские СГ очень хорошо разделимы: как минимум так же, как персидские СГ, если не больше. Части СГ могут быть разделены ваккернагелевскими клитиками, наречиями образа действия, объектами, выраженными местоимениями и почти всегда объектами выраженными именной группой, а также адъюнктами (см. примеры (19)–(20)).

Второй набор диагностик посвящен модификации неглагольных компонентов маркерами множественного числа и прилагательными (для именных НК), или маркерами сравнительной степени и наречиями (для адъективных НК). Наши данные свидетельствуют о том, что у шугнанских СГ есть ограниченные возможности к модификации. Именные НК в целом позволяют модификацию прилагательными, а адъективные могут быть модифицированы наречиями и принимать компаративный суффикс (см. примеры (21)–(22)); модификация множественным числом менее приемлема.

Третья часть нашего анализа затрагивает анафорические свойства именных НК. Оценки носителями демонстративов с НК в качестве антецедентов были непоследовательны, хотя некоторые контексты разрешались большим количеством информантов (см. пример (23)). В примерах ниже части СГ заключены в квадратные скобки.



|  |  |
| --- | --- |
| ［Yax］=um | ［čū-d］. |
| лёд=1SG | делать-PST |
| 'Я замёрз.' | |



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Uz=um | tar | pārk | ［bůy］ | gol-en | /wev | ［čū-d］. |
| PRN.1SG=1SG | EQ | парк | запах | цветок-PL | D3.PL.O | делать-PST |
| 'Я понюхал цветы / их в парке' | | | | | | |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Uz=um | xu | nān | xax̌ | ［fānd | δo-d.］ |
| PRN.1SG=1SG | REFL | мать | сильно | ложь | дать-PST |
| 'Я сильно соврал (букв. дал большую ложь) моей матери' | | | | | |



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Tar | čīd | mu | ［gārm-di | su-t］, | čim | tar | dargo. |
| EQ | дом | PRN.1SG.O | тёплый-COMP | стать.M.SG-PST | чем | EQ | снаружи |
| 'Мне стало теплее дома чем на улице' | | | | | | | |



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| X̌ac | ［yāx\_i | ču-d］. | (%Yu\_i | šito) |
| вода(F) | лёд(M) | делать-PST | D3.M.SG | холодный |
| 'Вода замёрзла (%Он [лёд] холодный)' | | | | |

### 2.2.7 Модель машинного перевода для шугнанского языка

Шугнанский язык, на котором говорит около 90 тысяч человек в Таджикистане и Афганистане, отличается очень ограниченным числом письменных источников, а также отсутствием стандартизированной орфографии, что сильно затрудняет цифровизацию и доступность современных средств обработки естественного языка. Наша работа по созданию первой языковой модели машинного перевода для языковой пары шугнанский-русский позволяет не только двигаться в направлении расширения возможных электронных ресурсов и распространения методов машинного обучения и искусственного интеллекта на малоресурсные языки, но и использовать переводчик для расширения параллельных корпусов за счет синтетических данных, что может быть полезно не только носителям языка и путешественникам, но также лингвистам и преподавателям.

Работа состояла прежде всего из подбора и предобработки обучающих данных и различных экспериментов с обучением модели. Для обучения модели машинного перевода необходимы были шугнанско-русские параллельные данные. Такие данные были взяты из двух основных источников: 1) корпуса шугнанского языка, содержащего устные тексты с переводами и переведенные на шугнанский язык фрагменты Евангелия от Луки; шугнанско-русского словаря Д. Карамшоева, содержащего большое количество сентенциальных примеров с переводами. Предварительная обработка включала удаление аннотаций, специальных символов и знаков ударения, приведение текста к нижнему регистру и нормализацию пробелов, а также унификацию орфографии. Поскольку у шугнанского языка отсутствует стандартизированное письмо, все данные были приведены к кириллической графике с помощью разработанного ранее конвертера, что позволило максимизировать совпадение токенов с русским и таджикским языками и обеспечить согласованность между различными ресурсами.

Мы использовали два тестовых набора. Первый, взятый из проекта BivalTyp, включает 123 короткие пары «шугнанский–английский», иллюстрирующие двухвалентные глаголы. Русские эталонные переводы были добавлены через английский язык. Второй тестовый набор состоит из 110 предложений из многодоменного датасета NLLB, незначительно адаптированных к контексту Памира и переведённых на шугнанский язык носителями языка из г. Хорога.

Мы провели серию экспериментов по дообучению многоязычных моделей машинного перевода с использованием различных вспомогательных языков. Схема была такой: объединить корпус данных «шугнанский–русский» с корпусом того же размера на вспомогательном языке и обучить общую модель во всех направлениях перевода.

В финальной системе в качестве вспомогательного языка мы использовали пушту, а также расширили обучающие данные с помощью обратного перевода шугнанского романа и дополнительных фрагментов Евангелия на шугнанском и таджикском языках. Такая конфигурация дала наилучшие общие результаты и позволила создать пригодную для практического использования систему машинного перевода между шугнанским и русским языками. Метрики финальной модели в двух размерах (600 миллионов параметров и 1.3 миллиарда параметров) можно увидеть на [таблица 2.2](#tbl-4.7.1) ниже. Переводчик доступен по ссылке: <https://huggingface.co/spaces/Novokshanov/Shughni-Translator>.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 2.2 – Метрики на объединенной тестовой выборке. Лучшие результаты выделены жирным.   | Source | Meteor sh→ru | Meteor ru→sh | chrF sh→ru | chrF ru→sh | | --- | --- | --- | --- | --- | | Baseline | 35.6 | 23.9 | 34.5 | 25.3 | | Pashto (aux) | 38.3 | 24.1 | 35.7 | 25.6 | | Slovenian (aux) | 38.9 | 23 | 36.1 | 24.4 | | Final 1.3B | **39.3** | 22.7 | **37.5** | 25.0 | | Final 600M | 35.6 | **25.8** | 35.4 | **28.2** | |

### 2.2.8 Акустический анализ лексической и фразовой просодии, вокализма и морфосинтаксиса вопросительных частиц шугнанского языка

Про ударение в шугнанском языке известно только то, что оно фиксированное и, как правило, падает на последний слог. Существуют, однако, исключения, возникающие по морфологическим причинам. Так, например, лично-числовые клитики всегда безударны, а отрицательный и прохибитивный префиксы всегда ударны. Детальные акустические исследования лексической просодии в шугнанском до сих пор не проводились, но по предварительным данным, ударение сигнализируется повышенной ЧОТ и интенсивностью.

Для проведения исследования была составлена анкета для сбора необходимого фонетического материала. Анкета состояла из нескольких типов заданий, а именно: 60 слов, состоящих из двух-четырех слогов, с одинаковыми по качеству гласными. Качество гласных контролировалось для того, чтобы устранить возможный спецэффект, связанный с вариативностью параметров у разных по качеству гласных. Также использовались минимальные пары типа *bozor* ‘базар’ — *bozor-i* ‘базарный’, где в первом слове ударный гласный находится в корне, а во втором на деривационном суффиксе, так, чтобы можно было посмотреть на акустику одного гласного в ударной и безударной позициях. Исследовалась также фонология ударения: проверялась способность нести на себе ударение у лично-числовых клитик, личных окончаний глаголов и деривационных суффиксов с помощью минимальных пар предложений, типа Bek=en *zīd* (encl 3pl) ‘Бека убили’ vs. *Bezen odam kitob-en na-x̌oyd* (stem) ‘Глупый человек ничего не читает’. Речь восемнадцати носителей шугнанского языка (возраст: 17–49, ср. = 22; 16 ж, 2 м) записывалась на наголовный диктофон в тихой комнате. Стимульные слова повторялись два-три раза в нормальном, естественном для них темпе. Акустические измерения были выполнены в программе PRAAT [29]. Для конечных согласных, различающихся по глухости / звонкости, измерялось три параметра : длительность гласного, средняя интенсивность и средняя ЧОТ. Все параметры измерялись в миллисекундах. По результатам предварительного анализа, лексическое ударение в шугнанском характеризуется повышенной ЧОТ и большей длительностью гласного. Эти два параметра являются основными коррелятами ударения типологически. Интенсивность гласного, впрочем, не играет значительной роли в сигнализации ударения. После проведения статистического анализа средних значений параметров было выявлено, что главным акустическим ключом к ударению в шугнанском является длительность гласного.

### 2.2.9 Разработка и пополнение цифровых ресурсов

Оцифровано 12 диалектологических карт Таджикистана, на каждой из которых представлены данные по 100 поселениям. Данные по каждой карте переведены в табличный формат, включающий сведения о поселении (название, координата), документируемой лингвистической характеристике и реализации этой характеристики в каждом из 100 поселений.

Также оцифровано 5000 словарных статей из словаря И. И. Зарубина (Шугнанские тексты и словарь. Москва, Ленинград: Изд-во Акад. наук СССР, 1960). Словарные статьи размечены словарные зоны (формы, значения, примеры, примечания, пометы, перекрестные ссылки) с сохранением иерархической структуры; вычитаны на предмет ошибок распознавания; выгружены на сайт «Путеводитель по памирским языкам» (<pamiri.online>). Настроены режим доступа к словарю и поиск по добавленному источнику. Словарная база данных pamiri.online, развиваемая в рамках проекта и содержащая оцифрованные версии словарей шести разных памирских языков, переведена в формат OnLex 3.0. Новый формат подразумевает более гибкую систему межстатейных ссылок и систему многоязычности, а также возможность добавления подзначений, транскрипций и медиафайлов. В нем также усовершенствован пользовательский и редакторский интерфейс (поддержка темной темы, адаптивность), добавлена поддержка английской версии сайта.

Произведено обновление и расширение функционала сайта pamiri.online: улучшенный поиск, отображение статей с возможностью ссылок на отдельные их части, увеличение числа словарных зон в соответствии со спецификациями платформы OnLex 3.0

## 2.3 Выражение пространственных отношений в языках Северной Евразии

### 2.3.1 Выражение Цели и Источника в языках Северной Азии

В рамках проекта в 2025 году исследовались пространственные отношения в языках Северной Азии. В ходе работы были подробно рассмотрены пространственные системы двенадцати языков ареала: тундровый ненецкий (самодийские), вах-васюганский хантыйский (хантыйские), бурятский (монгольские), эвенкийский (тунгусо-маньчжурские), долганский (тюркские), амурский нивхский (нивхские), айнский (айнские), колымский юкагирский (юкагирские), кетский (енисейские), центрально-сибирский юпик (эскимосские), ительменский (южные чукотско-камчатские), чукотский (северные чукотско-камчатские). Примеры кодирования пространственных отношений (всего около 1600 предложений) были собраны из грамматических источников, корпусных материалов и параллельных текстов и размечены по топологическому отношению, направительному значению и морфосинтаксическому типу средства.

Полученная база данных локативных конструкций языков Северной Азии использовалась для исследования асимметрии кодирования топологических отношений в Цели и Источнике, о которой см. [30]. Одной из важных тем, раскрытых в исследовании, является также синкретизм ориентаций (Цели, Места и Источника). Синкретизм Цели и Места — характерная черта ареала, представляющая собой, по всей видимости, контактное явление. В двух языках (бурятском и долганском) наблюдается полное совпадение Цели и Места (что, однако, не влияет на различительность Цели в сравнении с Источником). В тундровом ненецком, ительменском и центрально-сибирском юпике Цель, Место и Источник полностью различаются. В нивхском языке наблюдается довольно редкая типологически полная нейтрализация Места и Источника. В шести языках (эвенкийском, юкагирском, чукотском, кетском и хантыйском) обнаруживается частичный синкретизм Цели и Места, ограниченный для некоторых локализаций и средств.

Как синкретизм ориентаций, так и синкретизм локализаций, вызывающие асимметрию Цели и Источника, различаются для разных по степени грамматикализованности и формальной симметричности средств.

Наиболее грамматикализованные средства — падежные показатели — ожидаемо имеют менее специфицированное значение и являются более синкретичными: обычно они кодируют несколько локализаций и/или ориентаций. Синкретизм ориентаций вызывает семантическую асимметрию Цели и Источника, поэтому падежное кодирование обладает довольно большим значением асимметрии Цели и Источника.

Адлоги, наоборот, обычно означают только одну локализацию, за исключением редких случаев, обычно не склонны к синкретизму. Они наиболее асимметричны, в том числе из-за того, что неизменяемы.

Реляционные имена обладают меньшими способностями к синкретизму локализаций и ориентаций, чем падежные показатели, хотя и содержат их. Дело в том, что падежные показатели на реляционных именах имеют исключительно ориентационное значение, тогда как самостоятельные падежные показатели кумулятивны (означают и локализацию, и ориентацию), а значит, более грамматикализованы и более синкретичны. В отличие от адлогов, реляционные имена формально симметричны. Из-за двух этих особенностей система реляционных имен наименее асимметрична. Таким образом, асимметрия Цели и Источника следует не столько из семантики локализаций, сколько из грамматикализованности средств, которые их кодируют.

Сравнение средств кодирования пространственных отношений в языках Северной Азии позволяет проследить возможные пути грамматикализации пространственных средств в целом. Разные по морфологическому статусу средства в разных языках (а иногда даже в одном языке) позволяют уточнить классический линейный подход к грамматикализации (падежные показатели < адлоги < реляционные имена < нереляционные имена). Помимо этого, среди примеров кодирования пространственных значений в языках выборки встречаются конструкции, которые ставят под вопрос обобщение о порядке показателей ориентации и локализации и позволяют расширить набор возможных конструкций, используемых для кодирования пространственных участников.

В исследовании уделено внимание и взаимодействию пространственных значений с другими именными категориями. Влияние на кодирование пространственных участников оказывает референциальный статус ориентира и его одушевленность.

По результатам исследования подготовлены устные доклады «Нейтрализация Цели и Места в языках Северной Азии» на студенческую конференцию «Корпусные технологии и компьютерные науки в гуманитарном знании» (Конкорт — 2025, 2–3 июня 2025, Нижний Новгород, НИУ ВШЭ), «Синкретизм Цели и Места в языках Северной Евразии» на пятую конференцию по уральским, алтайским и палеоазиатским языкам (24–26 ноября 2025, Санкт-Петербург, ИЛИ РАН),«Source—Goal Asymmetry Across Topological Relations in the Languages of Northern Eurasia» на конференцию «16th International Conference of the Association for Linguistic Typology» (ALT—2026, 1–3 июля 2026, Lyon (France), Lumière University of Lyon).

### 2.3.2 Пространственные падежи в самодийских языках

Протосамодийский коаффикс \*-*kǝ*- и его связи с пространственными падежами в самодийских и других уральских языках по статье [31] были рассмотрены на регулярном семинаре «Падежные чтения», который проходит на факультете гуманитарных наук.

В статье приводятся две теории происхождения западноуральских пространственных *-s*-падежей: традиционная лативная теория и новаторская послеложная. Согласно лативной теории, элемент *-s*- является дативным/лативным коаффиксом, как и протосамодийский \*-*kǝ*-. Послеложная теория предполагает, что *-s-* является основой послелога, присоединявшего падежные показатели и впоследствии сросшегося с зависимым словом. Предполагается также, что прасамодийский аблатив \*-*kǝ-tǝ* может быть когнатом марийского аблативного послелога *gə̑č́*: ср. в нганасанском *tuj-kə-tə* и в горномарийском *tul gə̑č́* ‘из огня’.

В докладе были также рассмотрены новые данные, показывающие, что коаффикс -*xă*-, когнатный \*-*kǝ*-, не только входит в состав пространственных падежей, но и способен атрибутивизироваться: так, суффикс -*xăj°* в современном лесном ненецком языке и его когнаты в тундровом ненецком и других самодийских языках образуют прилагательные со значением ‘находящийся в Х-е’: *ṕetaλʹa-xăj°* ‘находящийся в лесу’.

## 2.4 Исследования фонетики и фонологии в языках Севера и Арктики

### 2.4.1 Фонологические особенности гортанной смычки в лесном ненецком языке

В рамках проекта в 2025 году исследовались фонологические особенности гортанной смычки в лесном ненецком языке.

Известно, что в ненецких языках (самодийские, уральские) существует гортанная смычка, которая появляется из-за нейтрализации шумных фонем в позиции коды. В тундровом ненецком гортанная смычка существует как [ʔ] перед сонорными, а перед шумными блокирует озвончение глухих шумных фонем, при этом происходит её синкопа: /pyiʔmya/ > pyiʔmya ‘штаны’; /xata/ > xada ‘бабушка’; /laʔta/ > lata ‘широкий’.

Данные для анализа были собраны в июле-августе 2025 г. в ЯНАО. В лесном ненецком гортанная смычка имеет два аллофона: смычку [ʔ] перед сонорными и придыхание [ʰ] перед шумными, при этом поствокальное озвончение взрывных отсутствует: /ṕiʔḿa/ > ṕiʔḿa ‘штаны’; /kata/ > kata ‘бабушка’; /laʔta/ > laʰta ‘широкий’. Однако точное распределение аллофонов /ʔ/ и устройство фонологических процессов, затрагивающих смычку в лесном ненецком языке, не было исследовано. В рамках исследования было дополнено описание гортанной смычки и показано существование вариативности в выборе аллофона смычки перед взрывным аллофоном /j/ – [dʹ], использующимся перед поверхностными гласными: /teʔ-dʹa-n°/ > teʔ-dʹa-n° ~ teʰ-dʹa-n° (‘собирать\_ягоды-SFS-1SG>PL’) ‘я собираю ягоды’.

В лесном ненецком в определённых контекстах происходит эпентеза гласного после смычки. Исходя из данных в примерах (24)–(26), контекст правила в нотации генеративной фонологии задан в (27).

1. /waʔw°/ > waʔaw° ‘постель’
2. /ḿaʔ-j°/ > ḿaʔa-j° чум-POSS.1SG ‘мой чум’
3. /ẃiʔ-λ°/ > ẃiʔi-λ° вода-POSS.2SG ‘твоя вода’
4. ø > V / Vʔ\_\_[+son]

В рамках исследования также моделируются фонологические процессы с гортанной смычкой (выбор аллофона, вставка гласного) в теории оптимальности. В то же время, теория оптимальности обладает таким недостатком как перепроизводство (overgeneration) — чем больше ограничений постулировано, тем больше грамматик существует. Однако на самом деле не всегда находятся грамматики с другими порядками постулированных ограничений. В то же время некоторые автосегментные теории (Government Phonology, Strict CV) также оперируют ограничениями, однако они не являются ранжированными, а применяются тогда, когда для них есть контекст. В будущем планируется моделирование процессов с гортанной смычкой в лесном ненецком в рамках теорий автосегментной фонологии.

По результатам исследований подготовлены доклады «Forest Nenets Glottal Stop Phonology» на Конференцию по типологии и грамматике для молодых исследователей (ИЛИ РАН, г. Санкт-Петербург) и «Glottal Stop in Forest Nenets: Optimality Theoretic Approach» на Пятую конференцию по уральским, алтайским и палеоазиатским языкам (ИЛИ РАН, г. Санкт-Петербург).

### 2.4.2 Ассимиляция сибилянтов в литовском языке

Исследовались некоторые морфонологические процессы в литовском языке (балтийские, индоевропейские). Материалы получены из грамматик литовского языка.

В литовском ассимиляции по глухости/звонкости ([voice]) и палатализованности ([front]) являются регрессивными, в то время как ассимиляция по месту ([place]) распространяется от более центральных морфем к более периферийным, то есть от корня к аффиксам (28).

1. /uʒ-sʲuto/→ [usʲ.sʲu.to] PFV-разозлиться.PST.3 ‘(он) разозлился’

Однако если благодаря результатам ассимиляций сибилянты совпадают по всем признакам, может произойти дегеминация — удаление одного из идентичных сегментов. Дегеминация обязательно происходит при присоединении к корню на сибилянт суффикса будущего времени (FUT) /sʲ/, однако при присоединении префикса к корню на сибилянт возникает вариативность – допускается как вариант без ассимиляции по месту, так и с ней, при этом с последующей дегеминацией (29)–(30).

1. /vʲeʒ-sʲ-u/ → [vʲe.ʃʲu] везти-FUT-1SG ‘я повезу’
2. /uʒ-sʲuto/ → [u.sʲu.to]/[uʃʲ.sʲu.to] PFV-разозлиться.PST.3 ‘(он) разозлился’

Описанную систему следует изучить тщательней для возможности её моделирования с помощью теории оптимальности и других фонологических моделей.

По результатам исследования подготовлен доклад «Lithuanian Sibilants Assimilation: Morphophonology in Stratal ОТ» на конференцию «Формальные модели в лингвистике — 2025 (НИУ ВШЭ, г. Москва).

### 2.4.3 Ударение в тунгусских языках

В 2025 году была подготовлена база для полевого исследования системы ударения быстринского диалекта эвенского языка в типологическом освещении. На данный момент не существует описания ударения в эвенском языке, за исключением отрывочных наблюдений в литературе. Для других языков тунгусской языковой семьи существуют описания разной степени подробности, фрагментарно представленные в десятках различных русскоязычных, англоязычных, японоязычных и китайскоязычных источников. Сколько-нибудь подробное и полноценное описание существует только для удэгейского языка [32]. Для того, чтобы в будущем полевом исследовании было возможно собрать релевантные примеры и описать систему ударения быстринского эвенского, были проанализированы и обобщены данные об ударении во всех тунгусских языках. Это позволит вписать планируемое исследование в контекст исторического развития ударения в тунгусских языках.

В ходе изучения и обобщения источников было выявлено, что в большинстве тунгусских языков присутствуют два типа ударения — силовое и музыкальное (тоновое). Ни один из видов ударения не является фонологическим, то есть ударение зависит от фонетических особенностей слова, а не от его значения или грамматической формы. Силовое ударение чаще всего соотносится с началом слова, тяжёлыми слогами (закрытыми слогами, слогами, содержащими долгие гласные или дифтонги), а музыкальное — с концом слова, однако в разных языках и зачастую разных описаниях одного и того же языка описаны значительно различающиеся системы. Иногда такие различия могут объясняться диалектным варьированием, а в некоторых случаях, по-видимому, неточностью и/или неполнотой некоторых описаний.

Исследование и обобщение источников позволило сформировать гипотезы об устройстве эвенского ударения и также подготовило базу для его описания в типологической и генеалогической перспективе после проведения полевого исследования. Итоги изучения и обобщения тунгусского ударения были представлены в виде доклада на регулярном семинаре эвенской экспедиции.

## 2.5 Типологическое многообразие эхо-вопросов

В 2025 г. завершено типологическое исследование эхо-вопросов. В статье были рассмотрены стратегии, используемые для формирования переспросов в более чем 30 языках различных семей и регионов.

Стратегия определяется как набор признаков, отличающих переспрос от стимульного предложения: 1) поверхностное кодирование части, вводящей цитирование; 2) поверхностное кодирование вопросительного семантического компонента; 3) присутствие / отсутствие местоименного дейктического сдвига. Часть, вводящая цитирование, может быть придаточным предложением, аффиксом, клитикой, частицей или оставаться невыраженной. Вопросительный семантический компонент обычно выражается интонацией, вопросительным аффиксом, клитикой, частицей или местоимением. Существует также несколько стратегий, которые не вписываются в эту классификацию. Одна из них — стратегия «удаления», при которой языковой материал (например, вопросительный маркер в стимульном предложении) удаляется (японский, кашмирский). Вторая стратегия заключается в замене личной формы глагола в стимульном предложении на неличную форму глагола в переспросе (нкоре-кига). Третья стратегия — добавление восклицательной частицы к стимульному предложению (коромфе).

Результаты исследования опубликованы в виде статьи в журнале «Вопросы языкознания».

## 2.6 Исследования по лексической типологии

По сравнению с грамматической типологией, типология лексических систем ставит задачу систематизации гораздо более обширного класса единиц. Подробное типологическое описание этого массива данных возможно лишь «малыми шагами»: предметом изучения в лексико-типологических исследованиях обычно становятся отдельные семантические поля, ср. классические зоны для подобного анализа — цветообозначения, термины родства, глаголы движения и др. Объектом нашего исследования стали лексические области, ранее не попадавшие в фокус типологического изучения — глаголам семантических полей ‘менять’ ([раздел 2.6.2](#sec-menyat)) и ‘мешать’ ([раздел 2.6.1](#sec-meshat)). В соответствии с заявленными целями и задачами проекта, эти области подробно изучались как на материале русского языка с привлечением (микро)диахронических данных, так и в широкой типологической перспективе. Все исследования проведены в русле фреймового подхода к лексической типологии [33].

### 2.6.1 Семантическое поле ‘мешать’

Семантическое поле ‘мешать’ было изучено как в широкой типологической перспективе, так и на материале близкородственных славянских языков. Результатом анализа стал набор параметров, которые могут определять выбор глагола для описания ситуации смешивания и, соответственно, лежать в основе лексических противопоставлений в исследуемой семантической зоне.

Материалом для типологического анализа послужили данные индоевропейских (французского, итальянского, испанского, английского, немецкого, норвежского), тюркских (турецкого, татарского), уральских (финского, севернохантыйского), а также японского и вьетнамского. Особое внимание уделялось материалу славянских языков — русского, украинского, польского, чешского, болгарского и сербского. Источниками данных стали толковые и переводные словари, моноязычные и параллельные корпусы текстов, а также опросы информантов, проводившиеся по специально разработанной анкете, которая в дальнейшем может использоваться для расширения типологического охвата исследования.

Анализ полученных данных показал, что для противопоставлений внутри поля смешивания релевантны три основных параметра: тип объекта (в том числе его таксономические и геометрические характеристики), цель действия и способ его осуществления.

#### 2.6.1.1 Тип объекта

С точки зрения **типа** объекта лексически выделяются прежде всего длинные тонкие объекты (волосы, провода, нитки, ср. топологический тип «веревки» в [34]) — для ситуации их соединения в единую массу в языках типологически регулярно используются специализированные лексические единицы, ср. рус. *путать* (*Кошка играла с клубком и* ***запутала*** */* ***\*перемешала*** *нитки*), а также исп. *enredar*, итал. *aggrovigliare*, нем. *verheddern,* вьетн. *rối*, яп. *karamu* и др. Регулярное лексическое разделение ситуаций такого типа и ситуаций с иными типами объектов имеет под собой ясные когнитивные основания: в отличие от прототипической ситуации перемешивания (например, соединения ингредиентов салата или растворения сахара в чае), «смешивание» длинных объектов происходит ненамеренно — в качестве субъекта здесь выступают неагентивные участники, ср. ‘кошка’ в ситуации с нитками или ‘ветер’ в (31):

1. *Ветер* *трепал и* ***путал*** *волосы над свежими, только что вымытыми розовыми личиками молочниц* <…> [А. И. Куприн. Жанета (1933)]

Противопоставление по интенциональности действия, по-видимому, является существенным для концептуализации ситуации, что и находит отражение в типологически частотном лексическом различении смешивания и запутывания. В то же время спутывание, как и собственно смешивание, по сути представляет собой соединение разных объектов — эта общность двух типов ситуаций находит отражение в допустимости их колексификации (хотя она встречается, по-видимому, значительно реже, чем дислексификация): в нашей выборке возможность использования для объектов типа волосы того же глагола, что и для прототипического смешивания ингредиентов, отмечается во французском (*mêler*) и татарском (*чуалу*), ср. употребления в разных типах контекстах для французского *mêler*:

1. французский

a. *Le vent* ***mêle*** *mes cheveux* ‘ветер запутывает мне волосы’.

b. ***Mêlez*** *la purée de marrons à la crème pour un sorbet maison* ‘Смешайте каштановое пюре со сливками для приготовления домашнего сорбета’

Еще один параметр объекта, который может влиять на выбор глагола смешивания, связан с консистенцией соединяемых ингредиентов. В этом отношении прежде всего лексически выделяются ситуации, в которых один из компонентов представляет собой жидкость (речь может идти как о соединении нескольких жидкостей, так и о растворении в жидкости нежидких ингредиентов). Так, в немецком смешивание, в котором задействованы жидкие компоненты, описывается глаголом *rühren*, тогда как воздействие на твердые ингредиенты задается лексемой *mischen*.

Наконец, некоторые частные типы объектов могут получать особое лексическое кодирование. К такого рода случаям относится перемешивание игральных карт, ср. рус. *тасовать*, исп. *barajar.*

#### 2.6.1.2 Цель действия

В отношении цели действия лексически могут разводиться два типа ситуаций смешивания (обратим внимание, что к спутыванию как неагентивной ситуации параметр цели неприменим). В прототипическом случае агенс посредством определенного — как правило, кругового — движения соединяет разные объекты для получения однородной смеси. Но то же самое движение может осуществляться и когда ингредиенты уже соединены — в этом случае цель действия оказывается несколько иной — например, чтобы смесь не подгорела, не застыла и т.д. В русском оба типа ситуаций могут задаваться глаголом *мешать*, как в (33 a–b), хотя могут и различаться на словообразовательном уровне, выражаясь разными приставочными дериватами, как в ((34 а–b), ср. понятие «слабой колексификации» в [35]):

1. русский

a. *Придвинул большой мольберт, взял палитру и стал сосредоточенно* ***мешать*** *краски, выдавливая их из тюбиков*. [Б. Н. Полевой. Силуэты. Новеллы (1974)]

b. *Он все так же попыхивал трубочкой и равнодушно* ***мешал*** *кашу*. [Виктор Некрасов. В окопах Сталинграда (1946)]

1. русский

a. *Так, Плиний Старший писал, что при выпадении волос надо* ***смешать*** *сушеных морских коньков, майорановое масло, смолу и сало.* [Александр Голяндин. Рассказы о животных, и не только о них: А у морского конька что за конек? // «Знание – сила», 2003]

b. *Зулейха бесшумно выдыхает, приподнимается, протягивает палку, чтобы* ***помешать*** *похлебку в ведре*. [Гузель Яхина. Зулейха открывает глаза (2015)]

В других языках два типа ситуаций могут различаться и на уровне глагольных корней, ср. *mix* vs. *stir* в английском, *mischiare* vs. *mescolare* в итальянском.

#### 2.6.1.3 Способ действия

Ядерные глаголы поля типа рус. *мешать*, англ. mix, японск. *мадзэру* и др. являются лексемами с недоспецифицированной семантикой. Подобно другим недоспецифицированным лексическим единицам (ср. о глаголах ‘искать’ и ‘находить’ в [36], о ‘прятать’ в [37], [38]), они задают цель действия (в нашем случае — получить однородную смесь), не указывая на конкретный способ ее достижения. Однако более периферийные глаголы поля могут лексифицировать как раз способ — а именно, описывать движение, при помощи которого достигается однородность смеси. Здесь встречаются прежде всего глаголы, выражающие итеративное каузированное движение — по круговой траектории (ср. итал. *girare* ‘вращать’) или из стороны в сторону (рус. *болтать*, вьетн. *lắc* ‘трясти’, японск. *фуру* ‘трясти’, болг. *клатя* ‘раскачивать’), а также глаголы общей семантики, не специфицирующие траекторию движения, ср. нем. *rühren* ‘двигать, шевелить’.

Подобные глаголы способа действия по сути представляют собой единицы других семантических полей, которые «заимствуются» в поле ‘смешивания’, наглядно демонстрируя принципы семантической смежности и проницаемости межполевых границ (см. о механизмах такого рода сдвигов в [39]). Однако в процессе семантической эволюции эти «чужие» глаголы могут утрачивать связи с источником и становиться специализированными глаголами смешивания (см. об аналогичных явлениях для глаголов поиска, находки и прятания в [36], [37]). Так, английский глагол *stir*, который синхронно безусловно является лексемой исследуемого поля, этимологически связан с идеей каузированного движения (ср. до.-англ. *styrian* ‘двигать’). Сходную эволюцию претерпевает и испанский *revolver*, восходящий к идее кругового движения, но синхронно выражающий прежде всего перемешивание.

Глаголы способа хорошо иллюстрируют и еще один теоретически значимый принцип лексификации семантических полей: параметры, которые можно выделить как основу лексических противопоставлений, не являются независимыми, как это представлялось, например, в исследованиях в рамках теории компонентного анализа (ср., например, [40]). Напротив, выбор значения для одного из них часто предопределяет значения прочих релевантных характеристик. Так, глаголы с семантикой ‘трясти’ предполагают в качестве объекта смешивания жидкость, глаголы вращательного движения часто соотносятся с определенной целью действия — поддержанием консистенции и т.д. Другим ярким примеров такой связи может служить особая группа глаголов способа действия — в отличие от обсуждавшихся выше лексем, эти глаголы выражают не каузированное движение, а воздействие, ср. ‘мять, давить’. Глаголы такой семантики тесно связаны с определенным типом результирующего объекта — прежде всего тестом, а также другими объектами, получаемыми сходным образом (например, глиной), ср., например, употребление в этой зоне польск. *gnieść* ‘мять’.

Подобные устойчивые связи между параметрами ложатся в основу выделения фреймов, которые используются в наших исследованиях в качестве инструмента для сопоставления лексических систем разных языков. Таким образом, языковой материал в семантической зоне смешивания объектов может служить еще одним свидетельством когнитивной обоснованности фреймового подхода к лексической типологии.

### 2.6.2 Семантическое поле ‘менять’

Система русских глаголов поля ‘менять’ заслуживает внимания сразу с нескольких точек зрения. Прежде всего, интересно само ее устройство: ее можно считать одновременно и доминантной (т.е. такой, в которой одна лексема покрывает все или почти все фреймы поля), и дистрибутивной (такой, где практически для каждого фрейма есть свое лексическое средство). С одной стороны, в ней выделяется бесприставочный глагол несовершенного вида *менять(ся)*, который покрывает существенную часть фреймов поля. С другой стороны, в перфективных контекстах этому глаголу соответствует сразу несколько приставочных коррелятов (*изменить(ся)*, *сменить(ся)*, *обменять(ся)*, *разменять, заменить, подменить* и др.), каждый из которых покрывает лишь узкий фрагмент поля и не может считаться чистой видовой парой к широкому по своей семантике и дистрибуции *менять(ся)* (в современном языке на эту роль все более основательно претендует глагол *поменять*). Тем не менее тот факт, что почти все лексемы поля восходят к одному корню, не позволяет считать систему дистрибутивной в классическом смысле: мы можем ожидать, что семантические различия между приставочными дериватами будут более тонкие, чем между глаголами с разными корнями, и с этой точки зрения интересно, как они соотносятся с типологическими закономерностями и подойдет ли для их описания набор фреймов, которые мы выделяем в результате типологического анализа нескольких языков.

Другой важный аспект этой системы, тоже связанный с её деривационной природой, это её подвижность. Известно, что лексические системы вообще существенно больше подвержены изменениям, чем грамматические. Поведение приставочных дериватов меньше изучено с этой точки зрения, однако, как показывает в том числе наше исследование, формирование целого ряда глаголов с одним корнем и разными приставками приводит к обилию близких синонимов, которые постоянно влияют друг на друга (то расходятся из-за общего стремления языка к устранению полных синонимов, то, наоборот, формируют близкие значения по аналогии), что делает систему особенно неустойчивой. Так, поле ‘менять’, почти целиком состоящее из приставочных дериватов, даже на коротком промежутке буквально в 100–-200 лет претерпело существенную перестройку. Были подробно изучены два эпицентра этих изменений: недавняя экспансия глагола *поменять(ся)* ([раздел 2.6.2.5](#sec-pomenyatsya)) и, наоборот, постепенное устаревание глагола *переменить(ся)* ([раздел 2.6.2.6](#sec-peremenitsya)).

В качестве типологического фона для описания русской лексической системы были исследованы соответствующие семантические области нескольких языков (включая малые языки России): хантыйского, карельского, адыгейского, осетинского, японского, китайского, казахского, итальянского, французского, английского, древнеанглийского, норвежского и др. Результатом этих исследований стал следующий набор фреймов — значений, которые могут противопоставляться лексически: Модификация, Превращение, Переключение, Переключение между объектами, Замена, Компенсаторная замена, Замещение, Компенсация, Обмен, Реципрокальный обмен.

#### 2.6.2.1 Фреймы Модификации и Превращения

К этим фреймам относятся ситуации, предполагающие изменение свойств одного участника (ср. *Ребёнок сильно изменился за лето*). Особенно сильные изменения могут привести к превращению объекта во что-то новое (ср. *Гусеница превратилась в бабочку*), но важно при этом, что в таких ситуациях трансформации по-прежнему происходят только внутри одного основного участника. Они допускают также наличие каузатора изменений, ср. ***Победа*** *изменила его до неузнаваемости* и ***Колдунья*** *превратила принца в лягушку.* Эта часть поля имеет тенденцию лексически обособляться в языках нашей выборки. Так, например, в финском языке глагол *muuttaa / muuttua* обозначает Модификацию и Превращение и не употребляется ни в каких других частях системы [41]. В казахском глагол *өзгер-* обозначает только Модификацию, а глагольный корень *айнал-* используется исключительно во фрейме Превращения [42]. Специальные лексические средства для обозначения этих фреймов есть в индонезийском, норвежском, персидском, саамском и многих других языках.

Аналогичная картина представлена и в русском языке: выделяются глаголы, которые обозначают только Модификацию, только Превращение или Модификацию и Превращение — в большинстве случаев не затрагивая другие фреймы поля. Фрейм Модификации обслуживает глагол *изменить(ся)*, не распространяющийся ни на какие другие фреймы, в том числе не покрывающий Превращение. Он может обозначать как каузированное (*деньги изменили его*), так и некаузированное изменение (*за год он сильно изменился*). Ср. тж. глаголы *модифицировать(ся)* и *варьировать(ся)*, последний из которых отличается наличием в своём значении итеративного компонента.

Фрейм Превращения единственный не способен выражаться никакими дериватами от основы *мен-* в современном русском языке. Для его выражения используется, прежде всего, специализированный глагол *превращать*(*ся*), характеризующийся высокой частотностью (135,47 ipm по НКРЯ) и покрывающий исключительно этот фрейм без дополнительных компонентов значения. Помимо *превращать(ся)*, к этому фрейму относятся также устаревающие глаголы *обращаться / оборачиваться*, которые, по всей видимости, уже не могут означать каузированное Превращение, а в контекстах некаузированного Превращения обычно стилистически окрашены. При этом глагол *обернуться* (но не *обратиться*) сформировал значение, очень близкое к семантике Превращения, но не идентичное ему: ‘вызвать, спровоцировать, иметь в качестве последствия’, ср. *ситуация обернулась катастрофой.* Кроме того, в контекстах Модификации и Превращения употребляется ещё ряд глаголов, объединяющих эти два фрейма — ср. *преобразить(ся)* и *преобразовать(ся)*, а также заимствованные аналоги (*трансформировать(ся)* и др.). Контексты преображения и преобразования без выраженного синтаксически результата (т.е. того, во что превращается объект) сложно однозначно отнести к одному из двух рассматриваемых фреймов: одно и то же событие может мыслиться и как значительное изменение, сохраняющее исходный объект, и как превращение, переводящее его в новое качество. Возможность этой неоднозначности, по-видимому, является связующим звеном между этими двумя фреймами, обеспечивая воспроизводящуюся их колексификацию (в терминах [35]).

В русском языке к этим же фреймам можно отнести словообразовательную модель с *пере*-: присоединяя эту приставку, слова, семантика основы которых не имеет непосредственной связи со значениями изменения или превращения, могут выражать значение Модификации и в некоторых случаях Превращения. Глаголы, основы которых имеют достаточно общее значение: *переделывать(ся)*, *переиначивать(ся)*, *переработать(ся)* и др. — выражают значение этого фрейма с точностью до ограничений, накладываемых основой. Например, *переделывать(ся)* требует агентивного субъекта, *переиначивать(ся)* предполагает, что результат изменения можно квалифицировать как ‘другой, иной’. Разнообразие основ, которые образуют глаголы поля ‘менять’ по описанной модели, при их широкой или некомпозициональной семантике указывает на то, что основным носителем значений Модификации и Превращения является префикс *пере*- — или сама модель. Подобные серии формируют, например, глаголы с приставкой *у-* (*умертвить*, *убить, устранить* и др.), *про-* (*пропустить*, *проморгать*, *прозевать* и др.), *о-/об-* в значении ‘ошибиться’ (*опечататься*, *обознаться*, *оговориться* и др.). Все эти модели можно считать особыми словообразовательными конструкциями (в терминах Грамматики Конструкций, см. [43]), в которых в качестве якоря выступает приставка, а в качестве переменной — глагольная основа. Особенность этих конструкций заключается в том, что они выражают не (квази)грамматические, как большинство подобных конструкций [44]), а лексические значения.

#### 2.6.2.2 Фреймы Переключения, Замены, Замещения и Компенсации

«Центральная» часть поля ‘менять’ — фреймы Переключения, Переключения между объектами, Замены, Компенсаторной замены, Замещения и Компенсации — в русском языке распределяется между глаголами *сменить* и *заменить*. Основное различие между этими двумя приставочными дериватами сводится к параметру внутренней или внешней причины изменения: изменения, происходящие по инициативе субъекта (ср. Переключение, Переключение между объектами, Замещение), обозначаются глаголом *сменить*, а все вынужденные замены (Замена, Компенсаторная замена, Компенсация) — глаголом *заменить*. Такое лексическое распределение в целом нехарактерно для языков нашей выборки. Как правило, языки «реагируют» на количество участников, семантических или синтаксических. Так, например, в адыгейском любая ситуация, в которой субъект меняет один объект на другой, будь то Переключение или Замена (ср. ‘сменить врача’, ‘заменить лампочку’, ‘поменять линолеум на паркет’), обозначается глаголом *zebleχʷən*, а ситуации, в которых объект замещается самим субъектом (ср. ‘сменить охранника на посту’, ‘подменить заболевшую сотрудницу’), описываются конструкциями с существительным *č̣əṗe* ‘место’. Норвежский реагирует на синтаксическое выражение участников: глагол с двумя валентностями (субъекта и объекта) *skifte* обозначает все ситуации замены исходного объекта на однотипный, не требующий специального синтаксического выражения (ср. ‘сменить врача’, ‘заменить лампочку’), а ситуации,в которых старый и новый объекты различаются и потому должны быть выражены поверхностно (ср. ‘поменять линолеум на паркет’, ‘сменить кожаные чехлы на пластиковые футляры’ и под.), обозначаются глаголом *bytte.*

#### 2.6.2.3 Фреймы обмена

Следующая зона — это фреймы с наибольшим количеством участников, предполагающие изменения не самих участников и их состава, а посессивных отношений между ними, т. е. фреймы Обмен и Реципрокальный обмен. Это ситуации, в которых два субъекта совершают обмен принадлежащими им объектами. Объекты могут быть разными, и тогда это будет фрейм собственно Обмена (ср. *выменивать у соседей овощи и фрукты на молоко и сыр*), а могут быть одинаковыми — такие ситуации мы относим к фрейму Реципрокального обмена (ср. *Дети поменялись игрушками*). Как видно уже по самим условным обозначениям этих фреймов, в русском языке они покрываются, прежде всего, дериватом от основного для этого поля корня *мен*- с приставкой *об*-, т.е. глаголом *обменять(ся)*. Кроме того, глагол *обменяться* сочетается с именами действий, приобретая при этом значение, близкое к грамматическому, и выражая реципрокальность действия, названного именем-объектом, ср. *обменяться рукопожатиями =* ‘пожать друг другу руки’, *обменяться взглядами* = ‘посмотреть друг на друга’. Этим же глаголом обозначается обмен одной валюты на другую (*обменять доллары на рубли*), но не размен крупных купюр на более мелкие (*разменять сто рублей*). Еще один глагол в этой группе — *выменять* — обозначает исключительно Обмен разными объектами, не допускает реципрокальных интерпретаций и не сочетается с постфиксом -*ся*. Более ранней, по-видимому, для *выменять* была модель управления, в которой в позиции прямого дополнения выступает новый Объект. Об этом свидетельствуют, во-первых, то, что в панхроническом подкорпусе НКРЯ до середины XIX в. фиксируется только этот вариант управления, во-вторых, то, что эту модель отражают толкования в ряде традиционных словарей. Однако со второй половины XIX в. в корпусе встречаются контексты, в которых прямое дополнение явно соотносится с отдаваемым при обмене объектом. К числу ранних примеров такого рода относится следующий фрагмент из «Леса» А.Н. Островского, см. (35):

1. *Жаль, фрака нет; был фрак, да я его в Кишиневе на костюм Гамлета* ***выменял****.* [А. Н. Островский. Лес (1871)]

Другой глагол — *променять* — тоже ограничен фреймом Обмена разными объектами и не сочетается с постфиксом -*ся*. При *променять* один объект оформляется как прямое дополнение, другой вводится предлогом *на*. Однако, в отличие от *выменять*, *променять* всегда помещает в позицию прямого дополнения исходный Объект, а не новый. Кроме того, такой обмен обычно представляется как неэквивалентный: исходный объект оценивается говорящим выше, чем новый, поэтому действия субъекта осуждаются. Оценочность глагола *променять* особенно заметна в контекстах с абстрактными объектами (36).

1. *Ах, мессир, можно ли* ***променять*** *холостую свободу на тягостное ярмо!* [М. А. Булгаков. Мастер и Маргарита, часть 2 (1929-1940)]

#### 2.6.2.4 Бесприставочный глагол *менять(ся)*

Бесприставочный глагол *менять(ся)* можно считать доминантным в этом поле. Он употребляется во фреймах Модификации (*с возрастом люди меняются*), Переключения (*менять мужей как перчатки*), Переключения между объектами (*менять шило на мыло*), Замены (*менять лампочку*), Обмена (*менять козла на барана*), Реципрокального обмена (*меняться игрушками*). Недоступен глаголу *менять(ся)* фрейм Превращения (*\*вода меняется в лёд*, *\*менять принцессу в лягушку*), который, как уже было отмечено, синхронно не покрывается ни одним глаголом с корнем *мен-*. Интересно, однако, что глагол *менять(ся)* также почти не используется во фреймах Замещения (?*один охранник меняет другого*) и Компенсации (\**менять заболевшую учительницу*). Судя по корпусным данным, сочетания глагола *менять* с одушевленными существительными в позиции прямого объекта сводятся к контекстам Переключения (37).

1. *С такой же завидной регулярностью они* ***меняли*** *спутниц, каждый раз приводя новых.* [Анна Берсенева. Полет над разлукой (2003-2005)]

Кроме того, глагол *менять(ся)* не характерен для контекстов Компенсаторной замены (*\*если сахара нет, меняем его медом*). По всей видимости, это коррелирует с тем, что у него нет модели управления с новым объектом в инструменталисе — основной для фрейма Компенсаторной замены. При этом в конструкциях с новым объектом, который вводится предлогом *на*, даже в контекстах Компенсаторной замены глагол *менять(ся)* кажется более приемлемым, ср. *?если сахара нет, меняем его на мед*.

#### 2.6.2.5 Доминантный глагол *поменять(ся)*

Как уже было сказано, глагол *поменять*(*ся*) покрывает большую часть семантического поля ‘менять’. Он регулярно используется в контекстах Модификации: как без присутствия в ситуации субъекта в декаузативном варианте (*цвет палитры немного поменялся*) или (*расписание уроков опять поменяли*), так и с выраженным каузатором изменения (38).

1. *А ей так у них понравилось, она так загорелась, что решила полностью* ***поменять судьбу****, связав себя с новой музой.* [Инна Калабухова. Приключение длиною в сорок пять лет // «Ковчег», 2013]

Глагол *поменять* используется в контекстах Переключения (*поменять причёску*) и Переключения между объектами (*поменять отчество на русское*). Доступен он и в контекстах Замены: как с совмещённым выражением двух Объектов (*поменять замки*), так и с поверхностным выражением исходного и нового Объекта (*поменять батарейки на новые*).

Глагол *поменять(ся)* также выражает все фреймы, связанные с обменом. В значении Обмена разными объектами в наблюдаемых в корпусе примерах используется переходный глагол *поменять*, который способен присоединять три (39) или четыре синтаксических аргумента, включая второго субъекта обмена, управляемого предлогом *у* (40).

1. *Я бы с радостью, говоря откровенно,* ***поменял*** *свою водку на твой йод.* [Юлиан Семенов. Семнадцать мгновений весны (1968)]
2. *Землю можно будет* ***поменять*** *у них черт знает на что <…>* [Е. Петров. Остров мира. Комедия в 4 актах (1947)]

Возвратный глагол *поменяться* выражает Реципрокальный обмен конкретными (*мы решили поменяться прочитанными книжками*) и абстрактными объектами (*он предложил поменяться номерами*).

Почти во всех контекстах *поменять*(*ся*) конкурирует по крайней мере с одним приставочным однокоренным глаголом. При рассмотрении того, как *поменять*(*ся*) соотносится с другими приставочными вариантами, значимыми являются диахронические изменения в его употреблении. Этапы развития этого деривата просматриваются на материале словарей разного времени.

В течение XX и начала XXI веков мы наблюдаем расширение употребления лексемы *поменять*(*ся*). Это расширение происходит поэтапно. В начале наблюдаемого периода этот глагол в целом воспринимается как разговорный и имеет только значение Обмена и Реципрокального обмена. Далее значение расширяется: глагол начинает употребляться во фреймах Модификации, Переключения и Замены. При этом употребления в недавно появившихся интерпретациях в некоторых словарях интерпретируются как разговорные, тогда как более старые значения признаются принадлежащими нейтральному стилю. Факт расхождения в наборах значений, выделяемых в словарях, и пометах при этих значениях также хорошо согласуется со сценарием недавнего расширения значения.

#### 2.6.2.6 Устаревающий глагол *переменить(ся)*

Глагол *переменить(ся)* в современном русском языке выходит из употребления. По данным НКРЯ его частотность сокращается уже начиная с XIX века, а во второй половине XX — начале XXI века он почти полностью вытесняется однокоренными *сменить(ся)*, *изменить(ся)* и др. в большинстве коллокаций.

В основном подкорпусе НКРЯ в текстах XX и XXI века *переменить* преимущественно описывает ситуации, относящиеся ко фреймам Модификации (*переменить мнение*) и Переключения (*переменить место жительства*), иногда Замены (41). Особенно характерны для него контексты, в которых в современном русском языке возможна вариативность *сменить*/*изменить*, ср., например, (42).

1. *Около каменоломни шина сдала и дальше пришлось идти пешком. Велосипед я взяла еще на* *месяц*, *но* *только* *он* *будет* *готов* *завтра*: *мне* *обещали* ***переменить*** *камеру*, *а* *потом* *я* *только* *вчера* *заметила*, *что* *на* *переднем* *колесе* *у* *меня* *не* *хватает* *двух* *спиц*. [И. Н. Кнорринг. Дневник (1924)]
2. *Он был немцем, но в начале первой мировой войны* ***переменил фамилию (***OK***изменил /*** OK***сменил фамилию)*** *на Астров, принял православие.* [Д. С. Лихачев. Воспоминания (1995)]

Если глагол совершенного вида *переменить* ещё не совсем вышел из употребления, то его видовая пара несовершенного вида *переменять* практически полностью перестала использоваться ещё век назад. Однако в период до ХХ века *переменять* покрывал больший спектр ситуаций, чем *переменить* в рассматриваемый нами более поздний период (с ХХ века). Помимо ситуаций Модификации, Переключения и Замены, этому глаголу были также доступны контексты Обмена разными объектами (43), Реципрокального обмена (44) и Замещения (45).

1. *Зато не далее как через два года, с Божьей помощью, церковь была сооружена вчерне. Съездили мы с тобой в Ромны, на Ильинскую ярмарку,* ***переменили старое серебро на церковные вещи****. А еще через год в новом храме началось и служение!* [В. П. Авенариус. Гоголь-гимназист (1897)]
2. *В назначенный день Образцов зашел ко мне часов в пять вечера, мы сейчас же вышли из дому и зашли в одну кондитерскую. Тут, взяв отдельную комнату,* ***мы переменились костюмами****. Образцов надел мою штатскую пару, а я его кадетский мундир, с тесаком на портупее и с кивером на голове.* [А. М. Достоевский. Воспоминания (1896)]
3. *А ныне, государи, указано* ***переменить Лукьяна Остафевича*** *иному воеводе.* [Крестьяне дворцовых арзамасских волостей и разных сел и деревень Троицкого присуда Арзамасского у. царям об оставлении их в ведении осадного арзамасского воеводы Л. О. Сытина (1682—1694)]

Особенно обращает на себя внимание тот факт, что, согласно Словарю русского языка XVIII века [45], *переменить* раньше покрывал и ситуации превращения, в современном языке недоступные ни одному глаголу с корнем *мен-*. Ср. примеры из словаря и из корпуса, в которых глагол имеет модель управления, свойственную современному *превратить* (V NP.Acc в NP.Acc). С этими примерами согласуется, во-первых, тенденция современного *переменить* обозначать только существенные изменения с заданным результирующим состоянием в контекстах Модификации (именно такие контексты ближе всего к Превращению), а во-вторых, ограниченная фреймами Модификации и Превращения словообразовательная конструкция с приставкой *пере-*, см. [раздел 2.6.2.1](#sec-frame-modifikatsii-i-prevrasheniya).

1. *<Стихотворец> твердо уповал, что он ету Тигресу своими елегиями и мадригалами* ***перемѣнит в овечку****.* [Словарь русского языка XVIII века, стр. 94]
2. *Естьли которой женщинѣ мущина съ оною часто свиданіе имѣющій понравится, то хотя бы она къ такому иной склонности и не чювствовала, кромѣ истиннаго почтенія къ его достоинству, но ежели такой человѣкъ будетъ хитръ и смѣлъ, то онъ скоро* ***ея почтеніе въ любовную страсть перемѣнить*** *можетъ.* [Ф. А. Эмин. Непостоянная фортуна или похожденіе Мирамонда. Ч.2 (1763)]
3. *Судбинѣ то свойственно, чтобъ нещастныхъ всегда гонить; и когда она меня столь сильно возненавидѣла, то тѣмъ только чувствительнѣйшую тоскующему моему сердцу здѣлать могла досаду, и неисцѣлимую въ ономъ отворить рану, что дражайшую вашу любовь, которая полумертвой мой духъ до сихъ поръ оживляла, въ смертоносную* ***перемѣнила*** *ненависть.* [Ф. А. Эмин. Непостоянная фортуна или похожденіе Мирамонда. Ч.2 (1763)]

# 3. Частные исследования русского языка и других языков России

## 3.1 Нестандартные конструкции с числительными в русской речи билингвов

### 3.1.1 Введение

В ходе проекта исследовалась вариативность в конструкциях с числительными в речи билингвов из разных регионов России на русском языке. Используя корпусный подход, анализирующий данные спонтанной речи 181 участника, родившихся между 1920 и 1997 годами, исследование рассматривает факторы, влияющие на эту вариативность, включая уровень образования, год рождения, тип числительного и коллокационная сочетаемость числительных с существительными. Результаты показывают, что вариативность значительно связана с частотой использования и владением языком как вторым, причем коллокационная сочетаемость числительных с существительными играет ключевую роль. Хотя влияние контакта не находят убедительного подтверждения, результаты дают представление об усвоении языка как о процессе, основанном на опыте, и имеют значение для понимания языковой вариативности в ситуациях контакта. Мы предполагаем, что кажущиеся заимствованными модели могут быть лучше объяснены другими социолингвистическими и приобретением факторами.

### 3.1.2 Конструкции с числительными в родных языках билингвов

Родные языки, которыми владеют билингвы, принадлежат к четырём разным языковым семьям: 1) нахско-дагестанская языковая семья (в корпусе дагестанского русского) –– аварский, андийский, багвалинский, тукитинский, акушинский, ицаринский, мегебский, муиринский, цудахарский, лакский, арчинский, рутульский, табасаранский, лакский; 2) тюркская языковая семья –– кумыкский и азербайджанский из корпуса дагестанского русского, а также башкирский и чувашский, представленные в одноимённых корпусах; 3) уральская языковая семья –– бесермянский, карельский, марийский; 4) индоевропейская языковая семья (индоарийская ветвь) –– кэлдэрарский цыганский.

Практически во всех рассмотренных языках, кроме карельского и цыганского, в конструкции с числительными существительное не имеет специальных морфологических показателей. Кроме того, насколько можно судить по имеющимся описаниям, ни в одном языке нет отличий между конструкциями с паукальными и непаукальными числительными.

В грамматиках нахско-дагестанских языков обычно отмечается, что форма существительного не маркируется, см. [46: 188] об аварском, [47: 138–139] о зиловском диалекте андийского языка, [48: 159] о багвалинском, [49: 45] об ицаринском, [Казенин 2013: 29] о лакском и [50] о табасаранском. В рассмотренных тюркских языках существительное также не маркируется по числу, как отмечено в [51: 51] и некоторых других работах. Пример из кумыкского языка: *бир китап* ‘одна книга’, *беш китап* ‘пять книг’ [52: 233]. Среди уральских языков то же касается бесермянского и марийского. Пример из марийского: *ик пӧрт* ‘один дом’, *кандаш пӧрт* ‘восемь домов’ [53: 104].

Регулярное маркирование существительного в конструкции с числительным характерно лишь для двух языков –– карельского (уральская языковая семья) и кэлдэрарского цыганского.

В карельском языке в конструкции с числительными используются существительные в партитивном падеже как при паукальных, так и при непаукальных числительных: *poika* ‘мальчик’, *kolme poiku-a* ‘три мальчика’, *kymmenen poiku-a* ‘десять мальчиков’ [54, P. 140]

В языке цыган-кэлдэрар, интервью с которыми включены в корпус цыганского русского, числительные сочетаются с существительными во множественном числе, например: *jèkh matòr-a* ‘одна машина’ –– *dùj matòr-i*» ‘две машины’ [55: 464].

Таким образом, в родных языках респондентов конструкции с числительными, как правило, устроены проще, чем в русском языке. Маркирование существительных в некоторых из них возможно как дополнительный вариант в определенных типах контекстов.

### 3.1.3 Разметка корпусных данных

На первом этапе каждое высказывание было размечено как стандартное или нестандартное. Для дальнейшей обработки данных учитывались высказывания только тех респондентов, у которых нестандартная конструкция зафиксирована хотя бы один раз. Итоговая выборка включала 1748 высказываний. Доля нестандартных конструкций оказалась выше всего в корпусах дагестанского русского, чувашского русского и марийского русского (см. ([рисунок 3.1](#fig-7))).

|  |
| --- |
| Рисунок 3.1 – Доля нестандартных конструкций с числительными в билингвальных корпусах. |

Каждый пример был снабжён социолингвистической информацией (пол, год рождения, уровень образования респондента, родной язык и языковая семья, к которой он относится). Лингвистическая разметка материалов включала несколько параметров. Отмечался тип числительного (паукальное, собирательное или иное). Указывалась лемма существительного.

Для нестандартных примеров указывалась использованная форма существительного. Учитывая, что почти во всех родных языках билингвов в конструкции с числительными существительное не маркируется ни по числу, ни по падежу, можно было ожидать в конструкциях на русском существительное именительного падежа единственного числа. Такие примеры имеются: *четыре класс* –– лакско-русский билингв, *три доярка* –– чувашско-русский билингв и др. Но вместе с тем оказалось, что возможны и другие формы, например, именительный множественного числа (*три ягнята* –– бесермянско-русский билингв, *трое сыновья* –– марийско-русский билингв). При паукальных числительных во всех корпусах, кроме бесермянского, есть также примеры с существительными в родительном падеже множественного числа (*тридцать два учителей нас было* –– башкирско-русский билингв). В дагестанском и башкирском корпусах при паукальных числительных также отмечены существительные в родительном падеже единственного числа (*пять дня* –– табасаранско-русский билингв). Более подробно результаты разметки по данному параметру представлены в ([рисунок 3.2](#fig-8)). На нем нет материалов цыганского корпуса, поскольку в нём имеется лишь четыре нестандартных примера, во всех используется существительное в именительном падеже множественного числа (*три девчата*). В этом случае можно предполагать интерференцию с родным языком, но число примеров кажется недостаточным, чтобы с уверенностью это утверждать.

|  |
| --- |
| Рисунок 3.2 – Распределение падежно-числовых форм существительного в нестандартных конструкциях с разными типами числительных. |

Графики в ([рисунок 3.2](#fig-8)) показывают, что во всех корпусах со всеми типами числительных преобладает использование существительного в именительном падеже. Существительные в родительном падеже встречаются в нестандартных конструкциях заметно реже. Таким образом, интерференция является не единственной причиной появления нестандартных форм в конструкциях с числительными. Такое влияние можно предполагать лишь в половине всех примеров (там, где используется существительное в именительном падеже единственного числа, а также в примерах из цыганского языка).

При разметке данных было замечено, что ряд очень частотных существительных (*час*, *день*, *килограмм* и т.п.) реже, чем другие, используются в нестандартной форме. Чтобы оценить влияние этого фактора был использован дополнительный параметр –– *коэффициент коллокационности* (dice coefficient). Сначала из Национального корпуса русского языка [НКРЯ](ruscorpora.ru) была извлечена мера коллокационности для каждого сочетания *числительное + существительное* в исследуемой выборке. Затем все коэффициенты для одной леммы существительного были усреднены по всем числительным и объединены с основным набором данных для использования их в логистической модели.

### 3.1.4 Cтатистический анализ и его результаты

В качестве основного метода использовалась логистическая регрессия со смешанными эффектами. Модели были созданы в программной среде R [R Core Team 2024] с помощью пакета lme4 [56]. Модель включала следующие переменные (основные эффекты): тип числительного, коэффициент коллокационности, пол, год рождения, образование. Случайными эффектами были переменная *респондент*, вложенная в переменную *языковая семья*. Результаты статистического моделирования представлены на следующих графиках ([рисунок 3.3](#fig-9)).

|  |
| --- |
| Рисунок 3.3 – Вероятность нестандартной формы в зависимости от типа числительного, коэффициента коллокационности и уровня образования респондентов. |

Каждый график соответствует группе числительных (слева направо): собирательные, паукальные и прочие. По оси *х* отмечен коэффициент коллокационности (dice coefficient), по оси *у* –– вероятность нестандартных контекстов. Цветом обозначены уровни образования: красным –– высшее, синим –– среднее или начальное. Как и ожидалось, вероятность нестандартной формы ниже у респондентов с более высоким уровнем образования. Определенную роль играет тип числительного: для собирательных вероятность использования нестандартной формы самая высокая (но следует учесть, что контекстов с ними в выборке было меньше всего, поэтому на графике большие доверительные интервалы). В конструкциях с высоким коэффициентом коллокационности (*два часа*, *пять километров*) вероятность нестандартной формы значительно ниже.

Далее был рассчитан обобщённый коэффициент детерминации (generalized R squared, см. [57]). Он позволяет выстроить иерархию значимости переменных, которая в данном случае выглядит следующим образом:

коллокационность > уровень образования > год рождения > тип числительного > пол

Наиболее значимой переменной оказалась коллокационность (совместная встречаемость единиц). Чем выше степень коллокационности, тем ниже вероятность появления нестандартной формы в речи билингва. Это можно объяснить с точки зрения когнитивного подхода. Высокочастотные выражения с большей вероятностью попадают в инпут билингвов, чаще успешно воспроизводятся и могут храниться в ментальном лексиконе как единые блоки. В случае с менее частотными сочетаниями неносители языка, вероятно, конструируют фразы «на ходу», что увеличивает вероятность появления нестандартных форм.

## 3.2 Опущение предлогов в речи русско-чувашских билингвов

### 3.2.1 Введение

Обусловленные контактом нестандартные грамматические явления в русской речи билингвов из разных регионов России изучались многими исследователями. Авторы [58] предложили краткий обзор нестандартных явлений в дагестанском варианте русского языка, [59] фокусируется на грамматических особенностях речи эрзянских билингвов, в [60] исследуется русская речь энцев. Некоторые исследования посвящены определенной области грамматики (см. в [61] обзор фонетических и морфологических особенностей в русской речи носителей эрзянского и мокшанского языков) или специфическим нестандартным явлениям, например, нестандартному гендерному маркированию (см. в [62] о подверженном контактному влиянию русском языке Северной Сибири и Дальнего Востока) или выпадению предлогов (см. [63] о дагестанском русском языке).

Данное исследование затрагивает русскую речь чувашских билингвов (далее –– чувашский русский язык), которая является примером разновидности русского языка, находящегося в контакте с тюркским языком. Существует множество описаний чувашского языка, в том числе [64], [65], [66], [67]. Особое значение для данного исследования имеет работа [68], в которой автор перечисляет различные грамматические явления в чувашском языке и пытается сопоставить их с грамматическими явлениями в русском языке; эта стратегия также прослеживается в [61]. Русский язык, на котором говорят в Чувашии, уже был исследован в работе [69], где были рассмотрены фонетические, морфологические и лексические особенности лекта. Мызников рассматривает эту разновидность русского языка с диалектологической точки зрения и не исследует конкретные явления [69]. Насколько нам известно, это единственная опубликованная работа, описывающая чувашский русский язык; более того, нет информации об исследованиях в области морфосинтаксиса чувашского русского языка. Эта область особенно интересна, учитывая различия между морфосинтаксисом в славянских и тюркских языках, которые потенциально могут привести к нестандартным явлениям. На основе данных из корпуса русской речи Чувашии [70] мы пытаемся выяснить, какие факторы влияют на явление опущения предлогов в чувашском русском и связаны ли эти факторы с языковым контактом.

### 3.2.2 Методы и данные

Феномен опущения предлогов хорошо изучен для различных языков. Это явление часто упоминается в связи с новогреческим, где опущение предлога *se* ‘в’ встречается в определенных локативных и указательных контекстах [71]. Тем не менее в контактных вариантах русского языка это явление не ограничивается такими контекстами: там опущение предлогов встречается также во временных и комитативных предложных группах. Авторы [72] утверждают, что в английском языке предлоги *at* (локация) и *to* (направление), в отличие от *from* (источник), часто опускаются. Обсуждая опущение предлогов в русском языке, на котором говорят в Дагестане, исследователи [63] представляют следующий список типичных объяснений опущения предлогов:

1. фонотактические ограничения, с которыми сталкиваются носители второго языка (L2);
2. морфосинтаксическая интерференция с другими языками;
3. отсутствие семантической маркированности;
4. исключительная синтаксическая структура.

Авторы опровергают возможность того, что пункты 1 и 4 имеют отношение к их данным; тем не менее мы включили все представленные объяснения в методологию исследования.

Данное исследование основано на данных, извлеченных из корпуса русского языка, на котором говорят в Чувашии [70]. Он состоит из интервью с 14 респондентами. Интервью проводились в деревне Малое Карачкино, Чувашия. Для респондентов первым языком (L1) является верхний диалект чувашского языка, вторым (L2) –– русский язык. В описании корпуса говорится, что овладение L2 респондентами произошло в подростковом или раннем взрослом возрасте [70]. Объем корпуса оценивается в 36 140 токенов, не считая речи интервьюеров. Корпус позволяет выполнять базовый поиск по токенам, леммам и тегам, а также выполнять поиск по CQL. Каждый результат поискового запроса сопровождается социолингвистической информацией о говорящем: пол, год рождения и уровень образования.

Корпус русской речи Чувашии [70] транскрибирован в соответствии со стандартным письменным русским языком, поэтому нестандартные явления, в том числе опущение предлогов, не всегда отражаются в транскрипции. Таким образом, все предложные группы были извлечены двумя способами: путем автоматического и ручного (поиск предложных групп с опущенными предлогами в текстах) извлечения всех контекстов с предлогами. Затем соответствующие фрагменты аудиозаписей были прослушаны. Случаи, в которых было невозможно определить, имело ли место опущение предлога или нет (из-за плохого качества звука, фонового шума или других технических проблем), были удалены из выборки. Наконец, был получен набор данных, состоящий из 2529 контекстов. Все контексты в базе данных были дополнительно аннотированы в соответствии с параметрами, представленными на [рисунок 3.4](#fig-15).

|  |
| --- |
| Рисунок 3.4 – Значения внутри набора данных. Серые сегменты указывают на то, что данных нет. |

Столбцы “лемма следующей словоформы” и “коэффициент коллокационности” имеют особо важное значение для выявления связи между выпадением предлога и степенью сочетаемости с ним существительного в предложной группе. Идея о том, что нестандартное явление реже встречается в конструкциях, в которых слова часто встречаются вместе (например, в устойчивых би-граммах), была выдвинута [73] при исследовании нестандартных числительных в билингвальных разновидностях русского языка. Авторы утверждают, что очень часто встречающиеся конструкции, скорее всего, будут храниться в виде целых единиц в ментальном лексиконе, поэтому носители L2 русского реже используют нестандартные формы в контексте частых сочетаний числительных и существительных. Чтобы проверить эту гипотезу на нашей выборке, мы использовали данные Национального корпуса русского языка[[1]](#footnote-90), который использует коэффициент dice в качестве меры коллокационности ([57]). В тех случаях, когда словосочетание не было найдено в НКРЯ, было установлено значение коэффициента коллокационности 0.

### 3.2.3 Результаты

Всего было рассмотрено 2529 предложных групп, из которых в 340 (13,44%) предлог был опущен. Как показано на [рисунок 3.5](#fig-16), не все предлоги имели тенденцию к выпадению. Только семь предлогов были пропущены хотя бы один раз: *в*, *у*, *на*, *с*, *из*, *до*, *через*. Последние два предлога были опущены в выборке только по одному разу. Также стоит отметить, что *в* и произносится, и опускается значительно чаще, чем другие предлоги: 291 (33,64%) из 865 предлогов *в* были опущены (см. [рисунок 3.6](#fig-17)).

|  |
| --- |
| Рисунок 3.5 – Количество произнесений и пропусков для каждого предлога. |

|  |
| --- |
| Рисунок 3.6 – Количество произнесений и пропусков для каждого предлога, склонного к выпадению. |

Данные о предлогах, которые не выпадают, мы посчитали избыточными, поэтому последующий анализ был проведен на основе набора данных, который включает только те предлоги, которые имеют тенденцию к выпадению; этот набор данных состоит из 2069 предложных групп.

Чтобы выяснить, какие из факторов, приведенных на [рисунок 3.4](#fig-15), влияют на опущение предлогов, мы мы применим логистическую регрессионную модель со смешанными эффектами. Для построения более оптимальной модели следует уменьшить число категориальных факторов со множеством переменных: начальный звук следующей словоформы, ее падеж, семантика предложной группы.

Выпадение простых неслоговых предлогов как *в* в билингвальных разновидностях русского языка часто объясняется влиянием фонотактических ограничений L1. Интересно, что в чувашском русском также есть слоговые предлоги, которые имеют тенденцию к выпадению (*у*, *до*). Хотя это указывает на меньшую значимость фонотактики, все еще остается вероятность влияния ограничения на выпадение неслоговых предлогов.

Следует посмотреть на распределение начальных звуков в следующих после предлога словоформах. На [рисунок 3.7](#fig-18) иллюстрируется отношение числа примеров с опущением предлога к числу примеров без опущения. Существует два способа анализа [рисунок 3.7](#fig-18). Рассматривая вертикальные линии (предлоги), можно сделать вывод, что некоторые предлоги (в частности, *в*) имеют одинаковую вероятность выпадения в контексте практически любой фонемы, представленной в данных. Большинство предлогов имеют определенные фонемные предпочтения: для *у* это [i], для *с* это, что неудивительно, [z], для *на* это [tʃʲ], для *из* –– [s] и [u]. Еще один способ взглянуть на [рисунок 3.7](#fig-18) –– по горизонтали: более светлые линии указывают, в контексте каких фонем чаще всего выпадают предлоги. Согласно рисунку, линии с наиболее светлыми фрагментами соответствуют взрывным звукам: [b], [p]/[pʲ], [d]/[dʲ].

|  |
| --- |
| Рисунок 3.7 – Отношение числа примеров с опущением предлога к числу примеров без опущения для пар предлог-начальный звук следующей словоформы. |

Итак, во-первых, важно добавить в логистическую регрессию параметр самого предлога, поскольку некоторые предлоги имеют тенденцию выпадать в контексте любой фонемы; игнорирование этого факта привело бы к менее точным вычислениям. Во-вторых, вместо того, чтобы добавлять в регрессию множество категориальных переменных, я добавлю только бинарную переменную: взрывной звук далее или нет.

Еще один категориальный фактор с большим количеством переменных –– это падеж следующей за предлогом словоформы. На [рисунок 3.8](#fig-19) именительный падеж выделяется больше всего среди других: примерно в трети всех примеров, содержащих именительный падеж, также присутствует опущение предлога. Несмотря на то, что в большинстве пропущенных примеров используется локатив, это может быть просто связано с частотностью падежа.

|  |
| --- |
| Рисунок 3.8 – Обзор падежей, следующих за предлогом словоформ. |

Наконец, было принято решение преобразовать категориальный фактор семантики в бинарный фактор ‘неопределенная семантика’/‘определенная семантика’. Таким образом проверяется корреляция между отчетливостью семантики контекста и вероятностью пропуска предлога в указанном контексте. Распределение семантики предложных групп показано на [рисунок 3.9](#fig-20)

|  |
| --- |
| Рисунок 3.9 – Обзор семантики предложных групп. |

### 3.2.4 Статистическое моделирование

Чтобы оценить значимость выделенных факторов, мы применили логистическую регрессионную модель со смешанными эффектами, используя пакет R (R Core Team 2024) lme4 [56]. Следуя процедуре пошагового выбора переменных в обратном порядке [74], мы пришли к оптимальной модели (см. [таблица 6.14](#tbl-6)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 3.1 – Переменные в логистической регрессии   | Переменная | Тип переменной | Тип эффекта | | --- | --- | --- | | предлог | категориальная | фиксированный | | спикер | категориальная | случайный | | возраст | непрерывная | фиксированный | | образование (лет) | непрерывная | фиксированный | | семантика неопределенная | бинарная | случайный | | начальный звук следующей словоформы взрывной | бинарная | случайный | | падеж следующей словоформы именительный | бинарная | фиксированный | | следующая словоформа имеет стандартное падежное маркирование | бинарная | фиксированный | | коэффициент коллокационности | непрерывная | фиксированный | |

Факторы, которые оказались значимыми, представлены в [таблица 6.15](#tbl-7).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 3.2 – Значимые факторы в логистической регрессии   | Фактор | Уровень значимости | | --- | --- | | предлог *в* | p < 0.05 | | образование (лет) | p < 0.1 | | начальный звук следующей словоформы взрывной | p < 0.000 | | коэффициент коллокационности | p < 0.05 | | следующая словоформа имеет стандартное падежное маркирование | p < 0.000 | | падеж следующей словоформы именительный | p < 0.000 | | семантика неопределенная | p < 0.01 | |

### 3.2.5 Выводы

Опущение предлога *в* в чувашском русском не следует стандартным социолингвистическим тенденциям (например, в описанных данных пожилые/менее образованные носители языка не используют нестандартный вариант чаще, чем другие). Более того, факторы, которые увеличивают или уменьшают вероятность выпадения для большинства предлогов, –– например, начальный звук следующей за предлогом словоформы –– не влияют на выпадение *в*. Полученные данные подтверждают гипотезу о том, что фонотактические ограничения и отсутствие семантической маркированности являются причиной выпадения некоторых предлогов в чувашском русском языке. Аналогичные факторы, влияющие на опущение предлогов, были выявлены в работе [63].

Интересно, что некоторые единичные примеры, мотивация которых неясна, следуют общим тенденциям чувашской русской морфосинтаксической системы (например, предлоги, которые были опущены только один раз, были опущены в контексте семантики направления или времени). Другими словами, носители чувашского русского языка распространяют тенденции, которые изначально были мотивированы контактом или неполным усвоением L2, на новые контексты. Это может свидетельствовать о динамичности морфосинтаксической системы чувашского русского языка.

## 3.3 Пограничные явления в речи русско-казахских билингвов

В рамках решения задачи проекта 2025 года о синхронном изучении пограничных языковых явлений на материале нестандартных вариантов русского языка были выявлены особенности функционирования глаголов движения в метафорических контекстах на материале речи русско-казахских билингвов.

Использование глаголов движения в метафорических контекстах –– то есть fictive / abstract motion, типологически универсально: для самых разных абстрактных процессов и состояний оно может считаться базовой метафорой, хорошо узнаваемой носителями языков. Однако реализуется эта концептуализация в конкретных языках в соответствии с собственными правилами и ограничениями. Это хорошо видно на примере анализа учебных текстов казахско-русских билингвов. Действительно, с одной стороны, изучающие русский язык как иностранный или неродной (L2), отталкиваясь от этих базовых когнитвных принципов, регулярно создают их и в своей русской речи, а с другой — исходя в том числе и из своего опыта использования метафор движения в родном языке — игнорируют частные правила концептуализации fictive motion, действующие в русском. Соответственно, часто эти метафоры носят отпечаток «нестандартности».

Сосредоточившись на анализе нестандартных метафор неспецифицированного движения самой общей семантики (*идти / ходить / входить / выходить*), зафиксированных в текстах казахско-русских билингвов, авторы исследования стремились выяснить, какого рода отклонения от нормы и почему допускают русско-казахские билингвы. Выявление зон отклонения и их причин интересно в типологическом отношении, чтобы обозначить контуры будущих типологических анкет для подробного исследования этой области в разных языках. Одновременно оно необходимо и для уточнения методик обучения русскому — если требуется устранить «зазор» между стандартной и нестандартной нормой у учащихся.

Рассматривая нестандартные метафорические контексты глаголов *идти* и *ходить,* авторы заметили, что русско-казахские билингвы склонны использовать глагол *идти* как универсальное обозначение любых проявлений динамики, а не только целенаправленных процессов, как это принято в русском языке. Что касается *ходить*, то этот глагол тяготеет к другому метафорическому «полюсу» — обозначению состояний, причем в текстах билингвов преодолевается ограничение, действующее в русском языке, — обозначать только длительные и имеющие яркие внешние проявления состояния.

Материал показал и то, что в метафорических контекстах важно, какому участнику ситуации отводится роль траектора: в зависимости от выбора траектора метафорический контекст также может оцениваться как стандартный или нестандартный. Например, в стандартных контекстах приближение ситуации как метафорического траектора к субъекту-наблюдателю представляет субъект как пассивного участника ситуации, а траектор — как приближающуюся неизбежность. Если в действительности субъект действует осознанно и контролирует ситуацию, это приводит к своего рода лингвистическому конфликту и квалифицируется как нестандартное речевое поведение.

Наконец, выбор глаголов направленного движения в метафорическое пространство или из него определяется тем, как концептуализируется само это пространство — как закрытое или как открытое. Для автора-билингва этот выбор осложняется тем, что в русском языке сосуществуют очень близкие метафорические модели, связанные с глаголами *входить / войти* и *выходить / выйти* –– это может быть предпосылкой для сбоев в выборе метафорической модели в целом и глагола движения в частности.

Материал касался только очень общих глаголов движения — то есть таких, в которых нет спецификации способа перемещения. Это ограничивает параметры варьирования для нестандартных носителей и сводит их к трем основным: во-первых, статика VS. динамика как результат развития метафорических контекстов, во-вторых, выбор траектора, в-третьих, концептуализация SOURCE / GOAL пространства как открытого или закрытого. При этом нестандартные метафоры движения как результат языкового творчества билингвов могут быть подвержены влиянию сразу нескольких факторов: это может быть влияние родного языка (например, семантическое своеобразие его глаголов движения); предпосылки разного рода в целевом языке (такие как наличие семантически близких метафорических моделей) — но также и универсальные, надъязыковые стратегии, как, например, замена менее частотного глагола более частотным глаголом движения.

Обращая внимание на случаи, когда родной язык авторов (казахский) и целевой язык (русский) демонстрируют разные способы концептуализации fictive motion, следует тем не менее подчеркнуть, что в самом русском языке эта зона является весьма подвижной: в ней метафоры возникают, устаревают и конкурируют между собой. Результаты изложены в работе [75].

## 3.4 Пограничные явления в русских условных конструкциях

В русле решения задач, связанных с описанием специфики современного русского языка на материале корпусных данных, выявлен принцип выделения особой группы условных конструкций. Результаты изложены в работе «Русские контрфактивные предложения как наиболее информативный тип условных».

В этом исследовании предлагается новый параметр, отличающий контрфактивные условные предложения от других типов гипотетических условных конструкций: способность разрешать неоднозначность, иначе — информативность. В рамках функционального подхода обосновывается гипотеза, согласно которой контрфактивные конструкции являются наиболее информативным типом среди условных конструкций: они более информативны, чем реальные и маловероятные конструкции. Эта гипотеза, как демонстрируется, позволяет объяснить ряд фактов, связанных с контрфактивными конструкциями как в русском языке, так и в типологической перспективе.

Аргументация, основанная на данных русского языка, строится на предположении о том, что различные манифестации контрфактивности, регистрируемые в русских условных конструкциях, получают объяснение, если принять гипотезу информативности. В статье на материале Национального корпуса русского языка (НКРЯ) рассматриваются следующие манифестации контрфактивности.

Бессоюзные условные конструкции с глаголом в форме сослагательного наклонения (ср. *Знала бы с утра, надела бы что-нибудь другое* [НКРЯ]) в русском языке с необходимостью являются контрфактивными.

Условные конструкции с союзом *когда бы* (ср. *Ты бы тоже замолчал, когда бы услышал его пророчество* [НКРЯ]) в современном русском языке с необходимостью являются контрфактивными.

Инфинитивный протазис в русском языке с необходимостью содержит частицу *бы*, если реферирует к ситуации в прошлом (ср. невозможность опустить *бы* в протазисе: *Если бы мне иметь тогда одну сотую нынешнего опыта, дело можно было бы закрыть* [НКРЯ]) и может не содержать *бы* при референции к настоящему или будущему (ср. факультативность бы в протазисе: *Если бы заснять на кинокамеру весь процесс строительства, <…> зрелище получилось бы завораживающим* [НКРЯ]).

Типологически ориентированная часть аргументации базируется на обобщении, согласно которому множество языков, использующих реальные бессоюзные условные конструкции, составляет подмножество языков, использующих контрфактивные бессоюзные условные конструкции. Предлагаемое объяснение этой закономерности опирается на гипотезу информативности и состоит в следующем. Бессоюзные условные предложения обладают большей семантической неопределенностью, чем условные предложения, вводимые союзом. В то же время, благодаря своей информативности, контрфактивные конструкции лучше эксплицируют условную интерпретацию, чем реальные конструкции.

Таким образом, обосновываемая в работе гипотеза информативности позволяет пролить свет на, казалось бы, не связанные между собой явления.

## 3.5 Русские пограничные подчинительные конструкции

В рамках задачи изучения пограничных языковых явлений, нестандартных речевых употреблений, вариативности в грамматике и дискурсе продолжено изучение употребления предложений с подчинительными союзами как самостоятельных, вне сложноподчиненного предложения.

В предшествующих исследованиях было показано, что самостоятельное употребление предложений с подчинительными (*потому что, так как, хотя*, и др.) и другими союзами (то есть), канонически вводящими часть сложного предложения, достаточно частотно в русском языке. Так, в выборке из Основного корпуса Национального корпуса русского языка за 2000–2021 гг. почти в 40% примерах с союзом *то есть* союз употребляется в начале самостоятельного предложения [76]. Частотность примеров с начальным союзом *потому что* в Основном корпусе НКРЯ за 1991–2024 г. составляет 4617 текстов, 13 459 примеров, в Газетном подкорпусе — 39 049 текстов, 50 187 примеров. Около 10% из них — это вопросо-ответные конструкции, однако остальные примеры — это случаи независимого употребления предложений с союзом *потому что*. Исследователями были выделены условия, при которых предложение с союзом *потому что* не может быть придаточным и оформляется как самостоятельное [77], [78]. Что касается союза «хотя», то, по мнению [79: 325], отделение точкой придаточных с союзом *хотя* от главного предложения может давать текст, а не парцелляцию как таковую, поскольку этот союз, как и некоторые другие, близок к частице.

В этом году на материале Основного и Газетного подкорпуса НКРЯ исследовался процесс формирования независимого употребления предложений с союзами «потому что» и «хотя». Рассматривались два наиболее частотных типа контекстов, в которых предложение с союзом «потому что» употребляется как независимое.

Наиболее частотный контекст предполагает, что предшествующее предложение имеет выделенную рему — контрастную, эмфатическую (выражается оценка, отрицание, противопоставление, сравнение и т.п., имеет место верификативное высказывание).

1. *В то время, когда Александр Александрович у нас работал, они с Мейерхольдом просто враждовали… Потому что Мейерхольд требовал искать новое искусство, а Блок был против всяких исканий и так нам прямо и говорил*… [Владимир Рецептер. Ностальгия по Японии (2000)]

В ходе микродиахронического исследования на материале Основного и Газетного подкорпусов НКРЯ с XVIII по начало XXI в. (2024 г.) удалось установить источник и ход формирования первого типа контекстов. Это разговорные вопросо-ответные диалогические структуры, включающие в себя вопрос о причинах одного из участников и ответную реплику, начинающуюся с «потому что» (*Почему? — Потому что* …). С течением времени такой вопросо-ответный диалог, предполагающий иллокутивное вынуждение ответа о причинах [80: 86-87], видимо, трансформировался в монологическую речь (вероятно, через конструкцию риторического вопроса типа объективизации [81: 265]), затем, видимо, вопрос перешел в имплицитную форму, а ответ, начинающийся с «потому что», стал самостоятельным высказыванием, которое иллокутивно вынуждается предшествующим предложением. Иначе говоря, оценка, отрицание, противопоставление, контраст, верификативное высказывание в предшествующем контексте иллокутивно вынуждают к объяснению причин, которое дается в форме предложения «потому что». Отметим, что связь рассматриваемых употреблений с разговорным диалогом отражает общую зависимость самостоятельного употребления предложений с подчинительными союзами в письменных текстах от устной разговорной речи: такие употребления свойственны опубликованным интервью, обнаруживаются в цитатах или в текстах, использующих стратегии устной разговорной речи в полемических целях (колонки, комментарии, блоги) [78], [82], [77].

Таким образом, было обнаружено, что источником самостоятельного употребления предложений с союзом «потому что» в большинстве контекстов, видимо, является не парцелляция, а другие речевые структуры.

Другой частотный контекст независимого употребления предложений с союзом «потому что» предполагает, что предложение вводит дополнительную, добавочную информацию с точки зрения развития темы; добавочная тема не получает дальнейшего развития в тексте:

1. *От ХПН. Хронической почечной недостаточности. Вместо того чтобы госпитализироваться на очередной сеанс гемодиализа, он пошёл на работу. Потому что сложный случай, и никому больше в руки несчастный, измученный лейкозом ребёнок не давался. Вернулся с температурой сорок, лёг на кровать и умер*. [Татьяна Соломатина. Мой одесский язык (2011)].

На основе анализа коммуникативной организации фрагмента (актуального членения высказываний, фокуса внимания и его смены) было показано, что в подобных случаях независимое употребление предложения с подчинительным союзом обусловлено необходимостью выразить уточнение, пояснение, кратко рассмотреть побочную тему. При этом предложение с союзом «потому что» не может быть присоединено к предыдущему без нарушения логики развития темы, иначе внимание переключается на рему сложного предложения — предикацию с союзом «потому что» и дальнейшее возвращение к общей теме фрагмента оказывается затруднительным:

1. …*Вместо того чтобы госпитализироваться на очередной сеанс гемодиализа, он пошёл на работу, потому что сложный случай, и никому больше в руки несчастный, измученный лейкозом ребёнок не давался. Вернулся с температурой сорок, лёг на кровать и умер*.

Очевидно, в таких контекстах самостоятельное употребление предложения с «потому что» является определенной стратегией и обусловлено дополнительным статусом высказывания относительно главной темы фрагмента. Иначе говоря, тут проявляется «стратегия вставки»: «Вставка — это временное отклонение от маршрута изложения, выполнение некоторого побочного хода, и затем возврат к исходному маршруту» [83: 140]. Эта стратегия была выделена в устном дискурсе, однако она нередко применяется и в письменной речи, особенно имитирующей устную или основанной на исходном устном сообщении.

Такой тип контекста особенно характерен для предложений с союзом «хотя». Они составляют 81% из всех предложений с союзом «хотя» в выборке из НКРЯ, как из Газетного, так и Основного корпуса.

1. *И они молча, тупо шли вперед и вперед, строго по компасу. Он — метров на десять позади, как и подобает сильнейшему. Хотя на самом деле сильнейшим он не был: часа полтора назад напомнил о себе старый перелом левой голени и тайком от Марии Антуан приволакивал ногу. Она вспомнила оставленную в машине холщовую сумку с лимонами и апельсинами. Как сладко в лютую жару очистить пахучую маслянистую кожуру апельсина и вонзить зубы в сочную, освежающую мякоть*!

Как и в примерах с «потому что», присоединение к предшествующему предложению нарушает логику развития темы:

1. *И они молча, тупо шли вперед и вперед, строго по компасу. Он — метров на десять позади, как и подобает сильнейшему, хотя на самом деле сильнейшим он не был: часа полтора назад напомнил о себе старый перелом левой голени и тайком от Марии Антуан приволакивал ногу. Она вспомнила оставленную в машине холщовую сумку с лимонами и апельсинами*.

Контекст оговорки, комментария, содержащего возражение, не отменяющее, но корректирующее сказанное ранее, соответствует семантике союза «хотя», который «вводит валентность «препятствия» — ситуации X, которая должна была помешать, наступлению Y, но не смогла» [84].

Тема, вводимая предложением «хотя», может распространяться на несколько предложений или даже разрастаться до целого фрагмента, абзаца, так что вставка трансформируется во фрагмент с новой темой, союз маркирует переход к новой теме (о маркерах поворота темы [85]) . Разрастание вставки встречается и в контекстах с союзом «потому что».

Представляется, что этот тип употреблений возник именно в результате парцелляции сложноподчиненных предложений с придаточными причины и уступки, однако не экспрессивной и не «конструктивной» парцелляции, а парцелляции как приема организации дополнительного сообщения, комментирующего предшествующее (базовую часть).

Парцелляцию обычно не относят к числу приемов, выражающих значение добавочного сообщения. Однако возможность его выражения у парцеллированных конструкций есть. Важнейшей функцией парцелляции исследователи считают создание нового рематического центра, отдельного высказывания [86]; [87: 279]; [88: 86]; [89: 4]; [90: 3]; [91: 7], то есть повышение коммуникативного статуса отделенного сообщения, которое, при сохранении семантической связи с базовой частью, может выражать комментарий к предшествующему (базовому) высказыванию. Эффект «комментария», «дополнения», введение побочной темы для пояснения сказанного ранее виден и в случае парцелляции в простом предложении:

1. *Процесс пошел. Вспять?* [81: 269].

При этом новая тема, вводимая парцеллятом (с точки зрения актуального членения — рема высказывания), не предполагает продолжения в следующих предложениях, в отличие от новой, актуальной информации, содержащейся в самостоятельном предложении, в т.ч. сложноподчиненном. Тем самым парцеллят действительно оказывается удобной формой для введения короткого комментария к предшествующему предложению, оговорки, пояснения — в конструктивно-семантическом отношении, вставки.

Таким образом, есть основания полагать, что источником возникновения такого типа контекста, в котором предложения с союзами «потому что» и «хотя» вводят дополнительное с точки зрения развития темы в тексте сообщение, вставку-комментарий, уточнение, оговорку, стала парцелляция, которая, однако, выступила не как экспрессивный прием, а как прием формирования дополнительного сообщения, связанного с предшествующим.

Итогом исследования стало выявление двух разных источников формирования независимого употребления предложений с союзами «потому что» и «хотя»: для первого из них — вопросо-ответные конструкции и парцелляция, создающее добавочное сообщение, вставку, для второго — парцелляция с функцией создания добавочного сообщения.

Результаты исследования изложены в докладе «Развитие независимого употребления предложений с союзами ХОТЯ, ПОТОМУ ЧТО: функция дополнительного сообщения» на IX Международном симпозиуме «Русская грамматика: полипарадигмальность как методологический принцип современных научных исследований» (г. Иркутск, 23–27 сентября 2025 г.) и в подготовленной статье «Развитие независимого употребления предложений с союзом «потому что» в VIII–XXI веках: от речевых структур к имплицитным смыслам» для журнала «Труды института русского языка им. В.В. Виноградова».

## 3.6 Пограничные конструкции: реакции на экстралингвистические ситуации

В работе «Прагматика в цифровую эпоху: база данных «Рутиникон» рассмотрен цифровой инструмент описания *рутин* — особого класса формульных фразеологических единиц, которые представляют собой реакции на стандартные экстралингвистические ситуации или комментарии к ним. Например,формула *Кого я вижу!* –– это реакция на ситуацию неожиданной встречи, а *Кто там?* — стандартная формула-реакция на стук в дверь. Сбор, классификация и исследование единиц такого рода представляет несомненный интерес с как точки зрения теории дискурса и прагматики, так и с точки зрения практики освоения живого разговорного русского языка инофонами. «Рутиникон» является естественным продолжением проектов «Русский Конструктикон» и «Прагматикон» и заимствует у проектов-предшественников некоторые принципы сбора и работы с данными. При этом в «Рутиниконе» собраны фразеологические единицы другого типа, чем в этих базах, и для их описания требуется принципиально иная структура разметки. Разработаны такой разметки, а также ее обсуждается потенциал для внутриязыковой классификации рутин русского языка и применимость к соответствующему материалу других языков — в перспективе последующих типологических исследований. Для иллюстрации привлекаются примеры из материала итальянского языка. Авторы исследования обратили внимание на необычность класса рутин как такового и его отличия от формул вежливости — близких к рутинам с точки зрения структуры ситуации и прагматики и описали процесс сбора языкового материала, классификация и описание которого строится параллельно с созданием специализированной базы данных «Рутиникон», продолжающей проекты «Русский конструктикон» и «Прагматикон». Особенность разметки состоит в том, что она сразу ориентирована на сопоставление материала разных языков, несмотря на то, что материал еще недостаточно представителен, а классификация — недостаточно разработана, чтобы предложить хотя бы фрагмент типологической анкеты.

## 3.7 Цифровые инструменты для исследования пограничных явлений

В рамках решения задачи о создании цифровых инструментов для изучения русского языка продолжалось совершенствование и пополнение коллекций созданных ранее ресурсов: Корпуса русских учебных текстов (КРУТ: <http://web-corpora.net/learner_corpus/>), Русского учебного корпуса (RLC: <http://www.web-corpora.net/RLC/>), Корпуса раннего детского письма (StartWrit: <http://www.web-corpora.net/children/>); Корпуса ошибок носителей русского языка, изучающих английский язык (REALEC: <http://realec.org/>).

На материале дополненной коллекции Корпуса раннего детского письма в сопоставлении с письменной речью наивных авторов ХХ века и текстами русского языка, датированными ХVIII–ХIХ вв., показаны черты сходства в орфографических решениях, принимаемых пишущим, который не знаком с современным речевым стандартом и кодифицированной нормой. Рассмотрены явления в области отражения на письме безударных гласных в корне и аффиксах, согласных в составе кластеров, на стыке морфем, на конце слов, передающих звуки, вариативно кодируемые на письме. Классификации таких нестандартных написаний посвящены доклады на Апрельской конференции НИУ ВШЭ, мероприятии ФГН (посвященном функционированию русского языка) и на конференции по детской речи в Санкт-Петербурге.

Другим направлением работы был образовательный проект — разработка в рамках студенческой мастерской электронного справочника по проблемам русской стилистики. Проект предполагает создание разделов по трудным вопросам грамматики, лексикологии, синтаксиса, дискурса. Особенное внимание уделяется вопросам вариативности и нестандартных речевых употреблений. В 2025 году была создана структура сайта, продумана структура каждого раздела и подготовлены главы справочника, посвященные: 1) случаям конкуренции дательного падежа и предлога для при глаголах и предикативах в русском языке, например: *мне важно/для меня важно.* В разделе предложены рекомендации для выбора той или другой формы в зависимости от условий контекста, 2) стандартному и нестандартному употреблению неопределенных местоимений серии на -то (*какой-то, кто-то*).

Разделы справочника разрабатываются с опорой на научные исследования последних лет. Так, раздел (1) подготовлен на основе работ [92], [93], [94] и др.

Раздел (2) подготовлен на основе исследований: [95], [96], [97], [98].

Образовательный ресурс «Справочник по практической стилистике русского языка» призван познакомить преподавателей русского языка и учащихся с результатами современных лингвистических исследований и осветить трудные вопросы грамматики, лексикологии, синтаксиса, текста.

## 3.8 Исследования ошибок в русской речи

### 3.8.1 Оптимизация инструкции разметки Русского учебного корпуса

Целью работы стало повышение точности и согласованности разметки и, как следствие, упрощение перевода размеченных данных корпуса в формат, пригодный для решения задач в области обработки естественного языка, таких как, например, машинное обучение для автоматического исправления ошибок (GEC). Необходимость пересмотра была вызвана выявленными проблемами в существующей системе разметки, которые приводили к низкой согласованности аннотаторов и затрудняли автоматизированный анализ данных.

#### 3.8.1.1 Недостатки предыдущей системы аннотирования Русского учебного корпуса и выявленные проблемы

Классификация тегов, по которой ранее проводилась разметка Русского учебного корпуса, была подробной и ориентированной на причины возникновения различных типов ошибок. Однако детальный анализ выявил ряд серьезных недостатков:

* низкая согласованность между разметчиками: сложность и размер набора тегов (40 основных + 7 дополнительных) приводили к тому, что разные аннотаторы по-разному маркировали одни и те же явления;
* размытые границы между тегами: отсутствие четких критериев различения таких тегов, как Ortho (орфографические ошибки) и Infl (ошибки в окончании), Refl (ошибки в возвратных глаголах) и Passive (ошибки в пассивных конструкциях), Gov (ошибки управления) и Prep (ошибки в употреблении предлога);
* неоднородность тегов: в сферу действия одного тега могли входить ошибки с разной лингвистической природой. Например, тег Impers (ошибка в безличных конструкциях) использовался для выделения и ошибок согласования, и ошибок в выборе залога и лексических средств, что значительно снижало его информативность.

#### 3.8.1.2 Редизайн системы аннотирования

В ответ на выявленные проблемы система аннотирования была полностью переработана. Были выявлены кластеры тегов, которые чаще всего непоследовательно использовались аннотаторами. Изменения и уточнения системы аннотирования проводились на основе анализа кластеров тегов. Основные изменения включают:

* **уточнение определений:** Для каждого тега были разработаны максимально четкие критерии применения с примерами и пояснениями для сложных случаев;
* **устранение избыточности:** Ряд тегов был удален (Disc, Subst, Del, Insert, GenNeg, ArgStr, Par), если их сфера действия совпадала или значительно пересекалась с другими тегами;
* **введение новых тегов:** Для повышения точности были введены новые теги, такие как Voice (залог) и Nominative (падеж подлежащего).

#### 3.8.1.3 Ключевые изменения в кластере орфографических тегов

* Ortho**:** ошибки, объяснимые с фонетической точки зрения (ошибочная передача звука на письме (напр, безударные гласные в корне слова). Имеет высший приоритет в спорных случаях;
* Misspell**:** комплексные, неанализируемые ошибки, затрагивающие две и более буквы одновременно (напр., *менеию*→*мнению*);
* Graph**:** ошибки, вызванные визуальным сходством графем (внутри или между алфавитами, напр., *р* в кириллице (*река*) / *p* в латинице (*pen*);
* Typo**:** механические ошибки, не имеющие системный характер и не объяснимые лингвистически (напр., пропуск буквы *Оссия*→*Россия*).

Также были сформулированы критерии для различения ошибок в окончаниях, связанных с орфографией и морфологией:

* Ortho (орфографическая ошибка)**:** качество редуцированного гласного в ошибке и исправлении совпадает (напр., *согласин* → *согласен*). Сюда же относится смешение *и/ы* и ошибочное написание мягкого знака в глаголах третьего лица единственного числа;
* Infl (ошибка в окончании, связанная с морфологическими причинами)**:** качество гласного при исправлении меняется, и полученная форма является несуществующей в парадигме слова (напр., *говорем* → *говорим*).

#### 3.8.1.4 Морфологический кластер

Тегами Refl и Passive могли помечаться как случаи неверного образования возвратной / пассивной формы, так и случаи их неверного употребления. В обновленной системе эти теги маркируют только неверное образование формы, в то время как нарушения употребления, имеющие, скорее, синтаксический характер, вынесены в сферу действия специально введенного тега Voice. Сужена сфера действия тега Impers: ошибки согласования в безличных предложениях отнесены к тегам AgrNum, AgrGender, а ошибки, связанные с использованием инфинитива — к тегу Infinitive.

#### 3.8.1.5 Синтаксический кластер

* Введен тег Voice: этот тег используется для маркировки ошибок в выборе залога на уровне синтаксической структуры предложения (напр., использование активной конструкции вместо пассивной);
* Введен тег Nominative: выделен из общего тега Syntax для выделения специфической ошибки маркирования подлежащего косвенным падежом, что часто происходит вследствие неправильного анализа синтаксических связей говорящим;
* Введено разграничение сферы действия тегов Gov (ошибка в управлении) и Prep (ошибка в предлоге):
  + Gov**:** нарушение падежного управления при глаголе или другом главном слове. Может включать как неправильный падеж при правильном предлоге, так и пропуск обязательного предложного управления;
  + Prep**:** ошибка в выборе предлога, а также употребление лишнего предлога или его пропуск. Может затрагивать как актанты, так и сирконстанты.

#### 3.8.1.6 Лексический кластер

* введено разграничение тегов Lex и Asp:
  + Asp: используется, когда ошибка заключается исключительно в выборе аспекта (совершенный/несовершенный) (напр., *открыться* → *открываться*);
  + Lex**:** используется для всех прочих случаев неправильного выбора слова, включая ошибки в приставках с фазовым значением (напр., *прочитал* → *дочитал*);
  + Lex **+** Asp**:** комбинация тегов используется в случаях, когда допущена ошибка и в выборе лексемы, и в выборе аспекта.

#### 3.8.1.7 Предварительная оценка обновленной системы аннотирования

Для оценки эффективности новой системы аннотирования были проведены два пилотных эксперимента по расчету согласованности разметчиков при аннотировании ошибок, связанных с орфографией.

* **Эксперимент 1** (для тегов Ortho, Infl, Misspell): согласие аннотаторов (Каппа Флейсса-Коэна) выросло с ~0.38 (предыдущая система аннотирования) до ~0.51 (обновленная система аннотирования).
* **Эксперимент 2** (для тегов Ortho, Typo, Misspell): согласие аннотаторов (Каппа Флейсса-Коэна) выросло с 0.16 (предыдущая система аннотирования) до 0.54 (обновленная система аннотирования).

### 3.8.2 Исследование замечаемости лексических ошибок непрофессиональными аннотаторами

В рамках исследования была предпринята попытка понять, насколько хорошо и последовательно непрофессиональные аннотаторы замечают и исправляют лексические ошибки и что влияет на их замечаемость и исправляемость. Лексические ошибки, в отличие от грамматических и орфографических, характеризуются высокой вариативностью и сложностью, что затрудняет их выявление и коррекцию. Наряду с орфографическими, эти ошибки являются одними из наиболее частотных [99], [100], однако орфографические нарушения, как правило, исправляются более единообразно. Особое внимание к непрофессиональным аннотаторам обусловлено перспективностью использования краудсорсинговых технологий для исправления ошибок при работе с большими объемами данных.

Исследование носило экспериментальный характер. В качестве материала были использованы 32 предложения с лексическими ошибками из Русского учебного корпуса (RLC). Для всех предложений экспертами-лингвистами были сформулированы исправления, представляющие собой «золотой стандарт». Некоторые из этих предложений были дополнительно скорректированы с целью минимизации количества возможных исправлений. В каждом из 32 предложений содержалось одно целевое слово, в котором была допущена лексическая ошибка, и для которого необходимо было выяснить, будет ли ошибка замечена, а само слово исправлено в зависимости от различных условий. Для каждого такого предложения было создано 4 вариации, каждая из которых содержит одну и ту же целевую лексическую ошибку и разные дополнительные ошибки:

* предложение без дополнительных ошибок;
* предложение с дополнительной ошибкой в целевом слове (где находится основная лексическая ошибка);
* предложение с дополнительной ошибкой вне целевого слова (в правом или левом контексте);
* предложение с двумя дополнительными ошибками (одна внутри и одна вне целевого слова).

Полученные 128 предложений были распределены по 8 онлайн-формам таким образом, чтобы вариации одного и того же предложения не встречались в одной форме (в каждой форме содержалось 16 предложений для исправления). Формы заполняли непрофессиональные аннотаторы. Перед респондентами стояла задача исправить ошибки в предложениях, не изменяя смысла написанного. Разрешалось вносить только необходимые исправления, даже если итоговое предложение звучало не совсем литературно, но приемлемо. В результате было собрано 262 ответа, что позволило получить от 20 до 35 исправлений на каждое предложение.

Каждое из 128 предложений было размечено по различным параметрам (наличие и тип дополнительных ошибок, коллокационная и семантическая близость между ошибочной лексемой и исправлением, POS разметка ключевых слов).

Анализ зависимости замечаемости лексических ошибок от их типа показал, что наиболее заметными являются ошибки в семантически не связанных парах и коллокациях. В обоих случаях это можно объяснить тем, что ошибки влияют на устойчивость итогового словосочетания, а низкая устойчивость словосочетания заметна аннотаторам. Под устойчивостью в данном контексте понимается вероятность, с которой определенный элемент предсказывает совместное появление остальных элементов сочетания. Анализ зависимости замечаемости лексических ошибок от части речи показал незначительную корреляцию: лексические ошибки в глаголах замечаются лучше, чем в прилагательных и существительных (см. [рисунок 3.10](#fig-zamechayemost)).

|  |
| --- |
| Рисунок 3.10 – Зависимость замечаемости лексических ошибок от части речи |

В целом результаты эксперимента показали, что дополнительные ошибки не оказывают столь значительного влияния на замечаемость и исправляемость лексических ошибок. Более выраженное влияние на эти показатели оказывают различные семантические характеристики и корпусные параметры. Во-первых, лексическая ошибка оказывается наиболее заметной в контексте, где отсутствуют дополнительные ошибки как внутри, так и вне целевого слова. Наличие ошибок как в целевом слове, так и вне его приводит к снижению заметности лексической ошибки. Расположение дополнительной ошибки не оказывает значительного влияния на заметность целевой ошибки. Кроме отсутствия дополнительных ошибок, лишь наличие дополнительной лексической ошибки вне целевого слова немного повышает заметность целевой ошибки. Во-вторых, больше всего из типов лексических ошибок выделяются ошибки, определяемые нами как нарушение паронимии (формальные ошибки), которые независимо от уровня замечаемости всегда хорошо исправляются. Это свидетельствует о важности их выделения в классификациях лексических ошибок. Несмотря на отсутствие прямой зависимости между метрикой семантической близости (semantics) и показателями замечаемости и исправляемости, качественное выделение семантически связанных и несвязанных пар показывает различия в ошибках в этих типах: ошибки в семантически несвязанных парах оказываются значительно более заметными. Ошибки в коллокациях также являются хорошо заметными и легко исправляемыми.

### 3.8.3 Исследование аффективных оценок русских существительных носителями языка

Исследование посвящено аффективным реакциям носителей русского языка на существительные с использованием метода слайдера [101]. Метод слайдера позволяет с высокой точностью регистрировать степень эмоциональной комфортности участников при восприятии слов. Целью работы было проверить универсальность некоторых закономерностей, выявленных ранее при проведении подобного исследования на английском материале, а также установить влияние индивидуальных психологических характеристик — эмпатии, застенчивости и общительности — на аффективные оценки.

В исследовании приняли участие 150 носителей русского языка (62 мужчины и 88 женщин) в возрасте от 18 до 65 лет. Для оценивания были взяты 280 русских существительных, подобранных по аналогии с английскими стимулами из словаря ANEW [102]. Участникам предлагалось перемещать на экране компьютера фигурку, выбирая положение на шкале, соответствующее степени приятности или неприятности слова. После выполнения основного задания участники проходили три психологических теста:

* Sociability Scale [103] — измерение уровня общительности;
* Basic Empathy Scale [104] — измерение эмпатии;
* BIS Scale [105] — измерение застенчивости.

Результаты эксперимента позволили установить следующие закономерности:

* женщины показали высокие показатели эмпатии и застенчивости, мужчины — высокие показатели общительности;
* обнаружена положительная корреляция между эмпатией и застенчивостью (r = 0,37);
* аффективные оценки демонстрируют так называемый позитивный сдвиг (positivity bias) — 60,5% реакций направлены в сторону приближения;
* дистанция коррелирует с частотностью слов (r = 0,29): чем чаще слово, тем оно воспринимается как более позитивное;
* застенчивость оказывает отрицательное влияние на экстремальность оценок (p = 0,002);
* общительность влияет на приближение фигурки к стимулам (тенденция p < 0,1);
* эмпатия не показала значимого эффекта;
* при имитации реакций противоположного пола застенчивость также снижает точность предсказаний.

Полученные результаты подтверждают воспроизводимость англоязычных закономерностей [101] на материале русского языка. Застенчивость проявляется как значимый психологический фактор, влияющий на степень эмоционального дистанцирования при восприятии слов. Положительная корреляция между частотностью и валентностью указывает на универсальные закономерности обработки лексики. Результаты эксперимента имеют большое значение для создания тональных словарей, а также для обработки лексики в учебных корпусах и корпусах ошибок, в том числе при анализе лексического выбора, обусловленного языковой интерференцией.

## 3.9 Исследование дискурсивных явлений в адыгских языках

В рамках работы проекта регулярно собираются и пополняются корпуса абхазо-адыгских языков –– как письменные, так и устные. В частности, в 2025 году готовились корпуса преимущественно письменных текстов на абхазском языке (совместно с Академией наук Абхазии) и кабардино-черкесском языке (совместно с кабардинским языковым клубом «Джарез»), а также собирались тексты для устных корпусов адыгейского языка (аулы Пшизов, Хачемзий, Егерухай, Хакуринохабль Республики Адыгея; совместно с Адыгейским государственным университетом) и абазинского языка (аулы Старо-Кувинск Карачаево-Черкесской республики).

На базе этих корпусов, а также ранее созданных корпусов оказывается возможным проводить сложные исследования, затрагивающие функционирование разнообразных грамматических механизмов в дискурсе и изучать явления, которые невозможно прояснять путем простого опроса консультантов –– носителей языка. Существенно, что при этом мы получаем возможность изучить и насколько сильно противопоставлены говоры в отношении рассматриваемых явлений русскому языку и другим идиомам, оказывающим на них влияние.

Примером такого исследования является изучение стратегий референциального выбора в адыгских языках (адыгейском и кабардино-черкесском). В рамках этого исследования было изучено, как долго и какими средствами топики/участники сохраняются в дискурсе, то есть продолжительно упоминаются и остаются в речи. Для этого в устных корпусах, включающих адыгейские тексты из аула Джамбечий и тексты на кубанском диалекте кабардино-черкесского языка из аулов Блечепсин и Ходзь были выделены элементарные дискурсивные единицы (ЭДЕ). С помощью ЭДЕ были измерены референциальная линейная дистанция между данным упоминанием референта и предыдущим, а также учтены такие параметры, как одушевленность, интерференция других участников, темпоральная или локальная смена действия, неполная кореферентность участников и протагонистичность. Впрочем, в том, что касалось темпоральной или локальной смены действия, можно было ориентироваться только на явную экспликацию такого изменения, и в результате в выборку попало мало таких случаев (около 18% в кубанском корпусе и 11% в джамбечийском), поэтому влияние этого фактора не включалось в анализ.

Примеры из обоих корпусов составили базу данных состоящую из всех анафорических цепочек с контекстом и характеристиками, перечисленными выше, для каждого выражения участника дискурса. Эти базы данных были проанализированы в трех вариантах. В первом варианте присутствуют разные виды референциального выбора: ноль, независимое местоимение, ИГ, редуцированная ИГ, другая ИГ, местоимение + ИГ. Поскольку такое дробное разделение именных групп обычно не используется в литературе, во втором варианте все добавленные категории с ИГ были объединены. Для третьей версии из выборки были удалены все первые упоминания референтов, поскольку такие упоминания не дают новой информации об активации референта в памяти и, кроме того, не обладают всеми приобретенными характеристиками, которые зависят от ближайшего контекста.

### 3.9.1 Референциальное линейное расстояние

Следуя А.Е. Кибрику ([106]), ниже используются следующие метрики активации для линейного расстояния: при расстоянии 1 указывается показатель ноль, а далее показатель уменьшается на 0.1, пока не доходит до показателя -0.5. В ([таблица 6.9](#tbl-1)) приведены средние показатели для каждого из референциального выбора в каждом варианте выборки (они зашифрованы следующим образом: Б1 –– выборка из Блечепсина, первый вариант; Д1 –– выборка из аула Джамбечий, первый вариант).

Средние значения линейного расстояния для текстов из Блечепсина больше, чем для текстов из аула Джамбечий, в том числе из-за большего размера самих кубанских текстов (чем больше ЭДЕ в тексте, тем больше возможностей для того, чтобы надолго забыть топик, а потом его вспомнить). Самые маленькие значения принимают независимые местоимения. Впрочем, для полисинтетических языков, где участники могут выражаться прономинальными аффиксами, использование независимых местоимений (особенно, в случае с адыгскими языками, где такие местоимения не несут никакую дополнительную к префиксам информацию) ограничено и маргинально.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 3.3 – Средние значения линейного расстояния.   |  | Б1 | Б2 | Б3 | Д1 | Д2 | Д3 | | --- | --- | --- | --- | --- | --- | --- | | ноль | 1.46 | 1.46 | 1.47 | 1.09 | 1.09 | 1.13 | | местоимение | 1.21 | 1.22 | 1.27 | 1.14 | 1.14 | 1.14 | | именная группа | 1.93 | 2.45 | 3.88 | 1.7 | 1.68 | 2.27 | | редуцированная именная группа | 3.27 | — | — | 1.17 | — | — | | другая именная группа | 3.8 | — | — | 1.86 | — | — | | местоимение + именная группа | 4.33 | — | — | 1 | — | — | |

Самая заметная разница между средними значениями возникает в третьей версии каждой из таблиц. Такая разница служит доказательством бóльшей репрезентативности именно третьего варианта данных для этой характеристики, поскольку исключение первых упоминаний (которые были помечены 0 в данных) устраняет искажения, связанные с отсутствием антецедента.

На ([рисунок 3.11](#fig-2)) и ([рисунок 3.12](#fig-3)) можно увидеть пороговые значения линейного расстояния, при котором вероятность использования именной группы превышает 50% — после превышения этого порога ЭДЕ произнесенных после ближайшего упоминания участника, говорящий скорее всего выберет именную группы как выражение референта. Как мы видим, и для блечепсинских, и для джамбечийских текстов этот порог не такой высокий: 2.25 ЭДЕ и 0.97 ЭДЕ соответственно.

|  |
| --- |
| Рисунок 3.11 – Пороговое значение линейного расстояния для именных групп в текстах из аула Блечепсин. |

|  |
| --- |
| Рисунок 3.12 – Пороговое значение линейного расстояния для именных групп в текстах из аула Джамбечий. |

Как видно, несмотря на признанное высокое влияние линейного расстояния на референциальный выбор, в условиях короткого устного адыгского дискурса только линейного расстояния недостаточно для того, чтобы предсказать этот выбор, из-за недостаточной количественной разницы между линейным расстоянием у именной группы и нулевой формы.

### 3.9.2 Одушевлённость

Данные по влиянию одушевленности приведены в ([таблица 6.10](#tbl-2)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 3.4 – Значения хи-квадрата и коэффициент Крамера V для одушевленности и вида.   |  | **Б1** | **Б2** | **Б3** | **Д1** | **Д2** | **Д3** | | --- | --- | --- | --- | --- | --- | --- | | **χ2** | 0.0 | 0.0001 | 0.0002 | 0.0097 | 0.0012 | 0.0052 | | **V** | 0.307 | 0.253 | 0.273 | 0.266 | 0.250 | 0.244 | |

Так как значения теста хи-квадрата в каждом из вариантов < 0.05, связь между видом и одушевленностью везде значимая, но значения коэффициента Крамера V показывает, что эта связь слабая. Следовательно, изолированно одушевленность слабо влияет на референциальный выбор. Самый большой процент неодушевленных референтов находится среди именных групп.

### 3.9.3 Протагонистичность

Статус протагониста часто совмещен с одушевленностью участника. Эти характеристики помогают более медленной деактивации референта в рабочей памяти ([107]), так что такие референты чаще будут выражаться нулем. Соответствующие данные по текстам из аулов Блечепсин и Джамбечий даны в ([таблица 6.11](#tbl-3)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 3.5 – Значения хи-квадрата и коэффициент Крамера V для протагониста и вида выражения.   |  | **Б1** | **Б2** | **Б3** | **Д1** | **Д2** | **Д3** | | --- | --- | --- | --- | --- | --- | --- | | **χ2** | 0.0020 | 0.0003 | 0.0072 | 0.2564 | 0.6116 | 0.8409 | | **V** | 0.255 | 0.237 | 0.208 | 0.175 | 0.068 | 0.044 | |

Для текстов из Блечепсина связь значима, но слабая, то есть статус протагониста действует совместно с другими факторами. Для текстов из аула Джамбечий связь между статусом протагониста и вида выражения не является значимой.

**Интерференция других участников** связана с наличием других топиков в похожей роли в трех ЭДЕ до рассматриваемого упоминания референта. В ([таблица 6.12](#tbl-4)) показывается, что связь между интерференцией и референциальным выбором есть. Так как значение коэффициент Крамера V находится в промежутке 0.3-0.5, мы можем назвать эту связь умеренной, то есть интерференция сильнее всех предыдущих характеристик влияет на референциальный выбор. Однако эта связь не является сильной, следовательно, интерференция тоже работает скорее в соединении с другими характеристиками.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 3.6 – Значения хи-квадрата и коэффициент Крамера V для интерференции и вида выражения.   |  | **Б3** | **Д3** | | --- | --- | --- | | **χ2** | 0.0 | 0.0 | | **V** | 0.504 | 0.348 | |

На ([рисунок 3.13](#fig-4)) и ([рисунок 3.14](#fig-5)) мы можем увидеть, что действительно при возникновении такой интерференции говорящий скорее будет выбирать именную группу, а не нулевую форму. В корпусе Джамбечий в таком случае иногда могут использоваться и независимые местоимения.

|  |
| --- |
| Рисунок 3.13 – Распределение типов анафор в зависимости от присутствия интерференции других топиков в трех предыдущих ЭДЕ, Блечепсин. |

|  |
| --- |
| Рисунок 3.14 – Распределение типов анафор в зависимости от присутствия интерференции других топиков в трех предыдущих ЭДЕ, Джамбечий. |

**Неполная кореферентность** с антецедентом возникает, когда анафора восходит к какому-то антецеденту, но их референты не полностью совпадают. Считается, что при проблемах с референции, доступность участника в рабочей памяти будет снижаться, поэтому чаще будут выражаться именными группами ([108]). Итоговые данные по этому параметру приведены в ([таблица 6.13](#tbl-5)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 3.7 – Значения хи-квадрата и коэффициент Крамера V для неполной кореферентности и вида.   |  | **Б1** | **Б2** | **Б3** | **Д1** | **Д2** | **Д3** | | --- | --- | --- | --- | --- | --- | --- | | **χ2** | 0.1664 | 0.0698 | 0.0119 | 0.0526 | 0.0245 | 0.0202 | | **V** | 0.164 | 0.135 | 0.197 | 0.226 | 0.186 | 0.210 | |

В тех случаях, когда связь оказывается значимой (в Б3, Д2 и Д3), значение коэффициента Крамера V не превышает 0.3, следовательно, признак неполной кореферентности не является определяющим фактором при референциальном выборе.

В дальнейшем вес разных факторов был соотнесен с помощью деревьев решений. В кубанском корпусе решающим фактором референциального выбора является интерференция –– говорящий в первую очередь пытается избежать неоднозначности в выражении участников дискурса. Кроме того, в случаях, когда участник-протагонист не испытывает затруднений с кореферентностью, выбор между именной группой и нулевой анафорой зависит от одушевлённости протагониста, которая выступает решающим фактором. Положительные значения активации, связанные с одушевлённостью, приводят к снижению скорости деактивации участника в рабочей памяти; как следствие, одушевлённый участник чаще выражается нулевой формой.

Перспективным направлением могло бы стать сопоставление полученных данных с материалами других полисинтетических языков как внутри абхазо-адыгской семьи, так и в других языках (ср. исследования ([109]; [110]; [111]) и др.). Кроме того, аналогичные исследования могут быть приведены для письменной речи, на которую намного большее влияние оказывает русский язык.

## 3.10 Исследования морфосинтаксиса и семантики глагольных форм в отдельных языках

### 3.10.1 Результативные конструкции в тундровом ненецком языке

В рамках проекта в 2025 году проведены исследования ряда аспектов глагольного морфосинтаксиса тундрового ненецкого языка, в первую очередь результативных конструкций, слабо описанныхв литературе по самодийским языкам. В работах по энецким и ненецким языкам отмечались результативные функции причастий (т. н. антериорного и пассивного антериорного в энецких, т. н. перфективного в ненецких), что типологически ожидаемо. Однако в этих языках встречаются и другие результативные стратегии.

В тундровом ненецком наряду с Перфективным причастием на *‑mi°*/*‑wi°* в предикативной позиции для выражения результативной семантики используются также Статив, образуемый с помощью показателей *‑ə* (подвергающегося чередованию *ə → i*) и *‑yo* (подвергающегося чередованию *o → u*) или чередований финальных гласных основы, Имперфектив на *‑tə*/*‑nə* и Дуратив на *‑(m)pə*.

Дуратив является единственным показателем, который может приводить к образованию переходных глаголов с результативной семантикой (типа *nyəqm°-bə-* ‘держать’ < *nyəqmə-* ‘взять’), однако непродуктивен в этой функции (о типологической редкости продуктивных переходных результативов хорошо известно). Из непереходных глаголов Дуратив в результативной функции сочетается прежде всего с производными глаголами, оканчивающимися на Трансформатив на *‑m*, а также показатели *‑(sya)ləm*, *‑xəm* и относящимися к акциональному классу достижений/семельфактивов.

Дуратив обычно присоединяется к трансформативному деривату, который образован от существительного, обозначающего профессиональную или этническую принадлежность, являющегося термином родства и т. п. В этом случае Дуратив от Трансформатива образует предикат, семантика которого значительно отличается от семантики исходного имени; ср. *wəs°kuy°mpə-* ‘важничать (букв. быть как купец)’ < *wəs°kuyəm-* ‘1) стать купцом; 2) стать важным’ < *wəs°kuy°* ‘1) купец; 2) важный’. Исходное имя может использоваться референтно, а дуративный дериват от трансформативного обозначает, например, поведение или указывает на некоторые значимые свойства субъекта.

Дуративы от глаголов с суффиксами *‑ləm*, *‑xəm* также не совпадают полностью по значению с исходными (в этом случае глагольными) основами; ср. *yabyexəmpə-* ‘пьянствовать (беспробудно)’ < *yabyexəm-* ‘стать пьяницей’ < *yabye-* ‘быть пьяным’.

Имперфектив в результативной функции присоединяется преимущественно к непереходным глаголам и встречается, таким образом, в основном в контекстах субъектного результатива. В сочетании с переходными глаголами показатели Имперфектива обычно приводят к антипассивизации и переводу свершений в класс деятельностей.

Результативная функция причастия с показателем *‑mi°*/*‑wi°* (которое в литературе именуется Причастием прошедшего времени, Перфективным причастием и пр. ярлыками) в предикативной позиции более исследована, в частности, известно, что в предикативной позиции Перфективное причастие является строго P-ориентированным (тогда как про тундровые ненецкие причастия в целом обычно говорится, что они имеют нейтральную залоговую ориентацию) и ведёт себя как типичный именной предикат. Однако в этом отношении также удалось получить новые данные и уточнить утверждения в существующей литературе.

Во-первых, несмотря на то, что Перфективные причастия от агентивных непереходных предикатов, по-видимому, в норме не встречаются в предикативной позиции сами по себе, они возможны в ней, если используется аналитическая конструкция с глаголом ‘быть’, то есть в контекстах отрицания, будущего времени и неиндикативных форм. Во-вторых, в аналитической конструкции с глаголом ‘быть’ Перфективные причастия в предикативной позиции могут быть образованы от глаголов с показателем Дуратива и модифицированы обстоятельствами длительности. В этих двух аспектах поведение Перфективных причастий в предикативной позиции в составе аналитической конструкции сближается с поведением Перфективных причастий в атрибутивной позиции, где также допустимы неэргативные и непредельные предикаты.

Данные наблюдения позволяют предположить, что аналитическая конструкция с Перфективным причастием в предикативной позиции выбирается не только по причине невозможности выражения той или иной морфологии на самом причастии, она ещё и проявляет иные свойства и иную сочетаемость (близкие к свойствам и сочетаемости атрибутивных причастий), что, вероятно, может свидетельствовать о различном объёме синтаксической структуры в этих двух конструкциях, а также о том, что в одном случае перед нами, действительно, результатив, а в другом, скорее, перфект (в терминах [112]). Однако проверка этих гипотез требует полноценного полевого исследования, которое может быть осуществлено в будущем.

Далее, было проведено исследование т. н. «стативной» морфологии в тундровом ненецком, суммирующее наблюдения в литературе и позволяющее в некоторых аспектах уточнить их.

Статив, реализующийся как сегментный показатель в сочетании с консонантными основами и как мена финального гласного в сочетании с вокалическими основами (в обоих случаях дериват переходит в класс т. н. «альтер(н)ирующих» глаголов, которые образуют т. н. «общую финитную» основу с помощью чередования), несмотря на свою общую продуктивность, обнаруживает ряд ограничений, как формальных (например, он не образуется от односложных основ на гласный), так и, по-видимому, семантических, при этом он может реферировать не только к наблюдаемым состояниям, но и к ненаблюдаемым, а также может вводить результирующий процесс, а не состояние (в сочетании с глаголами, обозначающими вхождение в состояние), и даже само вхождение в состояние (в сочетании с частью переходных предикатов). В последнем случае происходит обязательная антикаузативизация (декаузативизация).

Дистрибуция «стативных» дериватов и Перфективных причастий от непереходных глаголов в результативной функции также нуждается в дополнительном изучении. Что касается различий между Перфективными причастиями и Стативами от переходных глаголов, то последние не встречаются с выраженным агенсом и, по-видимому, не допускают его (хотя существование грамматического запрета должно быть проверено с информантами), и не могут, таким образом, образовывать пассивную конструкцию, в отличие от причастий.

Важным представляется также определить деривационный потенциал Статива, поскольку от некоторых предельных переходных глаголов он в принципе не образуется, в частности, от глагола *syerta-* ‘сделать’. Не зафиксирован Статив от каузативных глаголов с показателем *‑rəbta* и от многих каузативных глаголов с другими показателями. Стативы от эстимативных (≈ ‘считать, что P’) глаголов, маркированных показателем *‑kəbta*, напротив, образуются абсолютно регулярно. В ряде случаев засвидетельствованы только стативные дериваты, тогда как сами эстимативные глаголы неупотребительны; ср. *pyisy°kəbtə-* ~ *pyisy°kəbtyo-* ‘быть смешным’ (ср. *pyisyəh-* ‘смеяться’); *syad°kəbtə-* / *syad°kəbtyo-* ‘чувствовать неудобство, стыд’ (ср. *syadoh-* ‘пристыдить, смутить’).

### 3.10.2 Показатели эпистемической модальности и симилятивной деривации в самодийских языках

Было проведено исследование, посвящённое ряду показателей в самодийских языках. Они употребляются в качестве наклонения со значением эпистемической модальности и как показатель симилятивной деривации. Исследование проводилось на материале корпусов проекта INEL, грамматик самодийских языков и полевых данных из лесного ненецкого языка, которые были собраны автором в результате полевой работы в ЯНАО в течение 2023–-2025 гг. Наиболее подробно были проанализированы ненецкие материалы, энецкие материалы на уровне обзора корпуса также включены в выборку, а нганасанские симилятив и аппроксиматив были затронуты в формате пилотного исследования и находятся на стадии дополнительного изучения. Целью исследования было определить границы и пересечения дистрибуций показателей симилятива и аппроксиматива: так как в каждом языке показатели омонимичны и способны употребляться как в различном, так и в аналогичном морфосинтаксическом окружении, было необходимо подробно рассмотреть переходные состояния между симилятивной и аппроксимативной функциями этих маркеров.

Исследование показало, что соотношение между симилятивом и аппроксимативом в самодийских языках демонстрирует разную степень пересечения семантической и формальной дифференцированности. В ненецких языках эти категории выражаются частично омонимично; в энецких языках степень конструкционализации более высокая — формы симилятива и аппроксиматива чётко разграничены и имеют устойчивые значения; в нганасанском же аппроксиматив вовсе представлен единственной возможной формой и также значительно отличается от симилятива. Эти данные указывают на различный на синхронном уровне паттерн совмещения в этом маркере значений аппроксиматива и симилятива, и, вероятно, позволяют говорить о различии в стадии развития аппроксимативной функции этого показателя от симилятивного. Также в исследовании был предложен возможный путь развития модального значения от значения сходства ситуаций и представлена схема контекстного изменения функции симилятива, приводящая к развитию значения аппроксиматива: *симилятив на причастиях в позиции предиката > [утрата симилятивом конкретной контрфактивной семантики или актуализация модальной] > развитие аппроксиматива в полноценную глагольную форму, то есть появление у неё возможности спрягаться по разным парадигмам + (расширение контекстов аппроксиматива до непричастных форм).*

Работа над сравнением симилятивов и аппроксимативов в самодийских языках продолжается. В частности, планируется более подробное описание этих конструкций в лесном ненецком языке и корпусное исследование нганасанских симилятива и аппроксиматива.

### 3.10.3 Детранзитивирующий показатель *-(i)s’k* в удмуртском языке

Исследовались деривационные процессы в уральских языках. Основное внимание в первой половине года было уделено исследованию детранзитивирующего показателя ‑*(i)s’k* в бесермянском диалекте удмуртского языка. Целью было описать семантику и морфосинтаксис глагольных форм с суффиксом ‑*(i)s’k*, который известен в типологической литературе своей широкой полисемией.

В исследовании рассматривались функции суффикса ‑*(i)s’k*, ранее не описанные в литературе. Анализ проводился на данных корпуса бесермянского удмуртского; для каждого примера размечался набор участников и способ их оформления, а также тип актантной деривации (рефлексив, реципрок, декаузатив, имперсонал). Были описаны формы на ‑*(i)s’k* от глаголов различных валентностных классов: непереходных, переходных и трёхместных.

В ходе работы было показано, что в бесермянском удмуртском у ‑*(i)s’k* наблюдаются почти все значения, упоминаемые для литературного удмуртского: рефлексив и реципрок, антикаузатив (декаузатив) и антипассив (объектный имперсонал), за исключением канонического пассива — в корпусе нет примеров канонического пассива с ‑*(i)š’k.* Помимо канонических для детранзитивирующего показателя рефлексивных и декаузативных употреблений широкое распространение получает субъектный имперсонал, который зачастую имеет генерическое или хабитуальное прочтение. Подобные формы в корпусе часто описывают не единичные, а типичные ситуации, и их интерпретация сближается с субъектным имперсоналом, известным по другим пермским языкам. Например, глагол *ul*‑*ə̑nə̑* ‘быть, жить’ в корпусе допускает детранзитивную форму *ul*‑*iš’k*‑*ə̑nə̑* (55):



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| jun | š'ekə̑t | jun | š'ekə̑t-en | ul-iš'k-i-z |
| очень | тяжело | очень | тяжёлый-INS | жить-DETR-PST-3(SG) |
| 'Очень тяжело жилось' | | | | |

Подобные имперсональные употребления в бесермянском сопоставимы с типологически параллельными структурами в коми.

### 3.10.4 Каузативные показатели в ительменском языке

Отдельное внимание в 2025 году было уделено каузативным аффиксам в южном диалекте ительменского языка. Исследование проводилось на основе текстов, собранных А. П. Володиным в 1966–1973 годах, существующего корпуса ительменского языка ИЛИ РАН и словаря ительменского языка. Материалы собраны в южном диалекте ительменского языка. Для целей исследования часть текстов А. П. Володина объемом более 430 предложений была снабжена поморфемной нотацией.

В существующих работах, описывающих каузативизацию в ительменском языке [113], остаются почти незатронутыми вопросы о закономерностях употребления аффиксов, их семантической сочетаемости с глагольными основами и окончательном наборе каузативных средств с учетом их дистрибутивных особенностей. В ходе исследования была изучена литература по теме; данные обобщены в обзоре каузативных аффиксов в южном диалекте ительменского языка.

Морфологический каузатив образуется с помощью префиксальных показателей *ɬin /әn-*, что представлено в примерах (56) и (57):



|  |  |  |
| --- | --- | --- |
| mel | әnʲče-ƚ | n-ƚin-nu-miŋ |
| хорошо | рыба-INS | 3PL.S/A-CAUS-есть-1SG.P |
| '‘...Хорошо накормили они меня рыбой’ («Эмемкут борется с отцом»).' | | |



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Kutx | χeni-s-č | časit | p'e-ʔn | fse |
| Кутх | сказать-PRS-3SG.S | сейчас | ребенок-ABS.PL | всё |

|  |
| --- |
| q-әn-ŋiksi-xeʔn |
| 2.IMP.S/A-CAUS-заснуть-3PL.P |
| '‘Кутх сказал: сейчас уложи всех детей’ («Кутх и мыши»).' |

При этом префиксы обнаруживают избирательность. Так, префиксы *ɬin-/ɬәn-* демонстрируют тенденцию к сочетанию с агентивными глаголами, обозначающими более контролируемые непереходные действия (‘пить’, ‘есть’, ‘бросать’, ‘идти’), чем другие глаголы (‘тонуть’, ‘замерзать’). Однако значительно более продуктивным, по-видимому, является образование каузативов с помощью префикса әn-.

При этом суффиксы *-ŋ, -ʔɬ, -w*, традиционно рассматриваемые как часть циркумфикса, могут иметь самостоятельные значения и употребляться вне каузативных конструкций. Так, например, суффикс *-ŋ* встретился в сочетании с предикатом ‘спать’ в контексте (58), где, по-видимому, имеет значение интенсивности. Более того, в примере (59) представлен каузатив от того же глагола, имеющий в составе аффикс *-ŋ* с указанной некаузативной семантикой. Отметим, что каузативная деривация от предиката ‘спать’ встречается и без суффикса *-ŋ*, причем такие случаи представляются более частотными.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Kutx | χeni-s-č | časit | p'e-ʔn | fse |
| q-әn-ŋiksi-xeʔn | Кутх-AUG | очень | крепкий-ADV | PF-заснуть-INT-PF |
| '‘Кутх заснул очень крепко’ («Кутх и мыши»).' | | | | |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| qaʔt | Kutx | Miti | poka | p'e-n | k-әn-ŋiksi-ŋi-qzu-ʔin |
| вот | Кутх | Мити | пока | ребенок-ABS.PL | PF-CAUS-спать-INT-IMPF-PF |
| '‘Вот Кутх и Мити, пока укачивали детей, (тоже заснули)’ («Кутх и мыши»).' | | | | | |

Более подробное описание значений суффиксов при каузативной деривации требует дальнейших исследований. Доклад, посвящённый настоящему исследованию, был принят в очную секцию Пятой конференции по уральским, алтайским и палеоазиатским языкам (24–26 ноября 2025, ИЛИ РАН, г. Санкт-Петербург), тезисы будут опубликованы.

### 3.10.5 Конструкции с глаголом со значением ‘найти’ в аваро-андо-цезских языках

В 2025 году проводилось исследование конструкций с глаголом со значением ‘найти’ в цезском и других аваро-андо-цезских языках. В результате работы были обозначены четыре класса употреблений глагола со значением ‘найти’, помимо исходного, во всех рассмотренных языках: цезском, багвалинском, арчинском, мегебском и тукитинском.

Первое использование глагола ‘найти’ мы называем конструкцией обнаружения ситуации: она описывает ситуацию (текущий процесс или конечное состояние, в зависимости от аспектуального значения лексического глагола), наблюдаемую субъектом. Эти варианты использования могут иметь значение неожиданности, связанное с исходным значением глагола ‘найти’ в аваро-андо-цезских языках.

Второе использование встречается в конструкциях с презумптивной выводной семантикой: глагол ‘найти’ выполняет функцию эвиденциального/модального вспомогательного глагола, как ‘основываясь на моих представлениях о мире, вероятно, что P’.

Третье использование встречается в условных конструкциях: мы предполагаем, что использование вспомогательного глагола ‘найти’ в условной форме (‘если окажется, что P’) позволяет выразить различные аспектуальные характеристики ситуации с помощью подходящей формы лексического глагола (условное деепричастие лексического глагола само по себе не имеет аспектуальных значений).

Наконец, не на основании корпусного исследования, но опираясь на данные словарных примеров нескольких языков, мы можем утверждать, что формы глаголов со значением ‘найти’ в этих языках участвуют в образовании оптативных конструкций по схеме, схожей с образованием вышеупомянутых условных конструкций.

Результаты исследования были представлены в докладе «‘Find’ as an auxiliary in discovery constructions: data from East Caucasian» на конференции 58th Annual Meeting of the Societas Linguistica Europaea (г. Бордо, Франция). В ближайшее время планируется создание статьи по материалам корпусного исследования конструкций с глаголом ‘найти’ в цезском языке.

### 3.10.6 Некоторые маркеры будущего времени в шугнанском языке

В 2025 году продолжалось исследование фразовой энклитики =ta шугнанского языка как маркера временной референции будущего времени. На материале корпуса шугнанского языка проверялась гипотеза о том, что у этой энклитики раньше была хабитуальная семантика, которая в настоящий момент у носителей практически отсутствует.

## 3.11 Исследования морфосинтаксиса и семантики именных форм в отдельных языках

### 3.11.1 Эвалютивные суффиксы в быстринском диалекте эвенского языка

В 2025 году проводилось исследование системы эвалютивной морфологии в быстринском диалекте эвенского языка. Материал собран в экспедиции 2025 года в посёлках Эссо и Анавгай (Камчатка) и дополнен корпусными данными. Цель — описать распределение диминутивных и аугментативных суффиксов, их семантику и прагматику, а также соотнести эвалютивность с категориями определённости и посессивности.

Эвалютивные морфемы выражают отношение говорящего к обозначаемому объекту. Помимо «размера», они передают эмоциональные и оценочные оттенки: ласковость, иронию, пренебрежение. Для эвенского языка характерно взаимодействие этих значений с грамматикой определённости. В ламунхинском диалекте, описанном в [114], помимо размера эвалютивы маркируют информационный статус референта. В ламунхинском диалекте отсутствуют любые собственно оценочные значения. Быстринский диалект, напротив, сохранил большую экспрессивность, что делает его важным для реконструкции разных стадий грамматикализации.

Исследование сочетает полевые данные и корпусный анализ. В полевой программе использовались контролируемые эксперименты (называние картинок, обсуждение оценочных форм) и спонтанные нарративы, фиксировавшие естественное употребление суффиксов. Корпус, включающий около 450 контекстов, аннотирован морфологически; 56 % примеров содержат диминутивы, 44 % — аугментативы. В быстринском диалекте функционируют шесть эвалютивных суффиксов:

* диминутивы -*kakan* (INDEF), -*jakan* (DEF), -*čan* (POSS);
* аугментативы -*mkar* (INDEF), -*kaja* (DEF), -*nǯa* (POSS).

Структура системы симметрична: по три форманта в каждой категории, соотнесённые с неопределёнными, определёнными и посессивными контекстами. Суффиксы -*kakan* и -*mkar* обладают нейтральной семантикой, -*jakan* и -*kaja* — экспрессивной, нередко пейоративной. Суффиксы -*čan* и -*nǯa* связаны с посессивностью и сохраняют денотативное значение размера.

С исторической точки зрения некоторые формы восходят к тунгусо-маньчжурскому -*kan*, переосмысленному в сочетаниях -*kakan (-kan + -kan)*, -*jakan (-ja + -kan)*, -*kaja (-kan + -ja)*.

Эвалютивные морфемы быстринского диалекта функционируют не только как деривационные, но и как прагматические показатели. Суффикс -*kakan* служит базовым и наиболее частотным диминутивом (около 60 % всех диминутивных употреблений), нейтральным по оценке и свободным по сочетаемости. Он заполняет контексты, где *jakan* неприемлем из-за его отрицательной семантики.

Суффикс -*jakan* (≈ 25 % диминутивных контекстов) выражает уменьшенность, а с некоторыми лексемами появляется ярко выраженная отрицательная оценка — ‘плохой, испорченный, неприятный’. Выбор этого суффикса для таких лексем зависит не столько от определённости, сколько от субъективного отношения говорящего. Таким образом, семантические оттенки суффиксов -*kakan* и -*jakan* переписывают их поведение, обусловленное референциальным статусом именной группы.

Аугментатив -*mkar* является основным нейтральным показателем увеличенности (около ⅔ аугментативных употреблений), тогда как -*kaja* часто приобретает экспрессивную окраску (‘слишком большой, неуклюжий’). Суффикс -*nǯa* редок и преимущественно встречается в посессивных контекстах.

Эвалютивы активно сочетаются с личными именами: -*kakan* передаёт симпатию или жалость, значение -*jakan* зависит от контекста, а -*kaja* иногда выражает иронию или насмешку. Остальные морфемы обозначают исключительно размер референта. Такое употребление показывает переход от деривационного к прагматическому уровню функционирования.

Сравнение с ламунхинским материалом выявляет противоположные тенденции грамматикализации. В ламунхинском диалекте эвалютивы выполняют дискурсивные функции — различают «новую» и «данную» информацию, утратив экспрессивность. В быстринском — наоборот: эмоциональная и оценочная нагрузка преобладает, а грамматические ограничения ослаблены. Это указывает на сохранение ранней стадии развития категории, где эвалютивность остаётся частью прагматики.

Корпусные данные подтверждают баланс между нейтральными (≈ 70 % употреблений) и оценочными (≈ 30 %) формами. Эвалютивная морфология быстринского диалекта эвенского языка представляет собой динамичную подсистему, где грамматические и прагматические значения тесно переплетены. Нейтральные форманты (*-kakan, -mkar*) создают основу описания размеров, тогда как экспрессивные (*-jakan, -kaja*) реализуют эмоционально-оценочные функции. Распределение между -*kakan* и -*jakan* показывает систематическую семантическую саморегуляцию: первый нейтрализует отрицательные коннотации второго, сохраняя баланс между объективным и субъективным измерениями значения. В целом, эвалютивы быстринского диалекта демонстрируют переход от чисто деривационной категории к прагматической, отражая взаимодействие эмоциональности и грамматики в естественном языке.

По результатам исследования готовится статья к публикации в журнале Folia Linguistica.

### 3.11.2 Посессивные конструкции и показатели в кильдинском саамском, эвенском, уильтинском и эрзянском языках

В 2025 году готовилась к публикации статья, посвящённая категории посессивности в кильдинском саамском языке. Работа будет опубликована в тематическом сборнике, посвящённом саамским языкам, в журнале Acta Linguistica Petropolitana. Исследование основано на анализе и обобщении данных, собранных в предыдущих экспедициях на Кольском полуострове, а также на сопоставлении с существующими грамматическими описаниями.

В центре внимания исследования находится современное состояние посессивной морфологии кильдинского саамского языка и её взаимодействие с категориями рефлексивности и реципрокальности. Согласно традиционным грамматическим описаниям, система посессивных суффиксов в кильдинском саамском представлена тремя значениями лица, однако в современной речи отмечается значительная вариативность: часть носителей различает три лица, тогда как другая часть — только два, объединяя первое и второе лицо в единую морфему -ant. В корпусных и полевых данных А. О. Бузанова подтверждается наличие обоих типов парадигм, что указывает на диалектное и индивидуальное варьирование в пределах языка.

Особое внимание в работе уделено функционированию посессивных показателей в рефлексивных и реципрокальных конструкциях. Несмотря на то, что собственно именная посессивная парадигма в значительной степени утрачивается, в этих структурах она остаётся обязательной: формы типа *iž’-es’* (REFL-POSS.3) или *kaan’n’c’ kaan’n’c’es’* (друг друг-POSS.3) демонстрируют грамматикализацию посессивных показателей как части рефлексивной/реципрокальной стратегии. Таким образом, посессивная морфология в кильдинском саамском языке сохранилась именно в синтаксически ограниченных позициях, где она является морфологической частью средств выражения кореферентности. Кроме того, А. О. Бузанов рассматривает иерархию семантических классов существительных, допускающих посессивное маркирование. Анализ показывает, что все носители языка используют посессивные суффиксы (или притяжательные местоимения) с терминами родства, часть — с обозначениями домашних животных, и лишь немногие — с названиями предметов домашнего обихода.

В заключительной части статьи обсуждается переход от двойной/вершинной стратегии выражения посессивности к исключительно зависимостной. Совокупность данных позволяет утверждать, что кильдинский саамский язык демонстрирует позднюю стадию эрозии суффиксального посессивного склонения при сохранении его грамматических следов в области рефлексивных и реципрокальных структур. Утрата посессивного склонения происходит, вероятно, под влиянием русского, хотя аналогичные процессы наблюдаются и в других уральских языках.

Кроме того, в 2025 году по результатам проведённого в прошлом году исследования опубликована статья в журнале «Урало-алтайские исследования», в которой представлены основные выводы о функционировании рефлексивно-посессивных показателей в быстринском эвенском и уильтинском языках. Публикация закрепила результаты сравнительного анализа и заложила основы для типологической характеристики систем рефлексивной посессивности.

Стажёр-исследователь НУЛ социогуманитарных исследований Севера и Арктики К. Медушевский в 2025 году исследовал посессивные конструкции и посессивные суффиксы в больше-игнатовском диалекте эрзянского языка (мордовские, уральские). Данные были получены с помощью элицитации, в том числе онлайн, в июле 2025 г. В ходе исследования было выделено три типа посессивных конструкций:

1. двойное маркирование:
2. зависимое.GEN + вершина.DEF/POSS. б) зависимое.GEN.DEF/POSS + вершина.DEF/POSS.
3. зависимостное маркирование: а) Зависимое.GEN + Вершина. б) Зависимое.GEN.DEF/POSS + Вершина.
4. Соположение: Зависимое + Вершина.

В эрзянском в посессивных конструкциях типа 1 с местоимениями наблюдается вариативность: вершина может маркироваться либо показателем определённости (DEF), либо посессивным суффиксом (POSS) (60).



|  |  |  |  |
| --- | --- | --- | --- |
| Монь | кудо-м | / | кудо-сь |
| 1SG.GEN | дом-POSS.1SG | / | дом-DEF |
| 'Мой дом' | | | |

Зависимое, выраженное именем собственным, может принимать как генитив без значения определённости, так и определённый генитив (61).



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Андрей | те | Алёна-нь | / | Алёна-нть | мирде-сь. |
| Андрей | это | Алёна-GEN | / | Алёна-DEF.GEN | муж-DEF |
| 'Андрей --- это муж Алёны' | | | | | |

Результаты исследования были представлены в докладе «Посессивные конструкции в эрзянском языке» на конференции «СКИЛ — 2025» (Институт лингвистики РГГУ, г. Москва).

### 3.11.3 Атрибутивизаторы в лесном ненецком языке

В 2025 г. завершилось изучение атрибутивизаторов в лесном ненецком языке. В исследовании рассматриваются атрибутивизаторы в пуровском диалекте лесного ненецкого языка (самодийские, уральские). В этом языке есть пять продуктивных атрибутивизаторов, т. е. показателей, которые присоединяются к именным основам и образуют дериваты с внешним синтаксисом прилагательных (см. [таблица 3.8](#tbl-attr-purr)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 3.8 – Атрибутивизаторы в пуровском диалекте лесного ненецкого языка   | Суффикс | Семантический ярлык | Семантическая функция |  |  | | --- | --- | --- | --- | --- | | -j° | Относительный атрибутивзитор (relational attributivizer) | Материал: kuj°ʰko-j° ‘берёзовый’, происхождение: ṕetaλa-j° ‘лесной’, время: dʹŭλnu-j° ‘утренний’, назначение: ńĕ-j° ‘женский’. |  |  | | -xăj° | Локативный атрибутивизатор (locative attributivizer) | Расположение: ṕetaλʹa-xăj° ‘находящийся в лесу’, время: čŭkʹæ dʹaλʹa-xăj° (этот день-loc.attr) ‘сегодняшний’. |  |  | | -samæ | Проприетив | Обладание: ńamt°-samæ ‘рогатый’. |  |  | | -dʹaʔj° | Облинатив | Покрытие: kĕm-dʹaʔj° ‘покрытый кровью’. |  |  | | -šă-ta | Каритив | Отсутствие обладания: ńamt°-šă-ta ’безрогий. |  |  | |

В исследовании анализируется семантика показателей и выделяются два класса атрибутивизаторов. К первому классу относятся «относительные» атрибутивизаторы, выражающие некоторые посессивные отношения, не покрываемые генитивом. Второй класс образуют два проприетивных показателя, также выражающие посессивные отношения, но образующие конструкцию, синтаксически обратную посессивной: в ней посессор является вершиной, а обладаемое — зависимым. К этому типу также относится каритивный атрибутивизатор, выражающий значение, обратное проприетивному: ‘имеющий Х’ vs. ‘не имеющий Х-а’. Атрибутивизаторы в лесном ненецком образуют систему с более и менее семантически базовыми показателями.

В исследовании также рассматриваются морфосинтаксические свойства атрибутивизаторов и приводится структура именной группы, способной атрибутивизироваться.

Кроме того, приводятся генетические и типологические параллели, показывающие типичные и уникальные черты системы атрибутивизаторов в лесном ненецком языке.

Данные для исследования были собраны в экспедициях в г. Тарко-Сале и д. Харампур (ЯНАО) в 2023–2025 гг.

Статья о семантике атрибутивизаторов в пуровском диалекте лесного ненецкого языка планируется к подаче в журнал «Linguistica Uralica». В будущем планируется работа над второй статьёй, посвящённой системе атрибутивизаторов в лесном ненецком языке в контексте генетических параллелей с другими самодийскими и уральскими языками и типологических параллелей в языках Сибири.

## 3.12 Исследования вариативности, восприятия и форм бытования языка

### 3.12.1 Диалектная лексическая вариативность в кильдинском саамском языке

В ходе экспедиции в январе-феврале 2024 года были собраны данные по диалектной лексической вариативности кильдинского саамского языка. Исторически кильдинский саамский включал четыре основных идиома: ловозерский, вороньинский, чудзьяврьский и варзинский диалекты; в связи с переселением носителей разных диалектов в село Ловозеро в 1960-х годах, небольшим числом носителей и значительной языковой аттрицией у части из них диалектная вариативность кильдинского саамского до сих пор слабо изучена. Целью работы было составить общую картину диалектных лексических различий и описать сохранившиеся различия между ними. Работа проводилась с носителями трёх диалектов: ловозерского, вороньинского, чудзьяврьского.

Данные были собраны по двум анкетам, основанным в первую очередь на «Сопоставительно-ономасиологическом словаре диалектов карельского, вепсского и саамского языков» [115], поскольку это единственный источник, в котором представлены два диалекта: ловозерский и варзинский. Первая анкета включала в себя концепты, для которых указанные в «Словаре…» диалектные варианты не являлись когнатами, во вторую вошли пары, где одно из слов было русским заимствованием.

Анализ собранных данных производился с помощью различных методов многомерного шкалирования. Результаты показывают, что носители чудзьяврьского диалекта демонстрируют большую устойчивость по сравнению с носителями ловозерского и вороньинского диалектов, которые малоразличимы на уровне лексики. Исследователи пришли к выводу, что это связано с большей близостью между ловозерским и вороньинским диалектами по другим (грамматическим, фонетическим) признакам, которая ускорила ассимиляцию диалектов после переселения. Важными факторами также оказались социальные связи конкретных носителей и уровень владения языком. Так, носители с меньшей аттрицией оказывают влияние на лексику носителей с более низким уровнем владения кильдинским саамским; были также обнаружены отдельные случаи, когда лексика конкретного носителя сблизилась с лексикой носителя другого диалекта, поскольку два этих носителя чаще всего общаются по-саамски между собой.

В настоящее время готовится к публикации статья с результатами исследования на основе этих данных и сравнения их с проводившимися ранее исследованиями грамматической и фонетической вариативности. В ходе работы были как найдены дополнительные подтверждения известных принципов ассимиляции (культурная лексика сильнее всего подвержена изменениям, тогда как фонетика сохраняется наиболее стабильно), так и установлены отдельные особенности диалектной ситуации кильдинского саамского. Наиболее важными факторами вариативности в небольшом сообществе с большим количеством контактов между носителями оказались социальные: престижность диалектов, социальные связи отдельных носителей, а также лингвистические — близость диалектов между собой.

### 3.12.2 Морфосинтаксические аспекты утраты корякского языка

В 2025 г. были изучены лингвистические проявления языковой аттриции у корякско-русских билингвальных носителей. Исследование основывается на полевом материале, собранном в 2024 г. в с. Анавгай (Быстринский район, Камчатский край). Анализ полевых данных выявил ряд тенденций, связанных с редукцией морфосинтаксической системы: непоследовательность в падежном маркировании аргументов, колебания в оформлении субъектно-объектного согласования и редукцию системы пространственных падежей.

Так, имеются наблюдения о вытеснении пролатива локативом. Носители редко сами произносят пролативные формы, некоторые не узнают пролативные формы существительных, но чаще соглашаются, что между локативом и пролативом лучше выбрать пролатив.

Локатив более всего вытесняет пролатив в контекстах движения в замкнутом пространстве, например, ‘люди кочевали по тундре’, ‘медведь ходит по лесу’. Пролатив лучше сохраняется в контекстах движения вдоль вытянутого ориентира: ‘идти по дороге’, ‘идти вдоль реки’. Другие вместо пролативных форм используют деривацию с суффиксом *-taʝn* (62):



|  |  |  |
| --- | --- | --- |
| kmiŋ-u | ∅-ko-ječve-la-ŋ | jaja-tajnə-k |
| ребенок-PL | 2/3.S/A-NFUt-игра-PLZ-NPFV | дом-EDGE-LOC |
| 'Дети бегают вокруг дома' | | |

Помимо изменений в парадигматических системах, аттриция вызывает изменения и в синтаксисе. Они касаются оформления участников в переходных предложениях. В стандартном корякском переходные клаузы имеют эргативное оформление: A-участник маркируется инструменталисом, а S- и P-участники — номинативом. У носителей с аттрицией маркирование участников нейтральное — субъект и прямое дополнение в переходном предложении маркируются номинативом (63–64), что можно считать влиянием русского морфосинтаксиса:



|  |  |  |
| --- | --- | --- |
| laŋe | ∅-ku-kukejvə-ŋ | əpa-ŋa |
| девушка | 2/3.S/A-NFUT-варить-NPFV | суп-NOM.SG |
| 'Девушка варит суп' | | |



|  |  |  |  |
| --- | --- | --- | --- |
| əmma-na | ɣe-jəl-lən | ŋavakəkə-ŋ | əpa-ŋa |
| мать-AN | PF-дать-PF.3SG | дочь-DAT | суп-NOM.SG |
| 'Мама дала дочке похлебку' | | | |

### 3.12.3 Норма и вариативность в орфографиях эвенского языка

В 2025 г. создана статья об орфографии, использующейся эвенами в Быстринском районе Камчатского края.

В статье представлены два текста, написанных на эвенском языке двумя носительницами. В ней рассказывается, с какими сложностями сталкиваются носители быстринского идиома при записи своей речи. Описываются существовавшие в разные периоды варианты литературной орфографии (их более семи) и фонетические особенности диалекта, вызывающие сомнения при выборе написания. В статье рассматриваются все проблемные случаи, встретившиеся в двух текстах, они объясняются отличиями идиома от литературного языка, для которого создавалась стандартная орфография, и русской интерференцией. Спорные моменты сопровождаются подробными фонетическими комментариями, как что произносят сами носительницы, ведь помимо письменных вариантов имеются и аудиозаписи этих текстов. Обсуждаются разные стратегии носительниц в схожих случаях.

Так, целый ряд наблюдений связан с использованием графемы . Примечательно, как часто в одном из текстов встречается графема , вообще отсутствующая в стандартном эвенском алфавите. Её появление обусловлено двумя причинами. Первая группа употреблений связана с редукцией: она иногда возникает на месте редуцированных гласных [a] и [e]: *акындуву* (1), *быиӈытын* (6), *ырыглэ* (9) (ср. с *акандуву, бэиӈэтэн, эрэглэ* в литературной орфографии). Замечено, что звук [e] более склонен к редукции, чем [a], хотя точных статистических данных пока нет. Это можно увидеть, сравнив, например, *ывышки (эвэски)* и *ырыглэ (эрэглэ)* с одной стороны и *амышки (амаски)* с другой. Более того, нет засвидетельствованных случаев, когда в начале слова /a/ писался бы как и произносился бы как [ə]. Вторая группа употреблений связана с русской интерференцией — как фонетической, так и орфографической.

Графема часто встречается вместо [i] после некоторых согласных: , , и : *тытэл* (1), *биддыди* (11), *гуныди* (13), *Чукоткалы* (2). Однако эти употребления не последовательны: *нулгыдитын* (5), *ичушнин* (10), *былиддын* (7). Звук [i] никогда не записывается как после [ǯ], [č], [m], [b], [k], [r] или после гласных: *гуныди* (13), *Орочилду* (13), *мин* (1), *биддыди* (11), *амышки* (11), *уррин* (14), *аич* (10). Тенденция отражать на письме [i] как объясняется особенностями палатализации в русском и эвенском языках. В русском языке согласные сильно коартикулируют следующим за ними гласным, непалатализованный согласный не может сочетаться с [i]. В эвенском же языке согласные перед [i] не палатализируются. Из-за этого сочетания [di], [ti], [ni], [li], не соответствуют , , , в русском тексте. Поэтому носителям приходится каждый раз выбирать, отразить ли непалатализованность согласного в таком сочетании с гласным и написать или же сохранить качество гласного, написав , хотя согласный перед ним не смягчается.

Авторы представили черновой вариант статьи на семинаре НУЛ социогуманитарных исследований Севера и Арктики 21 октября 2025 года.

### 3.12.4 Восприятие монолингвами слов на русском и чукотском языках

Участники проекта совместно с коллегами из Института нейрохирургии им. Бурденко и Института высшей нервной деятельности РАН проводили исследование восприятия звучания собственного имени в сравнении с другими стимулами на русском и чукотском языках.

Здоровым русскоязычным испытуемым, в числе которых были 23 женщины и шесть мужчин, в ходе эксперимента давалась для прослушивания последовательность, включающая пять стимулов: собственное имя испытуемого, другое имя собственное на русском языке, близкое именам по звучанию имя нарицательное на русском языке, контрольное имя собственное, отличное по звучанию от этих слов, а также чукотское слово такой же слоговой структуры, как и все остальные слова в парадигме. Собственные имена и близкое по звучанию русское слово имели одинаковый по длительности и качеству второй слог, а первый слог различался (например, *Маша, Саша и каша*). Короткие формы имён, использованные в эксперименте, признавались испытуемыми в ходе предварительного анкетирования как употребимые по отношению к ним. Чтобы избежать влияния различной частотности слов на восприятие, были выбраны частотные русские имена с контролем частотности по частотному словарю. Использованное в эксперименте чукотское слово также состояло из двух слогов, но пять из шести звуков этого слова не имеют аналогов в русском языке и были непривычны для испытуемых, которые не говорят на чукотском.

Стимулы предъявлялись по 35 раз каждый в псевдослучайной последовательности, межстимульный интервал менялся от стимула к стимулу, чтобы предотвратить привыкание, и составлял 1300–1700 мс. Во время предъявления парадигмы у участников эксперимента регистрировалась ЭЭГ от 32 электродов, фиксированных на поверхности кожи головы. Целью была регистрация потенциалов, связанных с событиями, и сопоставление их для разных видов стимулов.

Анализ показал отличия потенциалов, связанных с событиями, в ответ на собственное имя человека и в ответ на другие слова. Ответ на чукотское слово характеризуют иные, чем собственные имена и другие слова на русском языке, значения ряда параметров уже на более ранней латентности, в начале первого слога, тогда как статистически значимые отличия между русскими словами у испытуемых фиксируются в конце звучания первого слога.

Можно предположить, что чукотский стимул, звучание которого не отвечало фонетическим и смысловым ожиданиям слушающих, воспринимался мозгом как музыка или шум, то есть неречевой стимул. Первичная автоматическая оценка этого стимула как неречевого производилась ещё до того времени, в которое начинается смысловая дифференциация стимулов.

Результаты исследования изложены в статье [116] в журнале «Физиология человека».

### 3.12.5 Цифровая документация эвенского языка: устные и письменные тексты

В 2025 г. продолжилась разработка эвенского морфологического анализатора с использованием конечных преобразовывающих автоматов [[2]](#footnote-183).

Помимо этого, расшифровывались устные тексты для цифрового корпуса эвенского языка. За весенний период экспедиционной командой расшифровано 8 песен на эвенском языке, 6 минут монологической речи, для которых не было перевода на русский язык, и 1 минута диалогической речи на эвенском языке. За осенний период все тексты, до этого переведённые и сопровождённые поморфемной нотацией, были перенесены в проект FieldWorks. Среди них и устные тексты, записывавшиеся с 2019 по 2025 год, и сказки К. С. Черканова, и несколько выпусков газеты «Айдит», которая выходила с 1990-х годов по 2005 год. Программа позволяет унифицировать имеющиеся тексты, в том числе графику, в которой записаны сами эвенские слова, варианты однословных переводов эвенских корней, глоссы, менявшие названия за семь лет существования эвенского экспедиционного проекта. Эта работа способствует дальнейшему оформлению данных в корпус в наиболее аккуратном виде.

### 3.12.6 Вариативность структур и восприятие поэтического текста

В 2025 г. развивалось стиховедческое направление работы лаборатории. В рамках этого направления подготовлена статья «Теневая рифма в поэме В. А. Гандельсмана “Видение”» к подаче в журнал «Зборник Матице Српске за славистику».

Теневая рифма — это рифма, которую образуют группы слов, уже рифмующиеся между собой. Теневая рифма часто бывает неточной. Поэма В. А. Гандельсмана «Видение» интересна тем, что это произведение крупной формы, в котором теневая рифма является частью строфики. В ходе исследования показано, какие типы теневых рифм встречаются в поэме. Самыми частотными оказались разноударные рифмы, то есть такие, в которых различается позиция ударного гласного. Эти рифмы встречаются в русскоязычной поэзии очень редко [117], поэтому поэма «Видение» — важный источник данных для их исследования. На материале рифм поэмы рассмотрен вопрос о рифменном домене разноударных рифм. В поэме «Видение» большинство разноударных рифм относится к двум типам: либо слова с ударением на первый слог рифмуются со словами с ударением на третий с конца слог, либо слова с ударением на второй с конца слог рифмуются с ударением на четвёртый с конца слог.

Исследован вопрос о рифменном домене в разноударных рифмах. Для рифм русского языка, как и для рифм большинства европейских языков, рифменный домен — ударный гласный и все звуки после него [117]. Однако если позиция ударного гласного в рифмующихся словах различается, то положение границ рифменного домена неочевидно. Для англоязычной поэзии, где разноударные рифмы распространены сильнее, чем в русскоязычной, рифменный домен в разноударных рифмах обычно отсчитывается от последнего ударного гласного [118]. Показано, что в поэме «Видение» рифменный домен тоже рассчитывается от последнего ударного гласного. Исследование рифменного домена опиралось как на фонетику, так и на графику, так как для русских разноударных рифм графическое сходство слов играет важную роль [117]. Оказалось, графический облик рифмующихся слов в зоне после первого ударного слова совпадает чаще, чем после второго ударного слога. Проблема, возникающая при оценке фонетического сходства разноударных рифм, связана с тем, что эти рифмы неточные. Поэтому исследовательница воспользовалась методом оценки фонетического сходства неточных рифм, разработанным в [118]. Оказалось, что фонетическое сходство в зоне после второго ударного гласного статистически значимо сильнее, чем в зоне после первого ударного гласного.

Защищена магистерская диссертация по теме «Фонетика диссонансной рифмы в русскоязычной поэзии второй половины XX — начала XXI веков». Диссонансные рифмы — рифмы, в которых различается ударный гласный. Описана роль, которую диссонансные рифмы могут играть в строфике стихотворений, а также то, чем, кроме ударных гласных, могут различаться слова в диссонансных рифмах. Показано, что закономерности, которые другие исследователи выделяли для рифмующихся согласных, не могут описать то, как рифмуются ударные гласные. Однако важную роль в том, как рифмуются ударные гласные, играет коартикуляция этих согласных с предыдущим гласным.

Кроме того, проведено типологически ориентированное исследование того, как рифмуются согласные в диссонансных рифмах русскоязычной поэзии. Оказалось, что ключевую роль в том, какие согласные в неточных рифмах рифмуются, играет их совпадение по месту образования. Эта тенденция выражена у переднеязычных и у заднеязычных согласных; у губных она выражена слабее. Исследование показало, что в неточных рифмах, к которым принадлежат и диссонансные, требование к фонетическому сходству ударных гласных слабее. Это противоречит предыдущим исследованиям [118], согласно которым в неточных рифмах степень фонетического сходства обоих типов звуков одинаковая. Кроме того, показано, что для диссонансных рифм нет строгого требования к фонетическому сходству инициали ударного слога.

Сделан доклад на конференции «Компьютерная лингвистика и корпусные технологии» 3 июня 2025 года.

В 2025 г. проведено экспериментальное исследование восприятия непонятных слов в стихе и прозе на русском языке.

В эксперименте участвовали 98 респондентов. Возраст респондентов варьируется от 17 до 74 лет, средний возраст — 23 года. Большинство испытуемых были женщины (86,6 %), образование респондентов не ниже основного общего (с таким уровнем образования в выборке 1 %), большинство — студенты высшего учебного заведения (57,8 %), следующий по распространенности уровень образования — среднее общее (15,5 %), среднее профессиональное есть только у одного испытуемого (1 %), остальные респонденты имеют законченное высшее образование: специалитет или бакалавриат, магистратуру, аспирантуру (8,2 %, 9,3 %, 6,2 % и 1 % соответственно, всего 24,7 %). Стоит также отметить, что 49,5 % респондентов связаны с лингвистикой или филологией.

Было отобрано восемь стихотворных текстов, четыре из которых содержали потенциально непонятные слова: повилика, вольфрам, боливар и раструбы. Тексты выбирались из существующих стихотворений. Для каждого из текстов был сделан его прозаический пересказ, отличающийся не только записью в строку, но и тем, что из них были убраны особенности порядка слов, связанные с особенностью поэтической формы, ритм и рифма. Данный метод был разработан М. Л. Гаспаровым и Т. В. Скулачевой — благодаря тому, что в прозе сохраняется тематика и подбор слов, характерные для стиха, становится возможным изолированно изучать эффект формы текста. Были составлены четыре опроса на платформе Google Forms, два из них содержали оригинальные стихотворные тексты, другие два — их пересказы в прозе; порядки текстов одного формата были обратны друг другу, чтобы исключить влияние фактора порядка. В обоих случаях содержание опроса было таким: тренировочный текст и по очереди текст без потенциально непонятного слова и текст с потенциально непонятным словом.

Опросы состояли из двух частей: в первой респондентам по одному показывали тексты и по два вопроса к ним («Оцените, насколько вам понравился текст» и «Оцените, насколько этот текст сложный» — оба по шкале от 1 до 7, где оценка 7 считалась ответом «очень понравился» и «очень сложный» соответственно), во второй части был предложен список из восьми слов (*повилика, вольфрам, боливар, раструбы, щеточка, румяна, приговор, забавы*), содержащихся в текстах, и респондентов просили написать лексические значения каждого без использования дополнительных материалов.

Для анализа полученных данных был проведен статистический анализ с помощью языка программирования R. Был выбран метод линейной регрессии со смешанными эффектами, в последние годы получивший широкое распространение у экспериментальных психологов и лингвистов. Использовался модуль glmer из библиотеки lme4. Для расчета значимости p-уровней была использована библиотека lmerTest.

Для начала важно отметить то, как распределились показатели того, насколько верные определения респонденты дали словам, которые были изначально обозначены как понятные и непонятные. При обработке результатов их ответы были вручную заменены на 0, 1 и 2, где 0 — указанное значение неверно, 1 — верно частично, 2 — указанное значение полностью верно. По итогам анализа все потенциально понятные слова действительно оказались понятными (показатели точности предложенных испытуемыми дефиниций каждого из четырех слов очень близки к 2). Однако среди потенциально непонятных слов у «вольфрама» данный показатель оказался неожиданно высоким (1,52 при среднем 0,84 у остальных непонятных слов). Это вынудило исключить его из анализа вообще — таким образом, фактор сложности в исследовании опирается на четыре понятных слова и три непонятных там, где это важно.

Регрессионный анализ показал, что между группой с понятными словами и группой с непонятными словами есть не только видимая, но и значимая разница в правильности ответов (β = 1,033, SE = 0,148, p < 0,001). При этом разница между прозой и стихом несущественна (β = 0,120, SE = 0,088, p = 0,177).

С другой стороны, был проведен анализ корреляции между стихом и прозой для привлекательности, субъективной сложности и понимания слов. Во всех случаях связь сильная: коэффициенты корреляции r = 0,802, r = 0,887 и r = 0,973 соответственно. Из этого можно сделать вывод, что оценивание текстов по трем указанным параметрам было одинаково для прозы и стиха. Это важно отметить, так как этот факт свидетельствует о том, что прозаические пересказы, используемые в качестве стимулов, не испортили впечатление от текста.

Анализ показал, что корреляция между сложностью, привлекательностью и пониманием почти отсутствует. Анализ проводился попарно, без деления на поэзию и прозу (так как, как было показано выше, их показатели совпадают). Очень слабая связь (r = 0,050) выявляется между тем, нравится ли текст респондентам, и тем, присутствуют или отсутствуют в нем непонятные слова; совершенно нет связи (r = -0,006) между содержанием во фрагменте непонятных слов и его субъективной сложностью. Слабая положительная связь прослеживается у субъективной сложности и привлекательности текстов (r = 0,269).

Для анализа того, как понимание слов влияет на оценки привлекательности и субъективной сложности, была проведена линейная регрессия со смешанными эффектами. На субъективную сложность фрагментов фактор понимания никак не влияет (β = 0,168, SE = 0,187, p = 0,369), однако же в прозе отсутствие непонятных слов снижает привлекательность, при этом в стихе такого эффекта не наблюдается. Вместе с этим, проза в целом нравится меньше поэзии (что может считаться нормальным, так как это может быть общий эффект того, что используемые в эксперименте прозаические фрагменты — это переделанные стихи), и то, насколько эта разница велика, зависит от того, понятные или непонятные слова содержат тексты.

Кроме описанного выше, с помощью статистического анализа было проанализировано и то, как соотносятся собранные метаданные респондентов и оценки, которые они давали текстам, а также как хорошо они знали лексические значения непонятных слов. Однако значимость рассматриваемых данных рассчитана не была, поэтому анализ имеет предварительный описательный характер.

В исследовании было описано влияние возраста: чем старше респонденты, тем более субъективно сложными и менее привлекательными им показались тексты. При этом понимание слов примерно равно как между возрастными группами, так и между разными форматами внутри одной группы.

Влияния уровня образования информантов на их оценки субъективной сложности и привлекательности фрагментов не выявлено. Однако есть закономерность в уровне понимания слов: студенты знают значения слов лучше, чем люди со средним и с полным высшим образованием.

Мужчины в среднем считают поэзию субъективно сложнее и привлекательнее, чем женщины, которые, в свою очередь, считают прозу более сложной и привлекательной; разницы в уровне понимания слов не выявлено.

У групп респондентов, имеющих и не имеющих отношения к филологии или лингвистике, разница в оценивании субъективной сложности и привлекательности фрагментов была существенна, более высокую среднюю оценку ставила первая группа, однако при этом большого различия в уровне понимания слов между группами не обнаружено.

# 4. Языковое многообразие в формально-лингвистических моделях и их эмпирическая база

По мере накопления эмпирических знаний о языках России все более актуальным становится построение новых и уточнение существующих теоретических моделей с учетом этих знаний. В то же время разрабатываемые формальные модели служат более мощным и современным инструментом анализа эмпирических языковых данных. Одной из задач настоящего исследования является представление данных языков России с использованием формальных моделей (формальной семантики, синтаксиса, прагматики). Эта задача достаточно объемная и многоаспектная. В 2025 г. для моделирования были выбраны следующие классы явлений: кодирование временных отношений в языке, анафора и связывание, передвижение в полипредикации, согласование с нетривиальными контролерами, полярность (логические операторы в языке). Помимо этого, разрабатывались ресурсы, служащие эмпирической базой для разрабатываемых моделей.

Языки России, которые попали в сферу интереса, — это русский, хантыйский, абазинский, кильдинский саамский, мишарский татарский и др.

## 4.1 Семантика и синтаксис глагольных категорий

Одна из магистральных тем, которая разрабатывалась в 2025 году, включала семантику и синтаксис глагольных категорий. К этой области относятся и собственно формы глагола, маркирующие время (финитные и нефинитные), так и лексические показатели времени — темпоральные наречия. Работа проводилась на материале кильдинского саамского и казымского хантыйского языков, собранном в 2022–2025 г. в ходе экспедиций в с. Ловозеро Мурманской обл. (кильдинский саамский) и с. Казым Белоярского района ХМАО (казымский хантыйский). В саамском изучалась нефинитная форма со значением прошедшего времени, в хантыйском — взаимодействие интерпретации временных форм с временными наречиями. Для обоих явлений разработаны эксплицитные формальные модели.

### 4.1.1 Индексикальные наречия несовместимы с относительным временем: данные из казымского хантыйского

Исследование по хантыйскому языку было частью многолетнего проекта, целью которого является теоретическое моделирование вложенного времени в казымском диалекте хантыйского языка. Основной проблемой, которая анализировалось в этом году, были ограничения на сочетаемость вложенного (подчинённого) времени и индексикальных наречий. Данная группа наречий сочетается только с абсолютным, но не с относительным маркированием времени.

#### 4.1.1.1 Время и временные наречия в казымском хантыйском

В казымском хантыйском относительные клаузы образуются при помощи причастных форм, которые могут быть маркированы либо прошедшим (NFIN.PST), либо непрошедшим (NFIN.NPST) нефинитным временем. Для нас релевантно то, что нефинитное время в относительных клаузах (ОК) может интерпретироваться и как абсолютное (т. е. упорядочивать время события ОК и время речи), и как относительное (т. е. упорядочивать время события ОК и время матричной клаузы). Ниже приведены примеры для прошедшего времени — в (65)–(66) проиллюстрирована относительная интерпретация, а в (67) — абсолютная.



|  |  |  |  |
| --- | --- | --- | --- |
| waśaj-en | joŋtəs-ti | imi | woχ-əs |
| Вася-POSS.2SG | шить-NFIN.NPST | женщина | звать-PST［3SG］ |
| 'Вася позвал женщину, которая шила (пока он звал)' | | | |



|  |  |  |  |
| --- | --- | --- | --- |
| waśaj-en | joŋtəs-əm | imi | woχ-əs |
| Вася-POSS.2SG | шить-NFIN.PST | женщина | звать-PST［3SG］ |
| 'Вася позвал женщину, которая шила (до того, как он позвал)' | | | |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ［in | jak-əm | ewij-en］ | siri | ma | wankʉtλi |
| сейчас | танцевать-NFIN.PST | девочка-DIM.POSS.2SG | раньше | я | часто |

|  |  |  |  |
| --- | --- | --- | --- |
| want-ijəλ-s-əm | in | šɛŋk | ăntɵ |
| видеть-FREQ-PST-POSS.1SG | сейчас | вообще | NEG.EX |
| 'Девушку, которая только что танцевала, я раньше часто видел, а сейчас вообще не вижу' | | | |

#### 4.1.1.2 Анализ

В результате было принято решение анализировать время в ОК как относительное, т. е. интерпретирующееся относительно ближайшего временного якоря и не привязанное к времени речи. Такой подход поддерживается независимыми синтаксическими данными о том, что ОК не содержат проекций выше TP [119; 120] и допущением о том, что временное местоимение t\*, отсылающее ко времени речи, может вводиться только в верхних синтаксических проекциях ([121–123], среди прочих). Ниже приведены лексические вхождения, которые постулируются для двух нефинитных времён.

1. Нефинитное прошедшее:
2. Нефинитное непрошедшее:

С учётом изначального допущения, что время в ОК глубинно относительное, абсолютные интерпретации анализировались как подъём ИГ, содержащей ОК, в логической форме (однако в большинстве примеров мы можем видеть и синтаксическое передвижение ИГ ближе к левой границе клаузы). Анализ абсолютного времени ИГ как подъёма квантора был изначально предложен Т. Стоуэллом [123] для английских относительных клауз и позже М. О’Лири [125] для английских существительных. Согласно такому анализу, все ИГ глубинно интерпретируются с относительным временем. В случаях, когда возникает абсолютная интерпретация, ИГ передвигается в более высокие проекции, которые расположены выше финитного времени, но ниже местоимения *t*\*. Таким образом, новым локальным якорем становится время речи, к которому отсылает *t*\*. Рассмотрим относительную интерпретацию ОК, проиллюстрированную в предложении (70). Как показано в схеме (71), временной аргумент ИГ *in situ* локально связывается временем матричного предиката.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| pasaŋ-ən | ma | ［χʉλ | λɛw-əm］ | iki | šiwaλ-əs-əm |
| стол-LOC | я | рыбу | есть-NFIN.PST | человек | увидеть-PST-1SG |
| 'Я увидела мужчину, кто (до этого) ел рыбу за столом' | | | | | |
| Временной аргумент: время матричной клаузы → поедание рыбы до времени видения | | | | | |

1. Локальное связывание: *рыбу едящего мужчину увидела*

Для абсолютного времени, показанного в (72), ситуация обратная. ИГ поднимается в домен времени речи и там локально связывается местоимением t\*, см. схему (73).



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ［išńi | jeλpijən | χɵχətˊλˊ-əm］ | amp-en | mʉŋ | muλχatəλ | wɵnt-ən | want-s-ew |
| ［окно | до | бежать-NFIN.PST］ | собака-POSS.2SG | мы | вчера | лес-LOC | увидеть-PST-1PL&gt;SG |
| 'Мы видели собаку, которая только что бегала под окнами, вчера в лесу' | | | | | | | |
| Временной аргумент: время речи → собака бегает под окнами до времени речи | | | | | | | |

1. Подъём и локальное связывание: *под окнами бегавшая собака увидели*

Анализ абсолютного времени в ОК как подъёма кванторов поддерживается независимыми данными о том, что на время не влияет референциальный статус ИГ, но при этом, если ИГ связана более высоким квантором, это блокирует абсолютную интерпретацию. Рассмотрим пример (74), где ИГ с ОК связана квантором *никто*. Заметим, что маркирование непрошедшим временем, которое интерпретировалось бы как абсолютное, здесь недоступно.

1. [Завтра в школе будет концерт. Нашим мамам сказали принести наши костюмы обратно в школу сразу после того, как закончится концерт.]

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| nɛm\_χujat | ［ńawrɛm-əλ | {λɵmtəλ-əm | / | #λɵmət-ti}］ | λɵmətsuχət | imuχti |
| никто | ребёнок-POSS.3SG | носить-NFIN.PST | / | носить-NFIN.NPST | одежда | сразу |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| juχλi | ăn | tɵt-λ-aλ | pɵs-ti | χăj-λ-əλaλ |
| обратно | NEG | принести-NPST-1SG&gt;SG | мыть-NFIN.NPST | оставить-NPST-3SG&gt;NSG |
| 'Никто (из матерей) не принесёт обратно одежду, которую носил её ребёнок, они оставят её постирать' | | | | |

1. [Завтра в школе будет концерт. Нашим мамам сказали принести наши костюмы обратно в школу сразу после того, как закончится концерт.]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| nɛm\_χujat | ńawrɛm-əλ | #χaλewət | λɵmtəλ-əm | λɵmətsuχət |
| никто | ребёнок-POSS.3SG | завтра | носить-NFIN.PST | одежда |
| 'Никто (из матерей) не принесёт обратно одежду, которую (#завтра) носил её ребёнок, они оставят её постирать' | | | | |

При этом любопытно, что в случае абсолютной интерпретации в ОК не могут вставляться индексикальные временные наречия. См. пример (75), где относительное прошедшее время несовместимо с наречием *χaλewət* ‘завтра’. Также см. пример (76), где маркирование относительным временем недоступно при наличии наречия *tămχatəλ* ‘сегодня’. Анализ абсолютного времени как подъёма кванторов в своём текущем состоянии не предсказывает того, что наречия также влияют на интерпретацию времени, и требует некоторой доработки.



|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ma | muλχătəλ | ［tămχatəλ | arij-əm | / | #ari-ti］ | ajλat | puχ | want-s-ɛm |
| я | вчера | сегодня | петь-NFIN.PST | / | #петь-NFIN.NPST | молодой | сын | увидеть-PST-1SG&gt;SG |
| 'Вчера я видела юношу, который сегодня пел [на концерте]' | | | | | | | | |

#### 4.1.1.3 Анализ наречного сдвига

На материале английского языка Э. Кешет [126] замечает, что существительное и его модификатор должны интерпретироваться относительно одного и того же времени. Поэтому пример (77) неграмматичен — сенатор не может быть одновременно и студентом Гарварда.

1. #*In 1964, every U.S. Senator (then) at Harvard got straight A’s.* [126]
2. *Every U.S. Senator who was at Harvard in 1964 got straight A’s in college.* [126]

Для того, чтобы предсказать это свойство адноминальных модификаторов, Кешет [126; 127] вводит *Обобщение об интерсективных предикатах* (*Intersective Predicate Generalization*, IPG), приведённое в (79).

1. Два предиката, соединённые при помощи Предикатной Модификации, не могут оцениваться в разных друг от друга временах или мирах.

Заметим, что интерсективные модификаторы включают в себя не только прилагательные, но и наречия. Если уточнить определение предикатной модификации, то IPG может также применяться и к случаям, когда наречие модифицирует временной аргумент у клаузы, см. имплементацию в [128]. Мы будем использовать обобщённую предикатную модификацию, которую вводит Кешет.

1. Operator: Для любых функций и с объединимыми типами, a) , если , или б) , если . [126]
2. (Обобщённая) Предикатная Модификация: Если — ветвящийся узел, — множество дочерей и и$ ⟦γ ⟧$ являются функциями с объединимыми типами , тогда . [126]

Тем самым, использованный объяснительный аппарат позволяет предсказать, что относительное время несовместимо с индексикальными наречиями, т. к. оно связано локально доступным временем матричной клаузы, а наречия интерпретируются относительно времени речи. При этом в сентенциальных актантах этот механизм работает менее эффективно, так как в них возможны конфигурации, где сосуществуют относительное время и индексикальные наречия. Дальнейшая работа над темой, как мы надеемся, позволит решить вопрос объяснения и для этой группы явлений.

### 4.1.2 Семантика причастия прошедшего времени в кильдинском саамском языке

#### 4.1.2.1 Введение

В кильдинском саамском языке (< саамские < уральские) есть нефинитная глагольная форма на -*(m)a* (также у нее есть диалектный вариант -*(m)eɲč*), встречающаяся в различных контекстах: она может использоваться в качестве приименного модификатора, или причастия (82); будучи в соседстве с отрицанием, она может маркировать коннегатив прошедшего времени (83); а также она встречается в перфекте (84) и плюсквамперфекте (85). В нашем исследовании мы предлагаем моносемический анализ для формы на -*(m)a*.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| munn | janda | and-a | ɲaallks-et' | udl-a | lɨhk-ma |
| 1SG.NOM | завтра | дать-NPST.1SG | сладкое-ACC.PL | быстро-COMP | делать-PTCP.PST |

|  |
| --- |
| paarn-et' |
| ребенок-DAT.PL |
| 'Я дам сладости тому ребенку, который сделал (это) быстрее.' |
| {Говорящий дает детям задание. Завтра он спросит, кто успел выполнить его быстрее других.} |



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| munn | emm | pued | / | puadd-ma |
| 1SG.NOM | NEG.1SG | прийти.NPST.CNG | / | прийти-PTCP.PST |
| 'Я не приду / не пришел.' | | | | |



|  |  |  |  |
| --- | --- | --- | --- |
| soonn | l'ii | puadd-ma | peer̥'r̥'t'-e |
| 3SG.NOM | быть.NPST.3SG | прийти-PTCP.PST | дом-DAT.SG |
| 'Он пришел.' | | | |



|  |  |  |  |
| --- | --- | --- | --- |
| soonn | l'aajj | puadd-ma | peer̥'r̥'t'-e |
| 3SG.NOM | быть.PST.3SG | прийти-PTCP.PST | дом-DAT.SG |
| 'Он пришел (до какого-то момента в прошлом).' | | | |

#### 4.1.2.2 Семантика формы на -*(m)a*

Общий элемент значения -*(m)a* в (82)–(85) — это семантика предшествования: в (82) событие, маркированное -*(m)a*, предшествует некоторой референциальной точке, заданной матричным предикатом; в (83) показатель -*(m)a* задает предшествование моменту речи; кросс-лингвистически в (плюсквам)перфекте тоже часто наблюдается семантика предшествования некоторой референциальной точке (среди многих других [129–133]). В случае перфекта (84) референциальной точкой можно, упрощая, считать момент речи, а в случае плюсквамперфекта (85) — какую-то другую точку в прошлом, заданную контекстом.

Далее мы будем, следуя [129] и [130], формулировать свои наблюдения в терминах неорейхенбахского подхода к временной семантике. Для нас будут релевантны четыре временные точки: время события (event time, далее ET), топикальное время (topic time, далее TT), время перспективы (perspective time, далее PT) и время речи (utterance time, далее UT).

В описании выше мы упомянули, что в примерах с отрицанием типа (83) -*(m)a* может располагать событие относительно момента речи. В таком случае его поведение похоже на абсолютное прошедшее время, так как стандартно считается, что абсолютное прошедшее время располагает TT до UT [129; 134]. Однако другие употребления -*(m)a* плохо сочетаются с анализом этого показателя как показателя абсолютного времени. Так, в примере (80) событие, маркированное -*(m)a*, располагается на оси времени после момента речи (но до референциальной точки в будущем), что не дает нам считать -*(m)a* абсолютным прошедшим временем. Пример (85), как и другие примеры с плюсквамперфектом, также говорит против такого анализа, так как в таких примерах нам пришлось бы считать, что в одной клаузе одновременно присутствуют два (самостоятельных) показателя абсолютного времени (на копуле и на причастии), чего, насколько нам известно, не зафиксировано в других языках (см. также [134]). Таким образом, так как нашей целью является моносемический анализ -*(m)a*, мы считаем, что -*(m)a* не может маркировать абсолютное прошедшее время (в том числе и в примерах с отрицанием типа (83)).

У нас остается два варианта того, чем может быть -*(m)a*: это либо относительное время, либо какой-то аспект предшествования (anteriority aspect). Различие между этими двумя категориями не такое очевидное. Единственная известная нам работа, которая предпринимает попытку их различить –– [134]. В качестве главного критерия в этой работе рассматривается способность показателя сочетаться в рамках одной клаузы с относительным или абсолютным временем. Если показатель обладает такой способностью, то он не может являться показателем времени, так как в одной клаузе не может одновременно присутствовать более одного показателя времени, неважно — относительного или абсолютного. Как мы видели в примере с плюсквамперфектом (85), -*(m)a* спокойно сосуществует с матричным глаголом прошедшего времени. Мы, вслед за другими исследователями (среди многих других [135; 136]), придерживаемся интуиции, что аналитический (плюсквам)перфект является моноклаузальной конструкцией. Таким образом, остается единственная альтернатива: -*(m)a* является ничем иным, как упорядочивающим аспектом (ordering aspect), а именно аспектом предшествования.

Соответственно, следующий вопрос, который необходимо решить, чтобы дать точную семантику -*(m)a*, — какими временными точками/интервалами оперирует -*(m)a*? Мы предполагаем, что эти точки не фиксированы в значении -*(m)a*. Как сдвигаемый временной интервал, так и референциальная точка задаются из контекста (синтаксического и прагматического). В пользу этого взгляда говорят примеры типа (86). В этом примере мы видим отрицание плюсквамперфекта и, соответственно, два -*(m)a* причастия. Если бы -*(m)a* строго задавало предшествование времени ситуации референциальной точке, то пришлось бы считать, что в этом примере одно и то же время ситуации дважды располагается до какой-то/каких-то референциальных точек. Такая семантика выглядит избыточной. Если же считать, что только одно из употреблений -*(m)a* семантически наполнено, то придется создать отдельное правило, по которому -*(m)a*- лишалось бы своей семантики и делало бы это только в таком контексте. Нам кажется это излишним усложнением системы, поэтому мы предлагаем более простое решение — вводить оба интервала в виде переменных, которые заполняются в соответствии с синтаксическим и прагматическим контекстом.

1. [137: 14]

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| a | sɨjj | voof's'e | jev | l'iijj-ma | m'iin | saajjm-a | pɒbbx-a |
| INTJ | 3PL.NOM | вовсе | NEG.3PL | быть-PTCP.PST | 1PL.GEN | сеть-DAT.SG | попасть-PTCP.PST |
| 'Они не попали в нашу сеть.' | | | | | | | |

Если в клаузе есть только один показатель -*(m)a*, он обозначает отношение предшествования TT ≺ PT. Отношение между РТ и UT задается маркером абсолютного времени (например, маркером прошедшего времени на копуле в (85)). Согласно [130], время перспективы является подтипом топикальных времен: высказывания, содержащие и ТТ, и РТ, одновременно отсылают к двум топикальным временам.

В таком случае в качестве семантики -*(m)a* мы предлагаем именно предшествование топикального времени времени перспективы, потому что и экспериенциальное, и эвиденциальное прочтения перфекта допускают как перфективный, так и имперфективный аспектуальный ракурс: (87, a) иллюстрирует перфективный, а (87, b) — имперфективный аспектуальный ракурс при эвиденциальном прочтении перфекта. Вслед за [129; 130; 138] и многими другими, мы считаем, что аспектуальный ракурс определяется отношением между ET и TT. Например, в случае перфективного видового ракурса (в одном из его вариантов) время ситуации должно быть целиком включено в топикальное время (ЕТ ⊆ TT). Если бы -*(m)a* обозначало отношение между ЕТ и ТТ (а именно ЕT ≺ ТT), то тогда на местоположение ЕТ одновременно накладывалось бы два ограничения ЕТ ⊆ TT (от аспектуального оператора) и ЕT ≺ ТT (от -*(m)a)*, что вело бы к противоречию (см. более подробное обсуждение семантики перфекта, предложенной Кляйном, в [130; 139]. Поэтому мы считаем, что, когда в клаузе есть только один показатель -*(m)a*, он задает отношение TT ≺ PT.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| munn | puud'd'-e | peer̥'r̥'t'-e | paarrn-e | l'ee-v |
| 1SG.NOM | прийти-PST.1SG | дом-DAT.SG | ребенок-NOM.PL | быть-NPST.3PL |

|  |
| --- |
| ɲuččk-ma |
| прыгать-PTCP.PST |
| 'a. {Когда говорящий пришел домой, он увидел следы от ботинок на диване.} Я пришел домой, [оказалось] дети прыгали на диване (какое-то время до моего прихода).' |
| b. {Когда говорящий пришел домой, он увидел, как пыль сыпется с потолка.} Я пришел домой, дети прыгали наверху (во время моего прихода). |

В предложениях с отрицанием плюсквамперфекта (с двумя -*(m)a*) значение обоих -*(m)a* зависит от структурного контекста: более высокое -*(m)a* устанавливает отношение между “более высокими” (такими, которые вводятся позже в структуре) временными интервалами. Если говорить схематично, структурно более низкое -*(m)a* устанавливает отношение x ≺ y, а структурно более высокое -*(m)a*, в свою очередь, берет правый интервал из этого отношения и устанавливает отношение предшествования этого интервала (интервала y) третьему интервалу z (y ≺ z). Отношение между какими именно временными интервалами устанавливает каждое из причастий (например, ЕT ≺ ТT или TT ≺ PT) — вопрос, который требует дополнительной проверки. Сейчас нам релевантно лишь то, что примеры с отрицанием плюсквамперфекта явно показывают, что временные интервалы (и первый, и второй), между которыми *-(m)a* устанавливает отношение предшествования, строго не заданы и должны варьироваться в соответствии с синтаксическим и прагматическим контекстом. Поэтому, как было сказано выше, мы предлагаем, что *-(m)a* задает отношение предшествования между двумя переменными.

Мы предлагаем, что *-(m)a* является упорядочивающим аспектом (тип <it, it>) и вводит отношение предшествования между двумя переменными (88). Мы моделируем это отношение как пресуппозицию, следуя стандартному анализу прошедшего времени (см., например, [140]).



В качестве левого члена отношения *-(m)a* вводит темпоральное местоимение, так как прошедшие коннегативы типа (83) не имеют экспериенциального прочтения (89)–(90), которое ожидалось бы, если бы *-(m)a* экзистенциально связывало левый временной интервал [134].



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| – | toonn | kues's'-l'aaɲč | poorr-ex' | luemmn-et'ma |
| – | 2SG.NOM | когда-INDEF.REF | есть-PST.2SG | морошка-ACC.PL |
| '– Ты когда-нибудь ел морошку?.' | | | | |



|  |  |  |
| --- | --- | --- |
| – | #emm | poorr-ma |
| – | NEG.1SG | есть-PTCP.PST |
| 'a. ожид.: --- Не ел.' | | |
| b. получ.: #Не ел [в какой-то определенный момент времени]. | | |

В (89)–(90) использование прошедшего коннегатива влечет за собой интерпретацию, согласно которой отвечающий утверждает тот факт, что он не съел морошку в какой-то определенной ситуации, а не то, что он ни разу не пробовал ее за всю свою жизнь.

Аналогичное поведение мы можем наблюдать у -*(m)a*, которое выступает в роли приименного модификатора. Контекст в (91) требует наличие у мальчика опыта точения топора, то есть требуется, чтобы он заточил топор хотя бы один раз в жизни. Однако -*(m)a* причастие не может употребляться в таком контексте. Вместо этого у такого предложения получается одна из возможных интерпретаций, согласно которой мальчик совсем недавно поточил какой-то топор и он сейчас все еще острый. Даже если из этого следует, что у мальчика есть релевантный опыт, такое предложение не может служить ответом на подобный вопрос, так как в нем утверждается лишь факт осуществления события в определенный момент в прошлом.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| #tam'p'e | l'ii | ɒɒkkše | taahkt-ma | paar'r'š'a |
| там | быть.NPST.3SG | топор.ACC.SG | точить-PTCP.PST | мальчик.NOM.SG |
| 'ожид.: {Мне нужно, чтобы кто-нибудь поточил мой топор. Кто когда-либо это делал?} Вот там [стоит] мальчик, точивший [когда-то] топор. {Он умеет это делать.}.' | | | | |
| получ.: Вот там вот [стоит] мальчик, который заточил топор. | | | | |

Интерпретации, которые получаются в (89)–(91), демонстрируют анафорическую природу причастия -*(m)a*, так как в обоих случаях топикальным временем является какой-то определенный (например, контекстно заданный) временной интервал в прошлом, в котором произошло соответствующее событие. Таким образом, анафорические интерпретации (89)–(91) подкрепляют использование в семантике (88) темпорального местоимения в качестве левого члена отношения (ср. анализ английского времени в [141], который базируется на идее [142]).

Тут следует заметить, что анафоричность доступных интерпретаций (89)–(91) нельзя свести к требованию на сохранение лексического результата, из которого следует существование конкретного предшествующего события. Предложение (91), помимо интерпретации с сохранением результата, может иметь такую интерпретацию, при которой результат не сохраняется в момент речи, то есть мальчик заточил топор в какой-то момент в прошлом, например, на прошлой неделе, а в момент речи топор уже является тупым, так как его много использовали. В данном случае утверждается лишь факт осуществления соответствующего события в определенный момент в прошлом. То же самое верно и для предложений с отрицанием: в (89) утверждается, что в определенный момент в прошлом имело место событие, когда говорящий не съел морошку, причем результат этого события может как существовать, так и не существовать в момент речи. Таким образом, мы приходим к выводу, что -*(m)a* действительно вводит теморальное местоимение в качестве левого члена отношения предшествования.

#### 4.1.2.3 Выводы

Анализ семантики причастия прошедшего времени в кильдинском саамском языке позволил прийти к нескольким важным выводам. Во-первых, было доказано, что это причастие не может обозначать время в конвенциональном смысле (ни относительное, ни абсолютное). В литературе существует анализ причастия прошедшего времени как нефинитного прошедшего времени [143]. Мы опускаем дискуссию о разнице между нашим анализом и анализом М. Юльен в целях экономии места. Во-вторых, разработан семантический анализ -*(m)a*, успешно покрывающий все употребления этого показателя — маркера упорядочивающего аспекта, который вводит отношение предшествования. Детали композиции мы также опускаем в целях экономии места. При композициональном анализе семантики предложений типа (82) проблем возникнуть не должно: причастие обозначает предшествование точке, заданной матричным временем. Для интерпретации отрицания (83) нам придется обратиться к анализам безвременных языков (см. обзор в [144]), так как в рамках нашего анализа в клаузах с отрицанием нет выраженных маркеров финитного времени. Одной из конкретных альтернатив можно считать анализ языка лиллуэт из статьи [145], где предлагается ввести нулевое настоящее/непрошедшее время. В случае перфекта (84) мы предполагаем, что перфектный оператор (выраженный копулой) вводит результирующее состояние в духе [131], которое заполняется в соответствии с определенным прочтением перфекта. На настоящий момент мы избегаем конкретных предположений о том, как складывается значение плюсквамперфекта (85), так как мы пока достоверно не знаем детали семантики этой формы.

## 4.2 Синтаксические передвижения в полипредикации

### 4.2.1 Введение

Одной из самых базовых операций в формальной лингвистике (помимо правил, ответственных за построение синтаксических структур) являются кросс-клаузальные нелокальные взаимодействия, также известные как синтаксическое передвижение. В отчётном периоде его свойства изучались на примере субъектов вложенных клауз в мишарском диалекте татарского языка, маркированных именительным и винительным падежом.

### 4.2.2 Маркирование субъекта вложенной клаузы при скрэмблинге в мишарском диалекте татарского языка

В тюркских языках возможность дистантного скрэмблинга часто варьируется в зависимости от падежа субъекта вложенной клаузы. Данное исследование было посвящено изучению взаимодействия дистантного скрэмблинга и дифференцированного маркирования субъекта в мишарском диалекте татарского языка.

Целью работы является проверка на материале мишарского татарского гипотез, выдвинутых в работе [146] на материале балкарского языка. Авторы утверждают, что возможность скрэмблинга субъекта и других элементов вложенной клаузы (например, прямого объекта) зависит от того, маркирован ли субъект номинативом или аккузативом, и связывают это с различными структурными позициями, которые занимают субъекты в зависимости от их падежа. Данные были собраны в селе Большое Рыбушкино во время экспедиции в мае 2025 года.

#### 4.2.2.1 Предыдущие исследования: данные балкарского языка

Авторы работы [146] исследуют связь между падежом субъекта и возможностью скрэмблинга в балкарском языке. В работе рассматриваются номинализации, субъекты которых могут быть номинативными, аккузативными и генитивными (последние для нас нерелевантны).



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ustaz-ø | fatima-nɨ | sabiji-ø | alma-nɨ | aša-ʁan-ɨ-n | ešit-ti |
| учитель-NOM | Фатима-GEN | ребёнок-NOM | яблоко-ACC | есть-NFUT-AGR-ACC | слышать-PST |
| 'Учитель подумал, что ребёнок Фатимы съел яблоко' | | | | | |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ustaz-ø | fatima-nɨ | sabiji-n | alma-nɨ | aša-ʁan-ɨ-n | ešit-ti |
| учитель-NOM | Фатима-GEN | ребёнок-ACC | яблоко-ACC | есть-NFUT-AGR-ACC | слышать-PST |
| 'Учитель подумал, что ребёнок Фатимы съел яблоко' | | | | | |

Авторы демонстрируют следующую закономерность: cкрэмблинг субъекта возможен для аккузативного субъекта, но невозможен для номинативного. Скрэмблинг прямого объекта возможен из клауз с номинативными, но не с аккузативными субъектами. Ниже изложены допущения, принимаемые в анализе авторов. Во-первых, структурные позиции субъектов: Номинативный субъект находится в позиции [Spec, TP] внутри вложенной клаузы. Аккузативный субъект, передвигается в [Spec, CP], минуя [Spec, TP]. Во-вторых, Phase Impenetrability Condition [147]: Чтобы составляющая могла покинуть вложенную клаузу, она должна сначала переместиться на её границу (в [Spec, CP]). Это делает её доступной для дальнейших операций в матричной клаузе.

Номинативный субъект не может переместиться в [Spec, CP] из-за нарушения условия анти-локальности [148; 149], которое запрещает слишком короткие передвижения (например, из [Spec, TP] в [Spec, CP]). Следовательно, его последующее передвижение в матричную клаузу невозможно.

Прямой объект при аккузативном субъекте не может переместиться в самый высокий [Spec, CP], потому что эта позиция уже занята субъектом. Согласно принципу “tucking in” [150], при перемещении нескольких элементов в спецификаторы одной вершины первой передвигается самый высокий из них, а все последующие занимают более низкие спецификаторы. Поскольку аккузативный субъект занимает [Spec, CP] первым, прямой объект не может передвинуться в матричную клаузу без предварительного передвижения субъекта.

#### 4.2.2.2 Дифференцированное маркирование субъекта в мишарском татарском

Прежде чем проверять предсказания анализа [146], рассмотрим конструкции с дифференцированным маркированием субъекта в мишарском татарском. Известно, что в финитных вложенных клаузах с комплементайзером *diep* и с матричным глаголом глаголом *di*- ‘сказать’ субъект может быть как номинативным, так и аккузативным (см. [151]).



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| alsu | marat(-nɤ) | kit-te | diep | ujla-dɤ |
| Алсу | Марат(-ACC) | уйти-PST | COMP | думать-PST |
| 'Алсу подумала, что Марат ушёл' | | | | |

Есть свидетельства, что аккузативные субъекты занимают иную, более высокую позицию в структуре, чем номинативные. В частности, ИГ в матричной клаузе может служить антецедентом для рефлексива в позиции субъекта вложенной клаузы только если он аккузативный.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| marat | üz-e\*(-n) | jaxšɤ | ešlä-de | diep | atj-tɤ |
| Марат | сам-3\*(-ACC) | хорошо | работать-PST | COMP | сказать-PST |
| 'Марат сказал, что он хорошо поработал' | | | | | |

Эти и другие данные (см. [152]) позволяют сделать вывод, что в мишарском татарском, как и в балкарском, аккузативные субъекты занимают позицию [Spec, CP], в то время как номинативные располагаются в [Spec, TP].

#### 4.2.2.3 Дистантный скрэмблинг в мишарском татарском

Скрэмблинг субъекта в матричную клаузу возможен только для аккузативных субъектов.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| marat\*(-nɤ) | alsu | kit-te | diep | ujla-dɤ |
| Марат\*(-ACC) | Алсу | уйти-PST | COMP | думать-PST |
| 'Алсу подумала, что Марат ушёл' | | | | |

Это ожидаемо, если номинативный субъект не может занять [Spec, CP] вследствие анти-локальности, как утверждается в [146]. При скрэмблинге других элементов вложенной клаузы наблюдается ассимметрия между прямыми объектами и PP-адъюнктами. Скрэмблинг прямых объектов возможен только если субъект вложенной клаузы номинативный, в то время как скрэмблинг PP-адъюнктов доступен независимо от падежа субъекта.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| mašina-nɤ | alsu | marat(\*-nɤ) | sat-tɤ | diep | ujl-ɤj |
| машина-ACC | Алсу | Марат(\*-ACC) | продать-PST | COMP | думать-IPFV |
| 'Алсу думает, что Марат продал машину' | | | | | |



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| uram-nan | alsu | marat(-nɤ) | akca | tap-tɤ | diep | ujlа-dɤ |
| улица-ABL | Алсу | Марат(-ACC) | деньги | найти-PST | COMP | думать-PST |
| 'Алсу подумала, что Марат нашёл деньги на улице' | | | | | | |

Дополнительно была проверена одна гипотеза, которая могла бы потенциально объяснить, почему дистантный скрэмблинг возможен с аккузативными субъектами. До сих пор мы исходили из допущения, что все предложения типа (96) с аккузативными аргументами имеют структуру с подъёмом, где ИГ порождается в позиции субъекта, после чего передвигается в [Spec, CP], получая аккузатив. Однако в мишарском татарском также засвидетельствованы предложения с пролепсисом — конструкцией, где в матричной клаузе порождается аккузативная ИГ, коиндексированная с местоимением во вложенной клаузе (оно может быть нулевым).



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| alsu | marat-nɤ | (ul) | kit-te | diep | ujla-dɤ |
| Алсу | Марат-ACC | он | уйти-PST | COMP | думать-PST |
| 'Алсу подумала, что Марат ушёл' | | | | | |
| Букв.: Алсу подумала про Марата, что он ушёл | | | | | |

Для проверки теории [146] необходимо рассматривать именно структуры с подъёмом. Для различения двух структур мы будем использовать местоимения отрицательной полярности в позиции субъекта вложенной клаузы, которые будут лицензироваться вложенным отрицанием. В мишарском татарском такие местоимения в матричной клаузе не могут лицензироваться отрицанием вложенной клаузы [151]. Таким образом, этот тест позволяет рассматривать именно структуры, где аккузативный аргумент порождается как субъект вложенной клаузы. Показательность нашей диагностики подтверждается примерами типа (97), где однозначно задействована структура с пролепсисом. В таких предложениях местоимение отрицательной полярности в позиции аргумента матричной клаузы не может лицензироваться вложенным отрицанием.



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| \*alsu | kem-ne | dä | ul | kil-mä-de | diep | ujl-ɤj |
| Алсу | кто-ACC | EMPH | он | прийти-NEG-PST | COMP | думать-IPFV |
| 'Алсу думает, что никто не пришёл' | | | | | | |

При этом отрицание матричной клаузы ожидаемым образом может лицензировать местоимение отрицательной полярности в позиции пролептического аргумента.



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| alsu | kem-ne | dä | (ul) | kil-de | diep | ujla-m-ɤj |
| Алсу | кто-ACC | EMPH | он | прийти-PST | COMP | думать-NEG-IPFV |
| 'Алсу думает, что никто не пришёл' | | | | | | |
| Букв.: Алсу ни про кого не думает, что (он) пришёл | | | | | | |

Итак, местоимения отрицательной полярности точно являются субъектами вложенной клаузы, когда лицензируются вложенным отрицанием. Тогда, следуя [146], можно ожидать, что в таких предложениях дистантный скрэмблинг несубъектных составляющих будет невозможен, если субъект аккузативный. Однако ср. (99), в котором дистантный скрэмблинг возможен.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| däres-kä | ukɤtučɤ | kem-ne | dä | kil-mä-de | di-de |
| урок-DAT | учитель | кто-ACC | EMPH | прийти-NEG-PST | сказать-PST |
| 'Учитель сказал, что никто не пришёл на урок' | | | | | |

### 4.2.3 Заключение

В области дистантных передвижений в полипредикации было проведено детальное исследование дистантного скрэмблинга в мишарском диалекте татарского языка в контексте дифференцированного маркирования субъекта. Мы показали, что дистантный скрэмблинг некоторых элементов возможен вне зависимости от падежа субъекта вложенной клаузы, что не предсказывается существующим анализом, предложенным для балкарских данных. Было также показано, что грамматичность примеров нельзя объяснить, предположив для них альтернативную структуру, где аккузативный “субъект” на самом деле порождается в матричной клаузе. Таким образом, новый материал мишарского татарского служит эмпирической базы для уточнения существующих подходов к анализу.

## 4.3 Грамматические свойства логических операторов

Отрицание и его взаимодействие с различными классами выражений — одна из центральных тем лингвистической теории, лежащая в основе целого ряда формальных моделей. В отчётном периоде в рамках проекта исследовалось поведение местоимений на -*либо* в русском языке, дистантное лицензирование отрицательных местоимений в абазинском языке и неканонические случаи лицензирования отрицательных местоимений в русском языке. Данные были собраны методом элицитации (в том числе в полевых условиях в рамках Абазинской лингвистической экспедиции) и в рамках формального эксперимента на оценку приемлемости.

### 4.3.1 Местоимения на -*либо* как нереферентные неопределённые местоимения

Русские неопределённые местоимения на -либо (например, кто-либо, что-либо) принято относить к отрицательно поляризованным единицам [153–156]. Это означает, что они обычно употребляются в контекстах нисходящего следования, но не в утвердительных перфективных контекстах. Однако наблюдается парадокс: хотя, согласно [157], единицы отрицательной полярности должны лицензироваться под отрицанием (самым сильным из таких контекстов) местоимения на -либо в этой позиции недопустимы (105). Эта проблема известна как «загадка бублика» («the Bagel problem», [155]).

1. *Если вы кого-либо встретите, позвоните мне*. [155: пример (10b)]
2. \**Я видел что-либо*. [155: пример (8)]
3. \**Он кого-либо не встретил*. [155: пример (11)]

Существующее объяснение [155] связывает невозможность употребления -*либо* под отрицанием с конкуренцией с отрицательными местоимениями на *ни*- (например, *никто*). Однако у этой гипотезы есть недостатки. Во-первых, в некоторых случаях местоимения на -*либо* и *ни*- в некоторых контекстах взаимозаменяемы, что противоречит идее обязательной конкуренции. Это происходит при атрибутивном использовании (например, *какой-либо*/*никакой*) или при наличии послеименного модификатора.

1. *Иван не видел там каких-либо студентов*. [153: 43]
2. *Между тем, итальянки не смогли назвать кого-либо, кто лично был свидетелем феномена*. (НКРЯ/Центральные СМИ)

Мы считаем, что «загадка бублика» может быть разрешена, если отказаться от трактовки местоимений на *-либо* как отрицательно поляризованных единиц. Так, примеры (108)–(109) показывает, что местоимения на -либо могут лицензироваться в модальных и императивных контекстах, что противоречит данным [154].

1. *Сейчас можно покупать что-либо внутри приложений. Сами приложения скачивать нельзя*. (НКРЯ/Центральные СМИ)
2. *Если есть возможность, пошлите кого-либо в ближайший населенный пункт или лесничество за помощью*. (НКРЯ/Региональные СМИ)

Предлагаемая гипотеза состоит в том, что ключевым свойством местоимений на -*либо* природе является нереферентность. Они лицензируются не самим по себе контекстом нисходящего следования, а наличием квантифицирующего оператора, который может быть явным или скрытым. В слабых контекстах нисходящего следования, как в примере (103), мы предполагаем наличие нулевого модального оператора вслед за [158]. Употребление -*либо* в примерах (106)–(107) также объясняется через модальность. Наличие послеименного модификатора создаёт условия для скрытого модального оператора [158; 159], который и лицензирует местоимение на -*либо*. Аналогичный механизм работает для английского *any* в конструкциях типа «John saw any student that was around». Обращает на себя внимание возможность замены местоимения серии -*либо* на местоимение серии -*то*, которое является положительно поляризованной единицей, то есть не может находиться в сфере действия отрицания [160].

1. *Между тем, итальянки не смогли назвать кого-то, кто лично был свидетелем феномена.*

Таким образом, местоимения на -*либо* скорее следует относить к классу нереферентных неопределённых местоимений, а не к классу полноценных единиц отрицательной полярности. Это объясняет их грамматичность в модальных контекстах и условных предложениях и неграмматичность под отрицанием и в утвердительных предложениях.

### 4.3.2 Дистантное лицензирование отрицательных местоимений в ашхарском диалекте абазинского языка

Под дистантным лицензированием отрицательных местоимений понимаются случаи, в которых отрицательная единица, располагающаяся в зависимой клаузе, лицензируется маркером отрицания (как правило, глагольного), находящимся в главной клаузе. Абазинский материал представляет особый интерес ввиду того, что система маркирования зависимых клауз в этом языке (как и вообще в абхазо-адыгских языках) существенно отличается от языков среднеевропейского стандарта. Абазинский язык не имеет типов зависимых клауз, которые можно было бы назвать в полном смысле слова нефинитными (такими как, например, клаузы с инфинитивом в европейских языках). Все собственно глагольные формы в зависимых клаузах не имеют полного набора признаков нефинитности — в частности, все они способны иметь собственный субъект. Особняком стоят зависимые клаузы с масдаром — однако масдар не является в полной мере глагольной формой (если точнее, разные типы масдаров располагаются на разных точках шкалы «имя — глагол»). В ходе анализа материала мы стремились выявить основные факторы, ответственные за возможность, невозможность или сомнительность предложений с дистантным отрицательным согласованием. Как выяснилось, релевантны следующие факторы:

* cемантико-синтаксический тип предиката;
* тип зависимой клаузы;
* расположение зависимой клаузы относительно главной (препозиция/постпозиция).

Можно было бы ожидать, что основным фактором возможности / невозможности отрицательного согласования будет именно формальный тип (способ маркирования) зависимой клаузы. Однако в действительности это не так, причём зависимость от формального типа не соблюдается, так сказать, в обе стороны. С одной стороны, зависимые клаузы, содержащие одну и ту же форму глагола, однако при разных матричных предикатах, часто в разной мере допускают дистантное отрицательное согласование. С другой стороны, нередко оказывается, что при одном и том же матричном предикате способ оформления зависимой клаузы может быть разным — однако степень допустимости дистантного отрицательного согласования остаётся прежним.

Фактивные глаголы — как когнитивной группы, так и группы эмоций — чаще всего не допускают дистантного отрицательного согласования (111). Напротив, глаголы, предполагающие ирреальность вложенного события или неполную степень его вероятности, сочетаются с дистантным согласованием лучше (112):



|  |  |  |
| --- | --- | --- |
| ??Сарá | й-гьы-сы-з-ды́ра-м | аджв-гьи |
| я | 3SG.N.ABS-NEG.EMP-1SG.ERG-POT-знать-NE | кто\_то-ADD |

|  |
| --- |
| ды-с-з-áй-уш-ны |
| 3SG.ABS-1SG.IO-BEN-прийти-FUT-ADV |
| 'Я не знаю, что никто придёт' |



|  |  |  |  |
| --- | --- | --- | --- |
| Сарá | зджьара-гьи́ | с-ца-рны́с | гьы-с-тахъвы́-м |
| я | куда-ADD | 1SG.ABS-идти-PURP | NEG.EMP-1SG.ERG-хотеть-NEG |
| 'Я не хочу никуда идти (Белла)' | | | |

По нашим данным, матричные предикаты упорядочиваются по способности допускать дистантное отрицательное согласование в соответствии со следующей нестрогой шкалой:

1. субъективная модальность, фазовые > ?каузативные > ?эпистемические мнения > речи > *фактивные знания,* фактивные эмоций

В очень ограниченном классе примеров стратегия подчинения всё-таки является релевантной. В частности, обнаруживается различие между конструкцией с масдаром (114) (не склонной пропускать через себя отрицательное согласование) и конструкцией с целевой формой на -*рныс* (115) (обычно пропускающей отрицательное согласование). Это различие соответствует другому: масдар в большей мере, чем любая другая форма в подчинённых клаузах, склонен требовать согласования на матричном глаголе. Тем самым, несмотря на свои промежуточные свойства, масдар всё-таки находится ближе к актантной именной группе, чем другие формы — в этой связи он, как большинство номинализаций, с трудом пропускает отрицательное согласование:



|  |  |  |  |
| --- | --- | --- | --- |
| \*Сара | й-гьы-с-тахъы-м | зджьар-гьи́ | у-ца-рá |
| я | 3SG.N.ABS-NEG.EMP-1SG.IO-хотеть-NEG | куда-ADD | 2SG.ABS-идти-MSD |
| 'Я не хочу, чтобы ты никуда шел' | | | |



|  |  |  |  |
| --- | --- | --- | --- |
| Сара | й-гьы-с-тахъы́-м | зджьар-гьи́ | у-ца-рны́с |
| я | SG.N.ABS-NEG.EMP-1SG.IO-хотеть-NEG | куда-ADD | 2SG.ABS-идти-PURP |
| 'Я не хочу, чтобы ты никуда шел' | | | |

### 4.3.3 «Обратное» отрицательное согласование в русском

Считается, что основным структурным ограничением на лицензирование отрицательных местоимений является наличие локального сентенциального отрицания в пределах клаузы. Значимым различием между приемлемым предложением Никто \**(не) смотрит телевизор* и неприемлемым предложением \**Лина не думает, что никто смотрит телевизор* является именно наличие локального сентенциального отрицания. Для данного правила зафиксирован ряд исключений, природа которых до сих пор не до конца ясна: при некоторых предикатах субъектного контроля (например, *стараться*) отрицательное местоимение в позиции контролёра способно лицензироваться отрицанием, находящимся в инфинитивной клаузе: *Никто старался об этом не вспоминать* [161]. Похожее наблюдение существует и для обширного класса глаголов объектного контроля: отрицательное местоимение-объектный контролёр в матричной клаузе также может лицензироваться отрицанием, находящимся в инфинитивной клаузе. Целью настоящего исследования являлась эмпирическая проверка данного наблюдения, в наиболее эксплицитной форме сформулированного в работе [162], и основной гипотезы из этой же работы: лицензирование матричного отрицательного местоимения зависимым отрицанием — «обратное» отрицательное согласование — представляет собой единый феномен, одинаково доступный для должным образом очерченных подклассов в составе предикатов субъектного и объектного контроля и недоступный для предикатов подъёма и реструктурирования. Стоит заметить, что, хотя само обобщение в цитированных выше работах сформулировано на основе данных НКРЯ, лежащие в его основе примеры оцениваются носителями в рамках неформальных опросов достаточно низко. Именно эта разница между объективно засвидетельствованным в корпусе и суждениями отдельных носителей делает экспериментальную проверку обобщения особенно актуальной. Для проверки реальности «обратного» отрицательного согласования разработан эксперимент на оценку приемлемости в соответствии со следующими принципами. Во-первых, приемлемость «обратного» отрицательного согласования необходимо сравнить с приемлемостью канонического отрицательного согласования; во-вторых, это необходимо сделать для субъектного и объектного контроля. Таким образом, мы имеем две независимые переменные с двумя значениями каждой: матричная либо вложенная позиция отрицания (MatNeg/EmbNeg) и контекст субъектного либо объектного контроля (SC/OC). Оптимальным дизайном в данных условиях является метод латинского квадрата . Зависимой переменной является оценка приемлемости по шкале Ликерта. Для исследования было выбрано четыре предиката субъектного контроля (*планировать*, *предпочитать*, *собираться*, *хотеть*) и четыре предиката объектного контроля (*просить*, *убеждать*, *умолять*, *учить*), употреблённых в прошедшем времени несовершенного вида в сочетании с четырьмя инфинитивными оборотами (*возвращаться домой ночью*, *работать по воскресеньям*, *разговаривать с мошенниками*, *засиживаться в гостях допоздна*). Таким образом, 64 стимульных предложения разбиты на четыре экспериментальных листа по 16 стимульных предложений в каждом. В качестве филлеров использовались 16 стимульных предложений из другого исследования, никак не связанного с отрицательным согласованием без лексических пересечений, а также четыре заведомо неприемлемых инвариантных предложения с нарушениями падежного управления. Каждое сочетание условий подлежит оценке каждым участником четыре раза, по одному разу в пределах экспериментального блока. Все стимулы приводились в диалогической форме. Пример экспериментального блока приведён ниже.

1. Что у вас за спор? — Иван просил никого не возвращаться домой ночью. (OC; EmbNeg)
2. Что у вас за спор? — Никто планировал не возвращаться домой ночью. (SC; EmbNeg)
3. Что у вас за спор? — Иван не просил никого возвращаться домой ночью. (OC; MatNeg)
4. Что у вас за спор? — Никто не планировал возвращаться домой ночью. (SC; MatNeg)
5. Маша полностью перестала общаться с Иваном. — Да, но я не поражён этого. (Filler; Subcat)

Эксперимент реализован с помощью инструментов PCIbex [163] на платформе Яндекс.Задания; после отсева за слишком высокую оценку предложений с нарушениями падежного управления мы приводим данные 181 участника (средний возраст 37,08, SD = 11.15). Согласно гипотезе, если за предполагаемую приемлемость «обратного» отрицательного согласования в контексте предикатов субъектного и объектного контроля, ожидается отсутствие значимого эффекта при манипуляции данной переменной — тип контроля во взаимодействии с позицией отрицания не должен значимым образом влиять на оценки приемлемости. Как показывает график, данное предсказание не выполняется: объектный контроль в сочетании со вложенным отрицанием оценивается выше объектного контроля в сочетании с матричным отрицанием и субъектного контроля в сочетании со вложенным отрицанием. Субъектный контроль с матричным отрицанием при этом оценивается выше всего, в том числе выше объектного контроля со вложенным отрицанием.

|  |
| --- |
| Рисунок 4.1 – Распределение оценок по условиям |

Эти выводы подтверждаются и линейной регрессионной моделью со смешанными эффектами, в которой в качестве постоянных эффектов используются две переменные — позиция отрицания и тип контроля, а также их взаимодействие; идентификатор участника и стимульное предложение представляют собой случайные эффекты (z\_score ~ neg\_pos \* control + (1 | person\_id) + (1 | item\_id)). Свободный коэффициент (OC, EmbNeg) равен -0,08 (SE = 0,068). Фактор позиции отрицания является значимым (β = -0,287, SE = 0,094, *p* < 0,001). Тип контроля также является значимым (β = -0,50158, SE = 0,094, *p* < 0,001). Наконец, взаимодействие типа контроля с позицией отрицания также значимо (β = 0,932, SE = 0,133, *p* < 0,001).

Результаты эмпирического экспериментального исследования не подтверждают гипотезу о единстве механизма, ответственного за «обратное» отрицательное согласование при субъектном и объектном контроле, и могут быть истолкованы следующим образом. Высокая приемлемость канонического отрицательного согласования при субъектном контроле (*Никто не планировал возвращаться домой ночью*.) ожидаема и в дополнительном объяснении не нуждается. Значимо более низкая приемлемость канонического отрицательного согласования в контексте объектного контроля (*Иван не просил никого возвращаться домой ночью*.) может быть объяснена тем, что отрицательные местоимения в русском языке предпочитают препозицию по отношению к лицензирующему отрицанию — минимально отличное предложение *Иван никого не просил возвращаться домой ночью* звучит намного естественнее. При «обратном» отрицательном согласовании отрицательное местоимение, во-первых, предшествует отрицанию-лицензору и, во-вторых, объектный контроль может быть переосмыслен в терминах исключительного падежного маркирования (Exceptional Case Marking), а отрицательное местоимение не занимает позиции контролёра в матричной клаузе, располагаясь внутри инфинитивного оборота [164]. Низкую приемлемость «обратного» отрицательного согласования при субъектном контроле, в свою очередь, можно попытаться объяснить, переосмыслив субъектный контроль как реструктурирование, которое по определению не допускает «обратного» отрицательного согласования.

## 4.4 Анафора и связывание

В рамках работы над данной темой анализировалось поведение трех типов местоимений. Во-первых, на основе данных русского языка рассматривались нестандартные случаи контроля взаимных (реципрокальных) местоимений — а именно, конфигурация, где контролёр местоимения находится в предложной группе. Во-вторых, на материале казымского диалекта северохантыйского языка описывался близкий тип местоимений — прономиналы — а именно, возможности их использования для анафоры к посессорам. Наконец, в-третьих, проводилось исследование синтаксического поведения указательных местоимений в казымском диалекте севернохантыского языка. Не в последнюю очередь исследование было важно в том смысле, что позволило охватить разные исследовательские проблемы, связанные с местоимениями. Если для взаимных местоимений, как и для других анафоров, существенна в первую очередь проблема контроля и ограничений на него (статуса контролёра, синтаксической дистанции между контролёром и местоимением и других признаков), то для указательных местоимений на первое место выходят другие свойства. Демонстративы, рассмотренные в рамках работы, не замещают именную группу, а относятся к модификаторам — поэтому существенным оказалось выявить их соотношение с типами вершин и с другими типами модификаторов.

### 4.4.1 Связывание реципрокальных местоимений из предложных групп на материале русского языка

В данной части работы мы ставили вопрос о нестандартных (предложных) контролёрах взаимных местоимений. Вопрос о том, возможно ли связывание реципрокальных местоимений из предложных групп, остается дискуссионным: несмотря на мнение в литературе, согласно которому предложная группа не может выступать антецедентом (см. [165: 46]), эмпирической проверки эта гипотеза до сих пор не проходила. Семантически нет ограничения, которое не позволяет получение референциальной интерпретации анафором у предложной группы, ср. предложения (121) и (122), где при одном и том же глаголе по-разному выражен первый аргумент с одинаковой тематической ролью адресата.

1. ОК Мы решили устроить необычное интервью — и спросили ихi [друг о друге]i. [<https://www.ap22.ru/paper/Artisty-teatra-dramy-Eduard-Timoshenko-i-Nikolay-Miroshnichenko-interv-yu-naoborot.html>]
2. ? Полицейский спрашивал [у соседей]i [друг о друге]i.

Считается, что русское реципрокальное местоимение друг друга является строгим анафором, требующим с-командующего антецедента в непрозрачной области, а фактором непрозрачности является любое подлежащее [166].

1. Родителиi думали, что детиj любят друг друга\*i/j

В работе [165: 45] выдвигается следующую гипотезу о правилах поведения местоимения: Антецедент местоимения друг друга должен занимать более высокую позицию в иерархии, чем само местоимение:

1. подлежащее > прямое дополнение > косвенное дополнение > ИГ с предлогом

Результаты исследования [167; 168] показали, что указанная иерархия не всегда соответствует реальному употреблению: были зафиксированы редкие случаи с предложным антецедентом, а также контексты без выраженного антецедента. Некоторые исследователи также считают, что связывание из предложной группы в русском языке не является невозможным [169–171]. Была поставлена цель эмпирически проверить, в каких условиях действительно возможно связывание реципрокальных местоимений из предложных групп.

1. OK Вместо того чтобы отобрать камни чудес, он отбирает у нихi [друг друга]i перемещая в прошлое.

Наше исследование, ставившее целью проверить возможность предложных контролёров, было выполнено в рамках экспериментального синтаксиса. Данный подход уместен для анализа данной проблематики, поскольку, как показано ранее, зачастую в той или иной мере допустимы самые разные конфигурации — в том числе с контролёром в предложной группе и с местоимением в предложной группе. Тем самым, едва ли возможно говорить о бинарной оппозиции приемлемых и неприемлемых местоименных употреблений. В пилотном эксперименте было выявлено влияние следующих факторов: аргументный статус антецедента и анафора, вложенность анафора в именную группу, а также порядок слов. Эксперименты были составлены в соответствии с принципами, сформулированными в [172]. Данные собраны с помощью платформы PCIbex [163], всего в наших экспериментах приняло участие 162 респондента.

#### 4.4.1.1 Эксперимент 1 Факторы: антецедент(+-адъюнкт), анафор (+-адъюнкт), анафор (+-вложенность)”

Всего эксперимент прошло 80 респондентов (средний возраст — 33.7, соотношение лингвистов/не лингвистов — 15:85) приняли участие в эксперименте, причем мы никого не исключили из наших данных для анализа, так как при подсчете среднего у неграмматичных филлеров и среднего у грамматичных филлеров оказалось, что ни у одного участника первое значение не превосходит второе по 7-балльной шкале и по нормированной шкале. Для анализа данных мы использовали среду программирования R [173]. Примеры условий: +адъюнкт (антецедент), +адъюнкт (анафор), +вложенность в именную группу (анафор); +адъюнкт (антецедент), +адъюнкт (анафор), -вложенность в именную группу (анафор); +адъюнкт (антецедент), -адъюнкт (анафор), В целом, все предложения с предложным антецедентом оцениваются как неприемлемые (по остальным условиям стандартное отклонение варьируется от 1 до 2, значит разброс оценок умеренный). Плохие филлеры оценены хуже, чем все экспериментальные стимулы, а хорошие филлеры сильно лучше: оценка близится к 7 — чего мы и ожидаем. Была использована линейная смешанная модель. Сравнение результатов моделей m0 (учитывает только случайные эффекты) и m1 (учитывает синтаксические условия) показывает, что условия значительно улучшают модель (χ² = 44.61, df = 9, p < 0.001). Таким образом, можно заключить, что тип синтаксического условия оказывает статистически значимое влияние на приемлемость. Далее была использована метрика Estimated Marginal Means (EMMeans) pairwise сравнение с коррекцией Тьюки, чтобы установить какие именно условия статистически значимо отличаются друг от друга. Для каждой пары мы вычислили разницу средних оценок, CI (Confidence interval), p-value. Можно сделать вывод, что фактор вложенности анафора и аргументного / неаргументного статуса по отдельности не влияет на оценку. Влияет только комбинация факторов: аргументность антецедента, и анафора, а также базовый порядок слов.

#### 4.4.1.2 Эксперимент 2

Был также проведен эксперимент, в котором варьируемыми параметрами выступали аргументность антецедента, вложенность анафора в именную группу и грамматическая функция анафора (прямое дополнение и непрямое дополнение в дательном падеже), т. е. была добавлена синтаксическая позиция анафора. В результате эксперимента не было выявлено статистической значимости ни для одного из параметров и их комбинаций.

#### 4.4.1.3 Заключение

Согласно нашим данным, связывание из предложной группы в русском языке с предикатами с двумя предложными аргументами/адъюнктами скорее невозможно. Предложный антецедент потенциально возможен, когда и анафор, и антецедент являются аргументами в предложной группе, и невозможен при наличии альтернативы (например, соответствующей иерархии в (124)). Результаты экспериментов указывают на необходимость расширения списка факторов с учетом тематической роли, аргументной структуры, типа предлога и других.

### 4.4.2 Синтаксическое поведение демонстративов в казымском диалекте севернохантыйского языка

Вторая часть работы над темой «Анафора и связывание» была посвящена данным севернохантыйского языка. В центре внимания находились демонстративы (указательные местоимения): был предложен вариант анализа синтаксического поведения демонстративов в казымском диалекте севернохантыйского языка (обско-угорские < уральские). Взаимодействие демонстративов с артиклями активно обсуждается в литературе, однако безартиклевые языки, в том числе севернохантыйский, также предоставляют интересный материал для анализа. Севернохантыйские демонстративы — высокие модификаторы в именном комплексе, которые нетривиально взаимодействуют с некоторыми другими высокими модификаторами [174]. Для исследования оказалось существенным, что, как показывают стандартные тесты на структуру составляющих, согласуемый посессор может образовывать составляющую с ИГ [175]. Кроме того, ИГ, модифицированные согласуемым посессором, могут могут быть комплементами послелогов, а послелоги не допускают выноса из комплемента (Буров, л.с.). С учетом крайней ограниченности возможностей именного эллипсиса в севернохантыйском, отдельного внимания заслуживают данные о поведении прономинальных форм демонстративов *tăm(i)* ‘это’, *tum(i)* ‘то’ и *śit*, которые образуются присоединением сегмента -*i*/-*t* к демонстративной основе. В отличие от приименных форм, они могут присоединять морфологические показатели имен — падеж, число и посессивную морфологию:

* приименные демонстративы



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| \*min | tăm-ən | imi-jən | χuti | tɵ-λ-ij-mən |
| мы.DU | этот-LOC | женщина-LOC | PTCL | приносить-NPST-PASS-1DU |
| 'Ожид.: Нас ведет эта женщина.' | | | | |



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| \*wɵn | χujat-ət | tăm-ət | junt-ət | wan-əλtə-λ-ət |
| большой | человек-PL | этот-ət | игра-PL | видеть-CAUS-NPST-3PL |
| 'Ожид.: Взрослые показывают эти игры.' | | | | |



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| \*tăm-ɛm | χir-ɛm | tɛλijewa | piti | woj, | wʉrti | woj | suχ-ən |
| этот-POSS.1SG | мешок-POSS.1SG | полный | черный | животное | красный | животное | кожа-LOC |

|  |
| --- |
| pun-s-a.t |
| класть-PST-PASS |
| 'Ожид.: Этот мой мешок наполнен шкурами черных и красных зверей.' |

* прономинальные демонстративы



|  |  |  |
| --- | --- | --- |
| tămi-λ-ən | aλ | jurɛm-a-λən. |
| этот.one-PL-POSS.2NSG | PROH | забыть-IMP-NSG&gt;(N)SG |
| 'Не забудь эти твои (вещи).' | | |



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| — | năŋ | măta | aj | ik-en-a | maw | mă-s-ən? |
| — | ты | который | маленький | мужчина-POSS.2SG-DAT | конфета | дать-PST-2SG |
| '— Какому мальчику ты дала конфету?' | | | | | | |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| — | tăm-en-a | / | tum-en-a | / | śit-en-a. |
| — | этот.one-POSS.2SG-DAT | / | тот.one-POSS.2SG-DAT | / | DEM.one-POSS.2SG-DAT |
| '— Этому/тому.' | | | | | |

Прономинальные демонстративы допускают модификацию только согласованными посессорами, т. к. только они находятся в структуре выше демонстратива:



|  |  |  |  |
| --- | --- | --- | --- |
| Waśaj-en | tăm-əλ | λʉw-eλ | tɵ-λ-ɛm. |
| Вася-POSS.2SG | этот-POSS.3SG | он(а)-DAT | нести-NPST-1SG&gt;SG |
| 'Я это Васино отнесу ему.' | | | |

Несовместимость кванторов и демонстративов в севернохантыйском языке, по-видимому, объясняется семантически: если демонстративные ИГ и демонстративы имеют типы e и <<e,t>,<<e,t>,e>> соответственно [176], а кванторы имеют стандартный семантический тип <<e,t>,<e,t>,t>> и принимают комплемент типа <e,t>, попытка их комбинации приводит к несоответствию типов. Ни квантованная ИГ (тип <<e, t>,t>) не может быть комплементом демонстратива (требуется комплемент типа <e, t>), ни демонстративная ИГ (тип e) не может быть комплементом квантора (требуется комплемент типа <e,t>).

При этом приименные демонстративы являются вершинами и находятся в D0. Аргументом против этой гипотезы является линейная позиция демонстративов: в севернохантыйском языке левое ветвление, следовательно, мы ожидаем, что вершины будут следовать за зависимыми. Однако можно считать, что неаффиксальные вершины (т. е. фонологические слова) ветвятся вправо. Это относится и к некоторым другим модификаторам, которые, вероятно, являются вершинами (например, кванторы, приименные неопределенные местоимения). Единственные левоветвящиеся вершины в ИГ — это аффиксальные (число, посессивные показатели, падеж).

Анализ прономинальных форм предполагает, что они имеют ту же структуру, что и приименные, т. е. они являются находятся в вершине D0. При этом предлагается постулировать в их внутренней структуре семантически пустое и просодически дефектное имя N0, которое озвучивается как -*t* в *śit* и -*i* или Ø в *tăm(i)* и *tum(i)*. Дефектное имя присоединяется к ближайшему фонологическому слову слева (т. е. демонстративной основе), но она по-прежнему несет словоизменительную морфологию. Подобный анализ можно распространить и на местоименные посессоры и числительные, которые ведут себя похожим образом в аргументных позициях.

Прономинальные демонстративы не могут быть модифицированы низкими зависимыми, поскольку это приводит к сбою линеаризации. Дефектное имя присоединяется к самому низкому возможному зависимому, которое является фонологическим словом — числительным, местоименным посессором или демонстративом. Если между N0 и “мишенью” появляется другое зависимое, N0 попытается присоединиться к этому зависимому (например, к лексическому посессору), но зависимое не присоединяет морфологию (например, поскольку это полноценная словоформа), и это приводит к сбою: структура не может быть линеаризована.

Возможность модифицировать прономинальные демонстративы согласуемым посессором может быть объяснена тем, что посессор передвигается в более высокую позицию в узком синтаксисе. Таким образом, не возникает проблемы линеаризации поднятого посессора слева от демонстратива, дефектного имени и морфологических маркеров. Заметим, что, по предварительным данным Д. Бурова, субстантивированные числительные также могут быть модифицированы только зависимыми, которые находятся выше в структуре. На данном этапе можно предположить, что случаи “именного эллипсиса” в севернохантыйском демонстрируют те же закономерности, что наблюдаются в поведении прономинальных демонстративов. Это говорит в пользу единого анализа, подобного предложенному. Более того, текущее предположение о вершинном статусе демонстративов прекрасно согласуется с семантическим анализом севернохантыйских демонстративов, предложенным в [176].

Тем самым, работа над темой «Анафора и связывание» позволила описать ограничения на употребления двух классов местоимений. Эти ограничения касаются, с одной стороны, статуса антецедента и соотношения между антецедентом и местоимением, а с другой стороны, соотношения внутри именной группы и / или группы детерминанта (DP) местоимений и других элементов. При этом если часть ограничений (например, ограничение на синтаксически периферийные антецеденты) были отмечены ранее, а сейчас получили дополнительную экспериментальную проверку, то другие ограничения (ограничения на позицию и сочетаемость демонстративов с другими элементами DP) в полной мере описаны не были. Результаты работы могут быть значимы не только собственно для грамматики местоимений, но и для общетеоретических исследований синтаксических процессов и классов синтаксических элементов.

### 4.4.3 Анафора к посессивным составляющим в казымском диалекте севернохантыйского языка

На материале северохантыйского языка описывался близкий к предыдущему тип местоимений — прономиналы. Конкретнее, рассматривались конструкции, где прономиналы и функционально аналогичные им средства используются для анафоры к посессорам. Как выяснилось, для анафоры к посессорам могут использоваться и личные местоимения (примеры (133)-(134)), и посессивные показатели.



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| măń° | ṕi-n° | dˊuλk°-dˊa-m° | tańa-na | ŋaλka | dˊeʰkuša-ŋ | mo-dˊaʔj° |
| 1sg | снаружи-DAT | выйти-SFS-REFL.1SG | там-LOC | большой | сосна-GEN | ветка-AUG |

|  |
| --- |
| tadˊa-ŋa. |
| иметься-GFS［3SG］ |
| 'Я вышел на улицу, а там большущая ветка сосны валяется' |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| pon° | čĭkˊi | dʹeʰkuša | ḿaʔ-naʔ | kæw°-xăna | patλi-°-š° |
| всегда | этот | сосна | дом-POSS.1PL | рядом-LOC | быть\_воткнутым-GFS［3SG］-PST |
| 'Эта сосна всегда росла возле нашего дома' | | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ?pon° | pĭʰta | ḿaʔ-naʔ | kæw°-xăna | patλi-°-š° |
| всегда | 3SG | дом-POSS.1PL | рядом-LOC | быть\_воткнутым-GFS［3SG］-PST |
| 'Она (эта сосна) всегда росла возле нашего дома' | | | | |



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| dˊuλnu-ŋ | ilˊja | mašina-ŋ | šoλ°-na | ŋăj-m | tata-ŋa |
| утро-GEN | Илья | машина-GEN | вращаться-PT.NPST | нога-ACC | заменить-GFS［3SG］ |
| 'Илья утром поменял колесо (чьей-то) машины, а вечером починил её двигатель.' | | | | | |

Работа над данной частью темы на сегодняшний день не завершена. Тем не менее, уже сейчас ясно, что выявленные типы анафоры важны для синтаксического описания периферийных составляющих — набор и свойства используемых для такой анафоры средств могут быть важны и для типологии анафоры, и для описания синтаксического поведения посессоров.

## 4.5 Согласование

Несколько исследований, проведенных в 2025 году, были посвящены согласованию — одной из базовых грамматических операций. Все они использовали материал русского языка и затрагивали проблему семантического согласования: по роду, числу и одушевленности. Были использованы корпусные и экспериментальные данные.

### 4.5.1 Согласование по роду

В языках, где есть грамматический род, возникает следующая проблема: многие существительные, обозначающие профессии и социальные роли, грамматически относятся к мужскому роду. Между тем, в современном обществе эти роли доступны и женщинам. Как и в других языках, в русском за последние сто с лишним лет появилось много новых слов женского рода для обозначения профессий и социальных ролей. Однако есть и другая стратегия: слова, исторически относившиеся к мужскому роду, допустимо использовать по отношению к женщинам, причем как с согласованием по мужскому роду, так и с согласованием по женскому (например, *Мария Ивановна — наш* / *наша педиатр*). Эта стратегия достаточно необычна кросслингвистически и имеет целый ряд интересных ограничений, поэтому ее описанию посвящен целый ряд работ, написанных в рамках как формальной, так и функциональной традиции (напр., [177–183]). Так, согласование по женскому роду считается допустимым только в именительном падеже. Если в предложении есть атрибутивное и предикативное согласование, как в примерах (58a-d), не все сочетания будут грамматически правильными. Общий принцип, который проявляется и в других конструкциях, заключается в том, что, если речь идет о женщине, всё согласование может быть семантическим (по женскому роду) или формальным, возможно и смешивать их, но только если семантическое согласование дальше от вершины. Так как такие примеры, как (137–140), были использованы в качестве стимульного материала в двух проведенных в данном исследовании экспериментов, в них показаны экспериментальные условия.

1. *Седой педагог дружески похлопал выпускника по плечу.* (условие MM)
2. *Седая педагог дружески похлопала выпускника по плечу.* (условие FF)
3. *Седой педагог дружески похлопала выпускника по плечу.* (условие MF)
4. \**Седая педагог дружески похлопал выпускника по плечу.* (условие \*FM)

При этом экспериментальных исследований, посвященных приемлемости таких конструкций (например, чтобы ответить на вопрос, будут ли такие примеры, как (137), (138) и (139), оценены одинаково) и их обработке в реальном времени (прежде всего, чтобы проанализировать, как строятся связанные с согласованием ожидания читающего), практически нет. В рамках данной работы было проведено три эксперимента, чтобы заполнить эту лакуну. Опираясь на экспериментальные исследования подобных существительных в других языках, было решено уделить особое внимание не только грамматическим факторам, но и одному экстралингвистическому: гендерным стереотипам, связанным с различными профессиями и социальными ролями. В целом проведенное исследование продемонстрировало, что семантическое согласование по роду в русском языке, которое казалось детально изученным, прежде всего, в рамках формальной традиции, устроено намного сложнее, чем предполагалось раньше.

#### 4.5.1.1 Эксперимент 1. Дизайн

Все эксперименты проводились удаленно на платформе PCIbex [163]. В Эксперименте 1 была использована методика пословного чтения с самостоятельной регулировкой скорости. Когда предложение появлялось на экране, все слова были замаскированы дефисами. Затем участник нажимал на клавишу, и появлялись первое слово. После второго нажатия появлялось второе слово, а первое снова маскировалось, и замерялось время реакции. В эксперименте приняло участие 62 респондента.

В качестве стимулов были использованы 32 набора предложений. В этом эксперименте исследовалось только предикативное согласование. В Группе 1Ж и 1М подлежащими были названия профессий: стереотипно «женских», как в примерах (141)–(gls-ex60b), и стереотипно «мужских» (например, *мясник*, *щвейцар*). Данные о стереотипах были взяты из статьи [184]). Чтобы исключить влияние дополнительных факторов, во всех экспериментах были использованы названия профессий, у которых в литературном русском языке нет пары женского рода. В Группе 2 подлежащими были существительные, обозначающие людей с различными качествами, как в примерах (143)–(146). Абсолютное большинство таких слов — парные, с каждым грамматика допускает только один вариант согласования, употребление слова мужского рода для обозначения женщины (и наоборот) также не допускается. На примере Группы 2 можно установить, как читающий реагирует на ошибку при согласовании по роду с одушевленным подлежащим. Затем можно сравнить с ней Группы 1Ж и 1М.

1. *Педиатр был обеспокоен из-за объявления карантина.* (условие M)
2. *Педиатр была обеспокоена из-за объявления карантина.* (условие F)
3. *Интриган был осторожен в этом вопросе.*
4. \**Интриган былa осторожна в этом вопросе.*
5. *Интриганка была осторожна в этом вопросе.*
6. \**Интриганка был осторожен в этом вопросе.*

Материалы также включали 64 грамматически правильных предложения-филлера. После трети случайно отобранных предложений задавались вопросы на понимание, чтобы гарантировать, что участники читают достаточно внимательно.

#### 4.5.1.2 Эксперимент 1. Результаты

Для статистического анализа были использованы обобщённые линейные модели со смешанными эффектами. Сперва были проанализированы предложения в Группе 2 (факторы: род предиката и грамматическая правильность). Только второй фактор оказался значимым — в неправильных предложениях читались дольше слова со второго по пятое. Таким образом, нельзя сказать, что согласование по женскому роду всегда труднее обрабатывать, чем согласование по мужскому. Между тем, в Группе 1Ж и особенно в 1М согласование по женскому роду обрабатывалось дольше.

Затем были проанализированы предложения в Группах 1Ж и 1М (факторы: род предиката и стереотипы). На втором слове значимым оказалось взаимодействие этих факторов (β = 34,55, SE = 13,56, p = 0,01), т. е. задержка, связанная с согласованием по женскому роду, была более выраженной в Группе 1М, чем в Группе 1Ж. Таким образом, гендерные стереотипы влияют на обработку, причем начиная с самых ранних этапов. На третьем слове значимым оказался фактор рода сказуемого (β = 27,11, SE = 11,32, p = 0,02): согласование по женскому роду обрабатывалось медленнее, хотя грамматика и допускает его использование.

#### 4.5.1.3 Эксперимент 2. Дизайн

Была использована та же методика, что в Эксперименте 1. Было привлечено 52 участника. В этот эксперимент было включено и предикативное, и атрибутивное согласование. Были созданы 32 набора стимульных предложений, таких, как в примерах (143)–(146) выше. В половине подлежащими были названия стереотипно «женских» профессий (Группа Ж), в половине — стереотипно «мужских» (Группа М).

#### 4.5.1.4 Эксперимент 2. Результаты

Среднее пословное время чтения в различных условиях представлено на [рисунок 4.2](#fig-f2) и [рисунок 4.3](#fig-f3). Прежде всего, бросается в глаза, что предложения в Группах Ж и М читаются очень по-разному. Поэтому мы анализировали их по-отдельности. Были построены две модели с условием в качестве фиксированного эффекта, после чего были сделаны попарные сравнения между всеми условиями.

|  |
| --- |
| Рисунок 4.2 – Среднее пословное время чтения (в мс) в Группе Ж (с названиями стереотипно «женских» профессий). |

|  |
| --- |
| Рисунок 4.3 – Среднее пословное время чтения (в мс) в Группе М (с названиями стереотипно «мужских» профессий). |

Попарные сравнения между условиями в Группе Ж (с названиями стереотипно «женских» профессий) показали, что читающие замедлялись после прилагательного в женском роде по сравнению с мужским, но большинство различий между условиями не достигали значимости. Таким образом, хотя согласование по женскому было грамматически правильным, его обработка все же занимала немного больше времени, как и в Эксперименте 1. Однако это замедление было очень кратковременным. Более того, читающие успешно использовали грамматическую информацию о прилагательном для обработки рода глагола: условия MM и FF имели практически одинаковое время чтения начиная с глагола. В условиях MF и \*FM читающие реагировали на несоответствие между родом прилагательного и глагола, но, особенно начиная с пятого слова, задержка была больше в неграмматичном условии \*FM.

В Группе M задержка, вызванная женским родом прилагательного, была гораздо больше, чем в Группе Ж. Более того, если в Группе Ж читающие оперативно использовали род прилагательного, чтобы сформировать ожидания относительно рода глагола, в Группе M они не смогли этого сделать. Читающим было так сложно приспособиться к несоответствию между стереотипами и родом прилагательного, что когнитивная нагрузка нарушила эффективную обработку нескольких следующих слов. В результате четвертое слово (глагол) в условии FF читалось практически так же долго, как и в неграмматичном условии \*FM. В условии MF время чтения было значительно дольше, чем в других трех, и эта картина сохранялась до шестого слова, хотя условие MF грамматически правильно, в отличие от \*FM. Условие \*FM также значимо отличалось от MM и FF на пятом и шестом слове, в то время как последние два не отличались значимо и в конечном итоге сходились. Таким образом, информация о роде прилагательного помогала обработать согласование по роду на глаголе, как и в Группе Ж, но этот эффект был значительно отложен.

#### 4.5.1.5 Эксперимент 3. Дизайн.

Были использованы те же материалы, что в Эксперименте 2 (24 стимульных набора из 32 и 24 филлера), но другой метод — оценка приемлемости по шкале Ликерта от 1 до 5. Были привлечены 98 респондентов. Предложения предъявлялись по одному.

#### 4.5.1.6 Эксперимент 3. Результаты

Предложения в неграмматичном условии \*FM получили самые низкие средние оценки (1,5 для профессий с «женскими» стереотипами, 1,3 для профессий с «мужскими»), а предложения в условии ММ — самые высокие (4,6 и 4,7 соответственно). Таким образом, участники предпочли согласование по мужскому роду даже для стереотипно «женских» профессий и продемонстрировали чувствительность к грамматической норме, что не было очевидно в предыдущем эксперименте. Средние оценки в условиях FF и МF оказались в середине шкалы, хотя грамматика современного русского языка допускает его: 3,1 и 2,5 для FF и 3,2 и 2,8 для MF. Особенно интересным было то, что эти данные не повторяют закономерности, наблюдавшиеся в Эксперименте 2, где предложения в условии FF обрабатывались значимо быстрее, а условие МF вообще оказалось наиболее трудным для обработки для названий «мужских» профессий. Можно сделать вывод, что, справившись с этими сложностями, носители опираются на другие принципы, оценивая приемлемость предложений.

Для статистического анализа были использованы порядковые регрессии со смешанными эффектами. Первая модель, включающая весь набор данных с условием в качестве фиксированного эффекта, сопровождалась парными сравнениями между всеми условиями. Подтверждая результаты описательного анализа выше, все условия значительно отличались друг от друга, за исключением FF и MF (MM vs. FF: β =4,92, SE = 0,54, p < 0,001; MM vs, MF: β = 4,53, SE = 0,54, p < 0,001; MM vs, \*FM: β = 8,39, SE = 0,53, p < 0,001; FF vs, \*FM: β = 3,47, SE = 0,42, p < 0,001; MF vs, \*FM: β = 3,86, SE = 0,43, p < 0,001).

Вторая модель была построена для проверки роли стереотипности в приемлемости согласования по женскому роду в различных условиях. Соответственно, она включала только условия с согласованием по женскому роду; условие и стереотипность рассматривались как фиксированные эффекты. Стереотипно «мужские» профессии значительно снижали оценки (β = -0,87, SE = 0,30, p < 0,01). Условие \*FM значимо отличалось от условия FF, взятого в качестве базового уровня (β = -2,85, SE = 0,18, p < 0,001), о чем уже известно из предыдущей модели. Наконец, взаимодействие между стереотипностью и условием \*FM было значимым (β = 0,68, SE = 0,33, p = 0,04), что показывает, что в этом условии эффекты стереотипности были перекрыты неграмматичностью.

### 4.5.2 Согласование по одушевленности

**Данные.** Категория одушевлённости в русском языке выделяется только на основании выбора показателя винительного падежа (для существительных мужского рода 2 склонения — в обоих числах, а для остальных имен — только в множественном числе). При этом, как и в случае рода, числа и падежа, в винительном падеже адъективные модификаторы (причастия и прилагательные), а также местоимение *который* согласуются с вершиной по одушевленности, ср. *(Мы увидели)* ***сидевшего*** *на столе кота*/***стоявший*** *на столе стакан*.

Однако в некоторых случаях одушевлённость на модификаторе отличается от одушевлённости контролёра согласования. В рамках данного исследования были проанализированы конструкции, где адъективная единица согласуется с сочинённой группой. Некоторые группы конъюнктов (названия стран, спортивных команд и других организаций), будучи грамматически неодушевлёнными, тем не менее допускают одушевлённый тип аккузатива на мишени согласования:

1. *Чтобы добраться до финала Кубка лиги, нам нужно было переиграть «Ливерпуль» и «Тоттенхэм»,* ***добравшихся*** *до решающего матча Лиги Чемпионов.* [Отдел «Спорт». Сарри: Английская премьер-лига — сильнейшая в мире // gazeta.ru, 10.05.2019]

Было установлено, что такие примеры не только допустимы, но даже преобладают в корпусе. По данным газетного подкорпуса Национального корпуса русского (<https://www.ruscorpora.ru>), *который* при согласовании с сочинёнными группами в 29 случаях имеет одушевлённую форму аккузатива и в 23 — неодушевлённую. Постпозитивное причастие в 17 случаях согласуется по одушевлённому типу и в 12 случаях — по неодушевлённому. Кроме того, опрос носителей (122 участника, оценивавших разные примеры) также выявил предпочтение одушевленной формы.

**Анализ.** Для этого нестандартного типа оформления было предложено следующее объяснение. Одушевлённое маркирование в случаях стран и организаций можно рассматривать как семантическое согласование. Это очень важный вывод, так как исследования семантического согласования, в том числе в формальных подходах, ранее не затрагивали согласование по одушевленности. Действительно, хотя грамматически эти имена являются неодушевлёнными (ср. *мы победили Пакистан* / \**Пакистана*), их сочетаемостные свойства обнаруживают более сложную природу. Например, во многих контекстах с этими типами имён возможно исключительно или предпочтительно местоимение *кто*, а не *что*:

1. *Кто победил* / \**что победило — «Спартак» или «Динамо»?*

При этом семантическое согласование возможно в связи с тем, что не создаёт грамматического конфликта. С одной стороны, сочинённые группы не имеют лексических характеристик, в том числе и грамматической одушевлённости. С другой стороны, при согласовании по числу со всей группой (ср. множественное число в примерах типа (147)) одушевлённость также не обязана согласовываться с каждым из конъюнктов — а значит, не возникает конфликт между одушевлённостью каждого из конъюнктов и мишени согласования.

Семантическая сущность данной модели согласования подтверждается ещё и тем, что выбор одушевлённого согласования сильно ограничен контекстом: в примерах типа (149), где страны или организации концептуализуются как место, конечная точка и другие типы участников, стандартно являющиеся неодушевлёнными, одушевлённый тип согласования также невозможен.

1. *В течение последнего времени российские власти направили запросы о выдаче Невзлина в Израиль и США,* ***которые*** *тот посещал накануне.* [Мосгорсуд разрешил не знакомить адвокатов Невзлина с его делом // Lenta.ru, 15.12.2005]

### 4.5.3 Согласование по числу с т.н. «малыми именами» и контроль

А. Перельцвайг [185] утверждает, что в русском языке существует особая категория именных групп, не проецирующих DP — т.н. «малые имена», которые лишены некоторых синтаксических свойств, присущих полным DP. В позиции подлежащего такие именные группы не могут контролировать глагольное согласование. Кроме того, «малые имена» не могут быть контроллерами PRO. Это утверждение было оспорено в последующих работах (напр., [186; 187] на основе интроспективных суждений и корпусных данных. В рамках данного исследования способность русских «малых имён» быть контроллерами PRO была оценена в двух экспериментах.

#### 4.5.3.1 Эксперимент 1. Дизайн

Эксперимент проводился удаленно на платформе PCIbex [163]. Участников просили оценить приемлемость каждого предложения по шкале Ликерта от 1 до 7. Были привлечены 130 участников. При подготовке стимульных предложений варьировались три фактора: (i) единственное или множественное число согласования на глаголе; (ii) наличие или отсутствие предиката субъектного контроля; (iii) наличие или отсутствие QP-модификатора (напр., *около*). Первый и второй факторы проиллюстрированы в примерах (150)–(151).

1. *Шестьдесят студентов хочет/хотят сдать экзамен* (±ед.ч.; +контроль).
2. *Шестьдесят студентов сдало/сдали экзамен* (±ед.ч.; –контроль).

Каждый экспериментальный лист содержал 16 филлеров и 16 стимульных предложений (таким образом, каждый участник сталкивался с каждым экспериментальным условием дважды). Предложения предъявлялись по одному. Оценки каждого участника были преобразованы в z-показатели для устранения потенциального смещения шкалы. Нулевая гипотеза заключалась в том, что стимулы в условии (+ед.ч.; +контроль) получат оценки приемлемости, сравнимые с неграмматичными филлерами.

#### 4.5.3.2 Эксперимент 1. Результаты.

Грамматичные филлеры получили среднюю z-оценку 0,70, в то время как неграмматичные дистракторы — -1,20. Оценки в условии с ед.ч. не соответствовали ни грамматичным, ни неграмматичным филлерам: средняя z-оценка для условия (+ед.ч.; –контроль) составила 0,08, а z-оценка для условия (+ед.ч.; +контроль) — -0,22. Для статистического анализа была использована обобщённая линейная модель со смешанными эффектами. Она показала, что эффект числа значим (β = -0,601, SE = 0,070, p < 0,001), эффекты контроля и наличия модификатора не значимы, однако значимы взаимодействия ед.ч. и контроля (β = -0,223, SE = 0,099, p < 0,05) и ед.ч. и наличия модификатора (β = 0,197, SE = 0,099, p < 0,05).

#### 4.5.3.3 Эксперимент 2. Дизайн

Целью этого эксперимента было дополнительно проверить влияние порядка слов на приемлемость предложений с согласованием по ед.ч. — как утверждается, например, в работе [185], согласование по ед.ч. является предпочтительным, если субъект находится в постглагольной позиции. Для этого был использован тот же дизайн, что в Эксперименте 1, только субъекты в предложениях располагались справа от глагола. В эксперименте приняло участие 54 респондента.

#### 4.5.3.4 Эксперимент 2.Результаты

Средняя z-оценка грамматичных филлеров составила 0,67, а неграмматичных — -1,30. Оценки в условии с ед.ч. оказались выше, чем в Эксперименте 1, но, как и прежде, не соответствовали ни грамматичным, ни неграмматичным филлерам: средняя z-оценка для условия (+ед.ч.; –контроль) равнялась 0,41, а для условия (+ед.ч.; +контроль) — -0,06.

#### 4.5.3.5 Общие выводы

Таким образом, по результатам двух экспериментов можно заключить, что оценки в условии с единственным числом и контролем не совпадают с оценками для неграмматичных филлеров. Это противоречит предсказаниям [185], хотя согласование по единственному числу всё же ухудшает оценки предложений с контролем — это значимый, хоть и количество небольшой эффект. Предварительный вывод состоит в том, что сильная версия гипотезы о «малых именах» должна быть отвергнута.

## 4.6 Развитие эмпирической базы исследований в области формальных моделей языковых явлений на материале языков России

В соответствии с тем, что в цели исследования входит не только предложить уточнения существующих моделей для разных языковых явлений, но и (1) расширить эмпирическую базу исследований, вовлекая естественно-языковые данные языков России, до сих пор не учитывающихся при моделировании, (2) расширить эмпирическую базу верификации теорий, необходим целый ряд ресурсов. Во-первых, это корпуса спонтанных текстов на так называемых малоресурсных языках; во-вторых, это различные ресурсы, обобщающие данные по разным языкам относительно одного языкового феномена; в третьих, это экспериментальные данные относительно оценки приемлемости предложений при тестировании той или иной гипотезы относительно исследуемых явлений (например, механизма согласования, механизма отрицательного согласования и др.).

В рамках исследований предыдущих лет были собраны в полевых условиях и размечены в среде FieldWorks Language explorer (FLEX, <https://software.sil.org/ru/fieldworks/>) корпуса устных текстов на хантыйском и мокшанском языках. В 2025 г. проводилась выверка поморфемной аннотации, дополнительная ручная проверка, выполненная автоматически в среде, а также подготовка данных к представлению их в среде Tsakorpus. Эта среда представляет собой корпусной менеджер, обеспечивающий поиск по корпусу как по лексемам, так и по грамматическим тегам. В связи с тем, что якутские данные уже представлены в рамках других проектов, решено было расширить эмпирическую базу за счет мокшанских данных, собранных с 2013 по 2015 г., также обработанных в среде FLEX, но нуждающихся в дополнительной работе по включению большего количества текстового материала (не представленного в книге [188]), а также в выверке поморфемного аннотирования с учетом дополнительной разметки по порядку слов и дифференцированному маркированию объекта и в подготовке к размещению в специальной среде Tsakorpus ([189]), обеспечивающей поиск. Также важным источником данных являются корпуса литературных вариантов малых языков. Корпус литературного даргинского языка, созданный в рамках других проектов, к которому была применена автоматическая разметка [190], нуждался в дальнейшем усовершенствовании, включении художественной литературы, для этого была необходима дополнительная ручная правка плохо распознанных литературных текстов.

В 2025 году была начата разработка типологических ресурсов. В этом году рассматривалась проблема нефинитных форм. Традиционно нефинитные формы в грамматиках «привязаны» к одной из частей речи, однако, для многих языковых групп характерно, что нефинитные формы имеют один и тот же формант при образовании, но могут выступать в разных синтаксических функциях. Например, это характерно для уральских языков. Несмотря на то, что эти свойства нефинитных форм неоднократно обсуждались в литературе, до сих пор отсутствовала систематизация различных случаев полифункциональности. Результатом такой систематизации является подготовленный набор данных для базы данных по свойствам нефинитных форм, в которой отражены различные свойства нефинитных форм в языках Северной Евразии: в каких синтаксических конструкциях они доступны, какие значения категорий объединяются, по каким грамматическим категориям изменяются и др. Второй параметр, который был выбран для таких обобщений — это явление согласования. Существующие базы, такие как база Корбетта (Surrey Morphology Group), являются важным инструментом лингвистического анализа и охватывают базовые типологические параметры, морфологические особенности согласования по лицу, числу, роду, а также субъектное и объектное согласование. Однако у неё есть недостатки: ограниченный набор параметров, ограниченность данных по приименному согласованию. Также ограниченность демонстрирует и авторитетный WALS (World Atlas of Language Structures) — данные по согласованию ограничены в ней лишь несколькими параметрами.

Разработан прототип базы, в которой будут учитываться расширенные типологические параметры, такие как дефолтное согласование — использование в согласовании форм, выбранных по умолчанию, когда стандартный контролёр отсутствует или неочевиден. Включено согласование существительного с числительным. Также предусмотрено описание параметра согласования при дифференцированном маркирования объекта, что актуально, например, для уральских языков, множественного выражения признака (multiple exponence), согласования с посессором и др. Особое внимание планируется уделить языкам России: уральским, нахско-дагестанским, абхазо-адыгским, тюркским, чукотско-камчатским и тунгусо-маньчжурским.

Также важным ресурсом является такой ресурс, где бы можно было увидеть, какие наборы данных используются при исследовании различных языковых феноменов для оценки грамматичности предложений с варьированием грамматических параметров.

Ресурсы по мере подготовки будут доступны по ссылке на страницу ресурсов Лаборатории по формальным моделям в лингвистике (<https://lingformlab.ru/>).

## 4.7 Заключение

Таким образом, в результате исследований были получены уточнения формальных моделей (как в области формальной семантики, так и синтаксиса) для следующих областей: семантика и синтаксис видо-временных категорий, анафора и связывание, свойства логических операторов, дистантные передвижения в полипредикации. Для этого привлекались данные русского, хантыйского, абазинского, кильдинского саамского, мишарского татарского и др.

Для исследований русского языка активно применялись экспериментальные методы, которые позволяют верифицировать модели, часто основанные на интроспекции. Кроме того, экспериментальные методы позволили выявить новые нетривиальные особенности согласования в русском языке, ранее не учитываемые в теории. Материалы других языков были получены в ходе полевых исследований методом элицитации. Эти методы получили дальнейшее развитие, в частности, применялись современные методы семантического опроса.

Для развития эмпирической базы для дальнейших исследований был разработан ряд ресурсов, некоторые из которых открыты для общего доступа (в частности, корпус текстов на хантыйском языке, типологическая база нефинитных форм в языках северной Евразии).

В результате разработан эксплицитный формальный анализ семантики наречного сдвига в казымском хантыйском, позволяющий предсказать несовместимость относительного времени с индексикальными наречиями: относительное время связывается локально доступным временем матричной клаузы, а наречия интерпретируются относительно момента речи. Также разработана эксплицитная семантическая модель причастного показателя -*(m)a* в кильдинском саамском, объясняющая все употребления этого показателя (показателя упорядочивающего аспекта, вводящего отношение предшествования). Исследование синтаксических свойств дистантного скрэмблинга в мишарском татарском показало, что такой скрэмблинг возможен вне зависимости от падежного маркирования субъекта, что ставит под сомнение описательную и объяснительную адекватность ряда существующих моделей. Основными результатами исследования грамматических свойств логических операторов стало уточнение эмпирической картины — необходимый этап при верификации гипотез в основе эксплицитной формальной модели соответствующих явлений. Выявлены факты, ставящие под сомнение анализ русских местоимений на -*либо* как выражений отрицательной полярности, и выдвинута альтернативная гипотеза, подлежащая дальнейшей разработке. В результате экспериментальных исследований уточнены ограничения на лицензирование отрицательных и реципрокальных местоимений в русском языке, а также ограничения на согласование по роду и числу, в том числе в зависимости от референциального статуса контролёра согласования. Исследование синтаксического поведения анафорических элементов в казымском хантыйском позволило описать ограничения на употребления двух классов местоимений — демонстративов и прономиналов, заложив основу для разработки модели их поведения.

Исследования в рамках уточнения формальных моделей отдельных языковых явлений, а также более глубокое их рассмотрение с использованием экспериментальных методов велись в следующих направлениях: семантика и синтаксис грамматических глагольных категорий, анафора и связывание, исследование согласования экспериментальными методами, грамматические свойства логических операторов.

# 5. Языковое и когнитивное многообразие в логико-философском анализе и формальных моделях

Исследование коммуникативных компетенций ограниченно рациональных агентов было направлено на поиск новых подходов к разрешению затруднений, препятствующих эффективному информационному обмену в условиях языкового и когнитивного многообразия. В центре внимания находились такие классические для логики и формальной эпистемологии проблемы как концептуальная неопределенность и референциальная непрозрачность, скептицизм в отношении значения, логические и семантические парадоксы, а также новые теоретические и практические вызовы, обусловленные развитием неклассических логических систем, систем искусственного интеллекта и социальных сетей.

Методологически исследование характеризовалось укорененностью в существующих логико-философских традициях и обращением к их истокам не в сугубо исторической, но в реактуализирующей эти традиции перспективе, которое сопровождалось их переосмыслением в контексте современных проблем семантики, семиотики и когнитивных наук.

## 5.1 Реактуализация семиотического наследия Ч.-С. Пирса в цифровой коммуникации

Подтверждена актуальность семиотической теории Ч.-С. Пирса для новейших разработок в области искусственного интеллекта и коммуникации в социальных сетях. Методология синехизма, разработанная в поздних трудах Ч.-С. Пирса, распространена на современные подходы к коммуникации и научным исследованиям, включающие искусственный интеллект. Согласно доктрине синехизма, «непрерывность управляет всей областью опыта в каждой ее части» [191: 343–344; ср. 192] и, следовательно, «все, что существует, непрерывно» [193: 70–71].

В общем виде синехизм был сформулирован Ч.-С. Пирсом как такой взгляд на науку, который, являясь «регулятивным принципом логики, предписывающим, какого рода гипотезы могут быть рассмотрены и исследованы» [194: 118], предотвращает появление исследовательских лакун, поскольку акцентирует внимание на непрерывности исследуемого.

Действие регулятивного принципа синехизма было прослежено в наиболее технологичной области современных биологических исследований, связанной с использованием искусственного интеллекта, позволяющего, в частности, строить модели, в которых способность клеток справляться с новыми задачами используется для решения обратных биоинженерных задач [195: 689]. Были выявлены следующие нетривиальные особенности рассмотренной технологии, реактуализирующие синехизм Ч.-С. Пирса и его теорию рассуждения. Во-первых, возможность самостоятельного исправления ошибок позволяет моделировать активный вывод [196–198] и, следовательно, ретродуктивное рассуждение (абдукцию) Ч.-С. Пирса [199]. Во-вторых, наличие «обратной проблемы» также свидетельствует о релевантности рассуждения по абдукции, в результате которого новая идея возникает в качестве рабочей гипотезы. В-третьих, прагматическая максима Ч.-С. Пирса управляет исследованием, ориентируя его на истолкование нового понятия в терминах следствий его использования. В-четвертых, интеллектуальный континуум современной науки включает в себя технологии искусственного интеллекта, которые превосходят человеческие способности в том, что касается распознавании паттернов. Как следствие ожидается появление новых подходов к осмыслению методов рассуждений и навыков решения задач, которые недоступны естественному интеллекту.

Показано, что методологический синехизм оказывается связующим звеном, обеспечивающим инвариантность принципов обучения и оптимизации на различных уровнях организации интеллекта, управления и коллективного принятия решений. В полном соответствии с пан-инференционализмом Ч.-С. Пирса, эволюционные задачи аналогичны на разных уровнях организации, а в схемах их решения наблюдается общий паттерн рассуждения как движения мысли от посылок к заключениям. Два самых ярких примера — диахроническая обработка информации (вычисление) и зависимая от опыта модификация внутренней структуры для дальнейшего взаимодействия с окружением (память).

Эволюционизм, образующий фундамент синехизма в сфере биологии, оказывается важным и для теории знаков Ч.-С. Пирса, поскольку знаки являются с его точки зрения объектами эволюции, поддерживающими «прирост значения» [193: 130]. По мере интенсификации коммуникации системы с ее средой, а также внутри системы возникают новые значения в рамках сигнальных систем языка и другая информация, способная запустить и поддержать инновационные циклы обучения. Предписываемая прагматизмом оценка понятий в терминах следствий их употребления сдвигает фокус внимания при изучении языка на прагматику и контекст.

Таким образом, реактуализация пирсовской методологии синехизма, оказывается значимой для самого широкого исследовательского контекста современной науки. Тезис синехизма о непрерывности поддерживает трансфер идей из разных научных областей.

Реактуализация семиотического наследия Ч.-С. Пирса была связана также с интерпретацией интернет-мемов как семиотических феноменов с позиции его теории знаков. Была представлена критика биологизаторской трактовки мемов как «генов культуры» [200; 201] и предложено анализировать их структуру и трансформацию через понятия знака, иконичности, типа, токена и привычки вывода. Особое внимание было уделено механизмам, которые позволяют интернет-мемам сохранять узнаваемость при изменчивости их формы.

Интернет-мемы были определены как единицы информации, имеющие преимущественно юмористическое содержание и обычно оформленные в виде коротких видеороликов или изображений с подписями (другие определения см. в [202]). Успех мема зависит от его способности мгновенно передавать сообщение; в этом смысле мем, несомненно, является знаком, имеющим семантику. Однако мем не просто рационально воспринимается и анализируется, он влияет на наши чувства, потому что его информационная составляющая сливается с эмоциональной. Чтобы понять мем, мы должны эмоционально взаимодействовать с идеей, которую он несет, «чувствовать» её, и именно поэтому может быть сложно передать сообщение, заключённое в меме, словами, то есть посредством набора предложений. (Важно, однако, не путать передачу идеи мема с его объяснением, которое вполне возможно сформулировать с помощью предложений.)

Интернет-мемы были рассмотрены как знаки, сочетающие визуальный и текстовый компоненты. Показано, что с точки зрения классификации знаков по Ч.-С. Пирсу интернет-мемы могут быть отнесены к иконическим знакам, которые передают эмоции и настроения через метафорическое сходство с объектом. Икона-метафора, которую представляет собой мем, подразумевает существование опосредующего элемента, и эта роль отведена культуре, с которой интерпретатор должен быть знаком. Мемы функционируют в рамках конкретной культурной среды, что ограничивает их распространение за её пределы. Это связано с необходимостью для интерпретатора понимать культурные коды, заложенные в конкретном меме.

Вместе с тем на любой интернет-мем можно смотреть как на тип, а можно –– как на токен данного типа. Тип — это общая идея мема, его структура, токен же –– единичное воплощение типа (например, мем с определённым изображением и текстом). Создание новых токенов, сохраняющих узнаваемость типа, представляет собой процесс трансформации мема. В исследовании использовался также термин «появление», введенный Ф. Беллуччи и Дж. Бертоном [203] для обозначения копий одного и того же знака: токен –– это конкретный экземпляр типа, а появление –– это копия токена. Отмечается, что можно говорить о трансформации мема только в случае возникновения его нового токена, а не в случае его повторного появления. Новый токен появляется, когда исходный тип этого мема сохраняется в новом токене, и интерпретатор распознаёт его как таковой.

В центр исследования был поставлен вопрос о том, что определяет те изменения в меме, которые обуславливают появление нового токена того же типа. Теория таких изменений разработана с использованием введенного Ч.-С. Пирсом понятия «привычка вывода» (habit of inference). (Детальное исследование роли этого понятия в семиотике Пирса см. у Ф. Белуччи и А.-В. Пиетаринена [204]; также см. [205]). По словам самого Пирса, привычка «может быть выражена в предложении, в котором антецедент должен описывать все возможные предпосылки, на которых она может действовать, а консеквент должен описывать, каким образом вывод, к которому она приведет, будет связан с этими предпосылками» [193: 354]. Пирс использует этот термин вместе с понятием руководящих принципов, которые «могут быть сформулированы в предложении, утверждающем, что каждое предложение C, связанное заданным общим образом с любым истинным предложением P, является истинным» [206: 107]. Было показано, что руководящий принцип функционирует на логическом уровне, а будучи помещен в реальный жизненный контекст, уступает место социально ориентированной и, следовательно, более естественной привычке. Именно привычка регулирует формирование вывода на уровне конкретной информации, и это образует логическое ядро мемов.

Функция руководящего принципа становится наиболее ясной в дедуктивных построениях с проявлением их валидности, однако отсутствие дедуктивной валидности не означает отсутствия самого принципа, поскольку он также отвечает за ценность индуктивного рассуждения с его обобщениями, а также абдукции, стимулирующей новые исследования. Фундаментальная структура дедукции, согласно Ч.-С. Пирсу, регулирует получение результата как применения правила к частному случаю; индукция отвечает за получение нового правила, а гипотеза (в более поздних работах абдукция) позволяет сказать что-то новое о частном случае, основываясь на применении общего правила к конкретному результату. С помощью этих логических компонентов были раскрыты способы трансформации мемов. Процесс понимания мема представлен как псевдодедукция, которая может быть соотнесена с отменяемыми рассуждениями, где роль правила играет тип мема, а конкретное проявление мема — токен — рассматривается как частный случай. Привычка вывода отвечает в этом процессе за то, как трансформируется мем, а также за корректировку процессов, определяющих эту трансформацию. Этот механизм гарантирует, что информация, переданная из токена, сохраняет свой исходный тип. Он функционирует таким образом до тех пор, пока интерпретатор понимает тип мема, а информация, предлагаемая токеном, соответствует типу мема. Если эта связь между типом и токеном прерывается, непрерывность производства и восприятия мема нарушается.

Было сделано заключение, что в паре «изображение –– подпись» тип мема с большей вероятностью сохраняется, если подпись остается без изменений. Хотя возможно также сохранение типа благодаря неизменности именно изображения, в этом случае есть вероятность, что тип не будет правильно распознан, и привычка вывода не сработает, в результате чего мем будет утерян. Последнее объясняется тем, что изображения потенциально могут быть интерпретированы многими способами, в отличие от словесных фраз, где варианты интерпретации ограничены. Выводы были подкреплены анализом двух примеров трансформации мемов: 1) мем «Я и мой начальник» (с изображением монархов) и 2) мем «Рассеянный бойфренд». В первом случае тип задаётся текстом: «Одевайся для работы, которую хочешь иметь», в то время как изображения меняются (разные монархи или персонажи). Во втором случае тип задаётся изображением, а текст меняется. Показано, что во втором случае вариативность выше, но мем остаётся узнаваемым, пока изображение передаёт идею «желания лучшего».

## 5.2 Неопределённость значения в мультиагентной коммуникации

Проблема координации субъективных точек зрения для достижения общей референции рассматривалась в связи с гипотезой Г. Фреге о «логических чужаках» — воображаемых существах, чьи законы мышления противоречат нашим собственным. Во введении к работе «Основные законы арифметики» Г. Фреге описывает мысленный эксперимент, иллюстрирующий опасность смешения логического и психологического, а именно, логических законов бытия истинным и психологических законов признания истинным. Он предлагает представить гипотетических существ, «законы мысли которых прямо противоречат нашим собственным и часто приводят к противоположным результатам даже на практике» [207: 14]. Благодаря работам Дж. Конанта, эти существа известны сегодня как «логические чужаки» (“logical aliens”) [208]. Рассуждения «логических чужаков» не являются ошибочными, поскольку логические ошибки возникают в том случае, когда агенты, подчиняющиеся правилам логики, нарушают эти правила. Однако «логические чужаки» нарушают правила логики просто потому, что не подчиняются им.

С точки зрения Г. Фреге, «логические чужаки» принципиально невозможны, поскольку логические «законы мысли являются наиболее общими законами, универсально предписывающими способ мышления, которым мы обязаны руководствоваться, если вообще хотим мыслить» [207: 12]. Показано, что такой подход сближает Г. Фреге с И. Кантом, признававшим конститутивность логики для мышления. Вместе с тем, выявлено важное различие в трактовке нормативности логики И. Кантом и Г. Фреге.

Кант поясняет конститутивный характер правил общей логики через аналогию с общей грамматикой, «которая не содержит ничего, кроме одной лишь формы языка вообще, без слов, составляющих материю языка» [209: 267]. Подобно тому как невозможно говорить на каком-то языке, нарушая его грамматические правила, невозможно нарушать формальные правила, предписываемые рассудку логикой. Логические ошибки как нарушения таких правил силами одного лишь рассудка оказываются невозможными: «Рассудок сам по себе не может заблуждаться потому, что, когда он действует только по своим законам, результат его (суждение) необходимо должен согласоваться с этими законами, а в согласии с законами рассудка и заключается формальная сторона всякой истины» [209: 215]. Проблема, однако, в том, что хотя в «формальном смысле» логические ошибки являлись бы невозможным в принципе мышлением, противоречащим своим собственным законам, такая невозможность сделала бы избыточной нормирующую функцию логики. Нормативность логики как «критики познания» связывается в кантовской аналитике способностей с дихотомией чувственности и рассудка. К заблуждению может привести лишь «незаметное влияние чувственности на рассудок», на правильное, то есть «согласное с самим собой» его применение [209: 269].

Следуя за И. Кантом в признании конститутивности логики, Г. Фреге критикует, однако, аналогию между правилами грамматики и законами логики, которые он трактует как наиболее общие законы бытия истины. В отличие от И. Канта, для которого «логический критерий истины» есть лишь «негативное условие всякой истины» [209: 95], Г. Фреге считает предметом логики именно истину. Логика нормативна для мышления постольку, поскольку оно направлено на истину: «Мы должны признать, что правила нашего мышления и нашего полагания истинным предписаны законами истины» [207: 12]. Как и для И. Канта, формальность логики является для Г. Фреге следствием ее общности. Однако он рассматривает ее нормативность в отношении мышления как производную от ее дескриптивности. По Г. Фреге, закон любой науки, не только логики, но и, скажем, геометрии или физики, нормативен для мышления в силу того, что он описывает то, что есть и, соответственно, предписывает, как правильно мыслить о том, что есть. Отличие логики от других наук только в предельной степени ее общности.

Таким образом, при всем различии их мотиваций, ни И. Кант, ни Г. Фреге не считают, что «логические чужаки» рассуждают иначе, чем мы, потому что они вообще не признают их рассуждающими. Такая позиция несовместима с признанием плюрализма логических систем.

Природу логического плюрализма помогает истолковать классическая дихотомия, в соответствии с которой различаются два вида логики — *logica docens* (логика преподаваемая и изучаемая) и *logica utens* (логика применяемая в рассуждениях). Показано, что современные дискуссии о плюрализме в области logica docens берут начало от тезиса У. Куайна о «вариативности значений» логических терминов, согласно которому «альтернативные логики практически неотделимы от простого изменения в употреблении логических слов» [210: 355]. Такие изменения могут затронуть не только значения логических связок и кванторов (таких как «или», «и», «все»), но также значения метатеоретических выражений (таких как «валидность», «непротиворечивость», «категоричность»). Однако, поскольку выбор теории и ее концептуальной системы происходят синхронно, тезис о «вариативности значений» логических терминов не исключает возможности рациональной ревизии *logica docens* и, следовательно, плюрализма логических систем.

Центральным оказывается, таким образом, вопрос о том, как справляется с плюралистическим вызовом logica utens. В основу исследования была положена идея, что в реальной коммуникации мы взаимодействуем с когнитивными агентами, уже вовлечёнными, пусть и несовершенным образом, в разнообразные логические практики. Успешность такой коммуникации обеспечивается интерпретационными рассуждениями, цель которых отличается от традиционной цели рассуждений в конкурентном логическом дискурсе. Интерпретационные рассуждения носят не состязательный, а кооперативный характер, ориентируясь на нарративную задачу «понимания истории» или «угадывания подразумеваемой модели» (см. [211]).

Проблема неопределенности значения была рассмотрена также в контексте критики аргумента от перформативного противоречия, направленного против скептического тезиса Витгенштейна-Крипке [212–215].

Реконструкцию аргумента от перформативного противоречия можно свести к следующему. Позиция скептика подразумевает невозможность или бесполезность делать истинные утверждения и, как следствие, формулировать какие-либо теории. В частности, скептицизм относительно значения подразумевает невозможность делать истинные утверждения относительно значения. Чтобы аргумент работал, нужно допустить, что скептик рассматривает свою позицию как некоторую теорию, включающую, по крайней мере, утверждение о невозможности делать какие-либо утверждения. Тогда, отрицая такую возможность, скептик тем самым ее реализует, вступая в перформативное противоречие с самим собой.

Базовый для критики данного аргумента тезис заключается в том, что речевой акт скептика может иметь в контексте его коммуникации иной прагматический характер, отличный от утверждений о реальности, рассматриваемых как истинные, и потому может не вступать в противоречие с позицией самого скептика. С точки зрения введенного Дж. Остином [216] различия между локуцией, иллокуцией и перлокуцией можно высказать сомнение в том, что предложения, локутивно эквивалентные утверждениям, используются скептиком именно в этом иллокутивном модусе, а не являются шуткой, театральной постановкой или вообще бессмыслицей. В результате такого сомнения можно трактовать речь в модусе скептической иллокуции как прагматически осмысленную.

В результате такой реконструкции были получены следующие выводы:

1. Чтобы представить скептический аргумент как направленный на локутивное значение и при этом совместить его с теорией речевых актов, нужно разграничить в семантике два аспекта: с одной стороны, *аспект значения*, т. е. соотнесение языковых выражений с некоторыми объектами; с другой стороны, *аспект значимости*, т. е. cамо наличие таких объектов, которые эти выражения могли бы означать, а также структурных связей между ними. Было обосновано заключение, что скептицизм проблематичен для первого аспекта. Однако подобный взгляд на язык никак не помогает исследователю решать его теоретические задачи. Семантические законы и правила имеют смысл лишь постольку, поскольку мы считаем, что они определенным образом относятся к «нормальной» языковой практике –– регулируют и объясняют эту практику. Скептицизм относительно значения вызывает беспокойство именно потому, что значение нам нужно для уверенности в том, что наши языковые практики регулярны, что наш язык «работает» так, как мы от него ожидаем. И если в какой-то момент наши ожидания не оправдываются, мы хотим знать, почему это произошло. Для этого недостаточно одной лишь абстрактной возможности приписать значения словам так, чтобы вся система языка оказалась семантически согласованной. Нужно понимать, *что именно* слова означают (в значении слова «понимать» основанном на теории «примитивной уместности» [217]). Против возможности такого понимания и нацелен скептический аргумент.
2. Если представить скептический аргумент на иллокутивном уровне, его суть состоит в неопределенности интерпретации, то есть в невозможности понять интенцию, стоящую за употреблением говорящим определенных выражений в конкретном случае. Было указано, что сама постановка проблемы скептицизма относительно иллокутивного значения возможна лишь после отражения скептической аргументации на уровне локутивного значения.
3. Что касается перлокутивного аспекта (реального практического влияния актов коммуникации на чувства, мысли и поведение реципиентов), то именно здесь обеспечивается тот базис, на котором основывается относительная устойчивость и интеллигибельность локутивного и иллокутивного аспектов. Вместе с тем именно перлокутивный уровень делает скептическую позицию философски несостоятельной и, по сути, обеспечивает эффект перформативного противоречия, возникающего не из-за правил языка, а из-за правил и конвенций, существующих в деятельности под названием «философия». Скептическая позиция является самоподрывной именно с точки зрения того, чем мы занимаемся, когда философствуем. Философская деятельность регулируется имплицитными конвенциями, одной из которых является разделение определенной коммуникативной интенции. Участвуя в философской дискуссии, человек обязуется принимать установку на то, что его высказывания будут пониматься как утверждения, претендующие на истинность, и в качестве таковых эти высказывания можно защищать, подвергать сомнению, выводить из них следствия и т. п. (см. [218]). Если мы не можем атрибутировать скептику интенцию сказать нечто значимое, то мы лишаем почвы не только аргумент от перформативного противоречия, но и саму дискуссию о скептицизме. Отсюда следует вывод, что тезис скептицизма относительно значения противоречит воздействию этого высказывания на адресата в рамках определенного института осмысленной речи, именуемого философией.

## 5.3 Парадоксы концептуальной неопределенности

Важным направлением логико-когнитивных исследований стал логический анализ парадоксов концептуальной неопределенности, в частности, парадокса кучи. Была дана критика одного из последних больших философских проектов в этой области, представленного в книге Э. Бэкона [219].

Парадигмальный пример концептуальной неопределенности выглядит следующим образом: пропозициональная функция (*Bi*) «Мужчина, число волос которого равно i, лыс» является экстенсионально неопределенной. *B1* очевидно истинно, *B100000* столь же очевидно ложно, но кажется, что нет четкой границы между позитивными и негативными случаями. Иными словами, верен принцип отсутствия четких границ:

где ’ выражает операцию взятия последователя. Принимая во внимание тот факт, что верно, что *B1*, но неверно, что *B100000*, в рамках классической логики следует, что

Это следствие крайне проблематично, так как противоречит весьма правдоподобному принципу отсутствия четких границ. Более того, экстенсиональная неопределенность *B* требует наличия пограничных случаев. Иными словами, пусть *D* –– это оператор определенности (дефинитности), тогда верно, что

Однако в рамках классической логики из этого вместе с предположением о том, что 50000 –– это такое i, относительно которого *B* не определено (т. е. это пограничный случай), следует, что:

Такое следствие также является проблемой, так как вступает в противоречие с естественным пониманием определенности, в соответствии с которым ее отсутствие означает отсутствие соответствующего факта. Именно две описанные проблемы являются, как было указано, основными логическими вызовами, которые возникают в связи с неопределенностью. Показана, вместе с тем, неуспешность критики Э. Бэконом предложенных решений указанных проблем с использованием аппарата неклассических логик. Дело в том, что Э. Бэкон разделяет неклассические подходы, отвергающие закон исключенного третьего, но сохраняет транзитивность импликации и *modus ponens*. Он игнорирует то, что статус транзитивности в контекстах с неопределенностью проблематичен, ведь *modus ponens* вместе с транзитивной импликацией ведут к отрицанию принципа толерантности:

Было показано, что критика Э. Бэкона игнорирует очевидную проблематичность транзитивности (затрагивающую как закон транзитивности импликации, так и соответствующее правило вывода) при описании того, что можно назвать «мало-помалу феноменами». Отмечено, что Э. Бэкон фокусирует внимание на *modus ponens* и универсальной инстанциации, тогда как для формулировки парадокса достаточно цепочки случаев толерантности вместе с транзитивностью импликации. Таким образом, Э. Бэкон полностью игнорирует ключевую роль транзитивности в возникновении парадокса, а также наличие нетранзитивных логических систем, разработанных для решения парадокса кучи [220].

Вторая аргументативная стратегия Э. Бэкона, используемая для критики неклассических подходов, опирается на слишком сильное условие, выполнение которого требуется для того, чтобы свойство было точным. Это условие также конфликтует с более правдоподобным условием, которое может быть эксплицировано с опорой на бэконовский теоретико-типовой фреймворк неопределенности. Предполагается, что если свойство *F* точное, то:

Э. Бэкон приводит следующий пример [219: 10]: так как обладать-по-большей-мере-высотой-5.895-метров является точным свойством, то в силу правила универсальной инстанциации следует, что гора Килиманджаро обладает свойством *F* или не обладает свойством *F*. Он утверждает, что гора Килиманджаро представляет собой пограничный случай, не принимаемый во внимание сторонниками неклассических подходов, на которые направлена его критика. Показано, однако, что нелигитимность критики Э. Бэкона даже в рамках его собственного подхода к неопределенности. Причину этого он видит в том, что для Э. Бэкона первичными носителями неопределенности являются не предложения, а объекты [219: 296–298]. Таким образом, сама гора Килиманджаро как объект оказывается для него носителем неопределенности. Таким образом, собственный проект Э. Бэкона представляет собой комбинацию метафизических фактов, ведущих себя классическим образом, и эпистемического незнания этих метафизических фактов. Иными словами, на самом деле есть *i* такой, что если количество волос на голове человека больше или равно этому *i*, то человек не является лысым, однако если число волос меньше, то он лыс. Большинство эпистемических агентов просто не идентифицировали этот *i*.

Показано, что подход Э. Бэкона сталкивается с парадоксом кучи более высокого порядка, возникновение которого обусловлено воспроизведением парадокса на метауровне в силу отсутствия определенных границ, которые теперь устанавливаются относительно самой определенности в степени *n*-1 и для изначального выражения в случае, когда *n* равно 1:

Принцип отсутствия определенных границ итерируется, иными словами, для каждого n (где n –– это количество D-операторов), при этом соответствующая версия отсутствия четких границ определенно подразумевается экстенсиональной неопределенностью рассматриваемого выражения. Показано, что для возникновения парадокса достаточно принять систему KT для оператора D. Он обратил также внимание на вторую проблему программы Э. Бэкона по разрешению логических проблем неопределенности: его подход, будучи эпистемическим, приравнивает неопределенность, в частности, в пограничных случаях, к отсутствию знания. Однако такой подход ведет к смешению *explanans* и *explanandum*: суть неопределенности в пограничных случаях не может быть сведена к отсутствию знания.

Он обращается к затруднению, сформулированному Д. Эджингтон [221: 312–314], которая полагает, что в случае неопределенности примерно 50 на 50 относительно того, является содержимое кружки кофе или чаем, применение весьма правдоподобного принципа практического рассуждения, который поддерживается самим Э. Бэконом, ведет к тому, что агент должен предпочесть подобную кружку кружке кофе, если он является поклонником чая. Сам практический принцип выглядит следующим образом: если агент предпочитает <P> несовместимому <Q>, то *ceteris paribus* он должен предпочесть <P> с вероятностью 0.5 достоверному получению <Q>. Иными словами, рациональность наших желаний требует неэпистемического прочтения неопределенности.

Таким образом, показана неуспешность критики Э. Бэконом неклассических подходов к решению логических проблем, возникающих в связи с неопределенностью, но и выявил проблематичность классического подхода, которая коренится в чисто эпистемической интерпретации неопределенности.

Также намечено также решение другого парадокса, связанного с неопределенностью и сформулированного К. Райтом [222]. Аргумент К. Райта направлен против агностического тезиса о пограничных случаях, который объединяет противоположные подходы в исследованиях неопределенности –– эпистемический подход и то, что называется подходом «третьей возможности». Исследователи, убежденные в существовании «третьей возможности», воспринимают неопределенность как один из вариантов отношения произвольного объекта к некоторому свойству *F*: объект либо обладает этим свойством, либо не обладает, либо является пограничным случаем. Пропоненты эпистемического подхода рассматривают неопределенность и пограничные случаи как примеры нашего человеческого –– возможно, принципиального –– незнания о реальном положении дел, которое состоит в том, что объект либо обладает некоторым свойством, либо не обладает. В то время как в рамках эпистемического подхода естественно считать, что из наличия пограничного случая, связанного с некоторым объектом a и свойством *F*, следует агностический тезис о невозможности знания, что *Fa*, и невозможности знания, что не-*Fa*, то в случае с подходом «третьей возможности» эта двойная невозможность знания гарантируется фактивностью знания, то есть невозможностью знать то, что на самом деле не имеет места.

К. Райт предложил следующий мысленный эксперимент: пусть имеется некоторое неопределенное понятие *F* и существует континуум объектов, упорядоченных относительно этого понятия. Мы начинаем двигаться вдоль этого континуума от известных пограничных случаев (которые находятся посередине спектра) к положительным примерам *F* (например, если *F* обозначает «красный», то положительными примерами будут объекты, которые однозначно воспринимаются нами как красные). Основная проблема, поставленная К. Райтом, состоит в невозможности четко фиксировать переход от неопределенных случаев к определенным, поскольку любое промежуточное положение в спектре *F* не позволяет уверенно утверждать, что объект обладает качеством *F* или отсутствием *F* (). Однако существование высших порядков неопределенности предполагает наличие градации неопределенности. Таким образом, утверждение К. Райта состоит в том, что, постулируя неопределенность высшего порядка, мы не можем объяснить постепенность изменений в восприятии свойств объектов в процессе движения вдоль континуума сравнения.

Предложено разрешение этого парадокса, связанное с различением двух видов неопределенности, отличных от неопределенности первого порядка –– собственно неопределенности высшего порядка и кросс-порядковой неопределенности. Если К. Райт сформулировал парадокс, следующий из принятия допущений о второпорядковой неопределенности, в терминах обязательства агента не верить в некоторый факт (), то Э. Дзардини переформулирует данный парадокс в терминах невозможности знания, опираясь на то, что агент обязан не верить в то, что невозможно знать (). Формально парадокс выглядит так:



Здесь (159), (160) и (161) являются посылками, которые указывают на то, что a представляет собой пограничный случай между F и пограничным случаем F, то есть имеет место неопределенность более высокого уровня. (162) получено из (159) и (161) при помощи правила введения конъюнкции. (163) следует из (162) за счет дистрибутивности K относительно конъюнкции, а (164) получено из (163) при помощи правила введения . (165) выводится из (160) посредством удаления оператора К в силу того, что знание является фактивным. (166) представляет собой полученное противоречие, являясь конъюнктивным объединением (164) и (165).

Выявлена проблематичность посылки (2), которая связана с тем, что влечет смешение следствия и характеризации, а именно наличие этого утверждения среди посылок, порождающих противоречие, обосновано лишь в том случае, если для любых таких *a* и *F*, что верно (160), истинно, что *a* –– пограничный случай *F*. Причина этого в следующем: если истинно и даже если известно, что из *A* следует *B*, то в общем случае не следует, что из cледует . Однако такое следование имеет место, если из (160) следует *A*. Иными словами, стороннику эпистемическго подхода или исследователю, убежденному в существовании «третьей возможности», следует лишь аккуратнее формулировать свою позицию относительно пограничных случаев, избегая формулировок, эквивалентных (160).

Другое, более сильное возражение против парадокса К. Райта связано с тем, что в посылки парадокса оказывается «вшит» необоснованный переход от того, что *a* является пограничным случаем между тем, чтобы быть пограничным случаем *F* и не быть пограничным случаем *F*, к тому, что *a* является пограничным случаем между тем, чтобы быть *F* и быть пограничным случаем *F*. Строго говоря, такой переход оказывается обоснованным, только если ни один *F* не является пограничным случаем *F*. И если такой тезис не является особо проблематичным для сторонника подхода «третьей возможности», так как для него единственный способ быть *F* –– это определенно быть *F*, то для пропонента эпистемического подхода это составляет проблему.

В связи с этим возникает вопрос, может ли сторонник агностического тезиса принять, что по крайней мере в некоторых случаях описанное выше замещение имеет место, то есть подставить вместо второпорядковой или иной *n*-порядковой неопределенности кросс-порядковую неопределенность? Ответ на этот вопрос отрицательный. Такое допущение вместе с оператором определенности *Def* при любой его интерпретации приводит к возникновению парадокса, чья структура аналогична изначальной структуре парадокса К. Райта:



Здесь (167) и (168) являются посылками, (169) выводится из (167) в силу фактивности *Def*. (170) представляет собой правый конъюнкт (169), (171) получено из (167) за счет дистрибутивности *Def* относительно конъюнкции. (172) является левым конъюнктом (171), а (173) –– объединением при помощи конъюнкции (172) и (168). (174) получено из (173) благодаря дистрибутивности *Def* относительно конъюнкции. (175) представляет собой конъюнкцию (174) и (170) и обнаруживает полученное противоречие. При этом данное противоречие ведет к отрицанию (167), а не (168), так как в данном контексте (168) можно обоснованно считать истинным: если *a* –– это случай кросс-порядковой неопределенности (между тем, чтобы быть *F* и быть пограничным случаем *F*), тогда *a* встречается в промежутке между не-*F* и *F*, после определенных случаев *F*, которые уже . Следственно, в силу монотонности этого свойства на заданном промежутке a также является .

Таким образом, было установлено, что каждый сторонник агностического тезиса (для которого неопределенность подразумевает невозможность соответствующего знания) должен относиться к предположению К. Райта о том, что мы можем сказать о некотором объекте, что он представляет собой случай кросс-порядковой неопределенности, с таким же сомнением, как и к утверждению о том, что пограничный случай является позитивным, и, соответственно, рассматривать этот тезис как необязательный с точки зрения рациональности.

Также ответил на возможное возражение, которое имеет следующий вид: если на время забыть о проблеме обоснования второй посылки аргумента К. Райта, переформулированного в терминах невозможности знания, не будет ли эта посылка лишь утверждением сторонника агностического тезиса о том, что если случай является пограничным, то тезис о познаваемости неверен? И не подразумевает ли вера в агностический тезис, в частности, то, что в случае кросс-порядковой пограничности (например, свойства *F*) мы не можем знать ни о том, что *F*, ни о том, что это пограничный случай *F*? Э. Дзардини отмечает, что сама формулировка данных возражений обнаруживает скрывающуюся за ними проблему, которая состоит в переносе оператора знания от кондиционала к консеквенту. Так, сторонник агностического тезис убежден в

и, соответственно, принимает все утверждения, являющиеся частными случаями схемы:

Если два полюса неопределенности –– это обладание свойством *F* и бытие пограничным случаем *F*, то сторонник агностического тезиса принимает все случаи подстановки следующей схемы:

Соответственно, если *а* –– пограничный случай между *F* и пограничным случаем *F*, то пропонент принимает и конкретный случай подстановки указанной схемы, однако неясно, на каком основании он должен быть согласен и с более сильным тезисом:

Единственный возможный аргумент в пользу консеквента выглядит так: если мы знаем, что *a* –– кросс-порядковый пограничный случай *F*, то мы также знаем, что

так как выше было установлено (при помощи аксиомы K для знания), что

Но можем ли мы действительно знать, что a –– это кросс-порядковый пограничный случай *F*? Как показал Э. Дзардини, следствием формулировки парадокса с использованием оператор определенности *Def* является невозможность знания о том, что *а* –– случай кросс-порядковой неопределенности для *F* при принятии агностического тезиса.

Таким образом, была предложена переформулировка парадокса К. Райта в терминах невозможности знания, а затем выдвинуты два контраргумента. С одной стороны, указана нелегитимность перехода от агностического тезиса, основанного на наличии пограничных случаев, к этому же тезису, но уже характеризующему эти пограничные случаи. С другой стороны, намечено решение парадокса, использующее дихотомию неопределенности высшего порядка и кросс-порядковой неопределенности.

## 5.4 Семантические и прагматические аспекты языкового и когнитивного многообразия

Показано, что введение в семантику выполнимых (satisfiable) объектов, предложенное Ф Мольтман [223–226] является ключом к интегративному взгляду на истину и успешность перформативов.

Проведено сравнение трех вариантов базовых онтологических сущностей для референциальных теорий значения.

1. Классические пропозиции. Например, в стандартной фрегевской концепции пропозиции –– это абстрактные, интерсубъективные и вневременные сущности («мысли»).
2. Акты и события. Рассмотрен подход, берущий за основу акты (как класс событий), и показано, что он сталкивается с трудностями при объяснении референции к таким сущностям как разрешения или обязательства. Кроме того, возникает семантическое напряжение между событийными примитивами и атeмпоральными пропозициями на уровне предложения.
3. Аттитюдные и другие выполнимые объекты. Этот подход, развиваемый Ф. Молтман, предлагается как «средний путь». Показано, что концепция выполнимых объектов является естественным расширением стандартной онтологии, которое позволяет более адекватно моделировать значения высказываний в контекстах, где важную роль играют не только истинностные условия, но и иллокутивная сила, модальность, интенциональность.

Выполнимые объекты определяются как сущности, которые (а) имеют условия удовлетворения и (б) зависят от агента. В их число входят аттитюдные объекты (утверждения, суждения, обещания, надежды, решения и т. д.); модальные объекты (обязательства, разрешения, законы, способности); интенциональные объекты (поиски, долги). Среди этих категорий особый интерес представляют аттитюдные объекты, которые характеризуются как конкретные сущности, зависящие от агента и его сознания, имеющие информационную природу, ограниченные по времени своего существования, вовлеченные в причинные отношения, а также воспринимаемые. Аттитюдные объекты как вид выполнимых объектов в свою очередь подразделяются на иллокутивные объекты (утверждения, просьбы, обещания); ментальные объекты (убеждения, намерения, мысли); локутивные объекты (акты презентации содержания).

Рассмотрен пример семантического анализа в рамках этого подхода. Предложение “John made the claim that Mary is a genius” было формализовано так:

Здесь *d* –– аттитюдный объект (утверждение), производимый Джоном, а *prop* задает его условия выполнения (в данном случае –– условия истинности).

Хотя аттитюдные объекты, как и пропозиции, несут содержание и имеют условия удовлетворения, они не тождественны пропозициям, поскольку, а отличие от последних, конкретны и зависят от агента. При этом они не тождественны также и актам, поскольку, в отличие от актов, имеют информационную структуру и условия выполнения. Важным аспектом в понимании данной концепции является также различение каузальности, основанной на содержании и не основанной на содержании. Например, удивление может быть вызвано самим фактом ответа (не-контентная каузальность) или содержанием ответа (контентная каузальность).

При этом отмечена недостаточная разработанность подхода Ф. Молтман в части разграничения перформативных и неперформативных высказываний. Было указано, что одни и те же лингвистические формы (например, императивные предложения) могут иметь как перформативное, так и неперформативное прочтение. Это требует построения гибких формальных моделей, способных эксплицитно отражать такие различия в условиях разнообразия языковых употреблений. Выделено три аспекта коммуникативной ситуации с перформативами: семантика, включающая перформативный компонент значения, то есть семантически закодированную спецификацию типов аттитюдных/модальных объектов; прагматика, включающая модификацию через перформативный вывод, то есть контекстуально обусловленное приписывание перформативной силы высказыванию; параметры оценки. Был предложен интегративный взгляд на оценку перформативов на примере различения трех ипостасей слова «обещание» (promise): обещание как иллокутивный объект (речевой акт), обещание как ментальный объект (намерение в сознании говорящего), обещание как модальный объект (обязательство).

Показано, что не существует имплицитных перформативов на семантическом уровне, так как настоящий перформатив должен содержать выражение, обозначающее аттитюдный объект перформативного типа. Так, в предложении (183) слово “promise” явным образом обозначает тип иллокутивного объекта –– обещание.

1. *Hereby, I promise to solve that problem*.

При этом так же явно обозначается высказывание, кодирующее условия удовлетворения (“hereby”), и спецификация условий удовлетворения (“to solve that problem”) для модального объекта –– обязательства, порождаемого иллокутивным объектом. Однако в некоторых случаях высказывание типа (184) может быть интерпретировано как перформативное (“I promise to solve that problem”) благодаря прагматическому выводу слушающего.

1. *I’ll solve that problem*.

Разработанная фактуальная интерпретация перформативов дает возможность унифицированного подхода к моделированию значений разнотипных высказываний (утверждений, команд, обещаний и т. д.) в рамках одной формальной системы, что особенно важно для обработки естественного языка в условиях его функционального и прагматического многообразия.

Другим направлением когнитивных исследований является семантика и прагматика контекстов с обновлением. Проведено сравнение двух видов конструкций: обычных вложенных конструкций с пропозициональными глаголами, такими как “Jacob said that Bella is a vampire”, и конструкций со слифтами — “Bella is a vampire, Jacob said”. Было показано, что несмотря на схожее лексическое наполнение, конструкции ведут себя по-разному в дискурсе, различаются дистрибуцией пропозициональных глаголов и степенью коммитмента говорящего, что было подтверждено экспериментально. Эти различия были объяснены с помощью осмысления дискурсивного поведения конструкций в терминах импозиции и ассерции.

Слифты [227] являются вводной пропозициональной предикацией, которая состоит лишь из материала главной клаузы:

1. Bella, Edward said/Jacob suspects/I warn you, is a vampire.
2. Bella is a vampire, Edward said/Jacob suspects/I warn you.

С точки зрения своего пропозиционального содержания слифты похожи на обычные вложенные конструкции:

1. Edward said/Jacob suspects/I warn you that Bella is a vampire.

Можно было бы предположить, что слифты и вложенные конструкции имеют одинаковую семантику и условия истинности. Однако степень коммитмента говорящего выше в случае со слифтами, чем с обычными вложениями [228; 229], они ведут себя по-разному в рамках дискурса, а также набор пропозициональных предикатов, которые могут встречаться в слифтах более ограничен, чем в случае с обычными вложениями, в частности фактивные глаголы не встречаются в слифтах.

Различия в поведении слифтов и стандартных вложенных конструкций в рамках дискурса анализируется и формализуется в терминах двух типов обновлений дискурса, формально –– обновления модели, репрезентирующей контекст разговора.

1. [A: Who said that Sam is a vampire?] B: Dean said that Sam is a vampire. A: No, he didn’t.
2. [A: Who is a vampire?] B: Dean said that Sam is a vampire. A: No, he isn’t./#No, he didn’t.
3. B: Sam, Dean said, is a vampire. A: No, he isn’t./#No, he didn’t.

В то время как ассерции представляют собой информативные и обсуждаемые обновления, импозиции не могут прямо отрицаться и обсуждаться в дальнейшем дискурсе, представляя собой необсуждаемые обновления. Предлагается анализировать слифт в (190) и вложенную конструкцию в (189) как импозиции с эвиденциальной функцией, которые указывают на источник информации в семантически вложенной пропозиции.

Различие в поведении слифтов и вложенных конструкций в рамках дискурса сводится к следующим нескольким пунктам: во-первых, если последующий вопрос может быть обращен к пропозициональной предикации в обычных вложениях, то в случае со слифтами это оказывается невозможно. Аналогичное различие имеет место в случае с последующим отрицанием (191, 192): простое отрицание не может отрицать информацию, «спрятанную» в слифте.

1. A: US papers reported that Sam is running for political office. B: No, they didn’t.
2. A: Sam, US papers reported, is running for political office. B: (#No / √Actually), they didn’t.

Во-вторых, стандартные вложенные конструкции, но не слифты могут взаимодействовать с обсуждаемым вопросом (QUD). (188) –– это пример, в рамках которого пропозициональная предикация представляет собой ассерцию, а не импозицию, так как обсуждаемый вопрос –– это вопрос об источнике информации. Аналогичная ситуация со слифтом невозможна.

В-третьих, слифты и вложенные конструкции с предикатами пропозиционального отношения отличаются своей способностью переносить фразовую акцентуацию –– фокусную и контрастивную. В данном вопросе Эти фразовые акценты указывают на то, который из обсуждаемых вопросов актуален в данный момент дискуссии [230–233]. Иными словами, в рассматриваемых случаях фразовая акцентуация –– признак ассерции, и, следовательно, не встречается в слифтах.

1. Who said who is a vampire? JACOBFOC said BELLAFOC is a vampire.
2. Who said who is a vampire? #BELLAFOC, JACOBFOC said, is a vampire.

Для адекватного семантического описания указанных феноменов используется формальный анализ импозиций и ассерций, предложенный в [234] для анализа аппозитивов. Центральная идея данного подхода состоит в том, что импозиции и ассерции различаются тем, как именно они становятся частью общего контекста разговора. Для адекватной репрезентации того факта, что как импозиции, так и ассерции являются информативными, они формализуются как ограничения имеющегося контекстуального множества. Однако если импозиции автоматически ограничивают данное множество, то ассерции делают это лишь в том случае, если они были приняты всеми участниками разговора.

Изначальное состояние дискурса (до произнесения интересующего нас выражения) соответствует множеству пропозиций, принимаемых участниками разговора, чье пересечение и образует контекстуальное множество [235]. По мере появления импозиций и принятия ассерций контекстуальное множество уменьшается. В модели с возможными мирами это означает обновление, в ходе которого часть миров исчезает, сокращая количество совместно рассматриваемых эпистемических альтернатив.

Уже было предложено [234] моделировать контекстуальное множество при помощи пропозициональной переменной, в которой «хранится» актуальное множество пропозиций и все его непустые подмножества. Пусть актуальное множество –– это множество из всего лишь трех миров {w1, w2, w3}:

1. ⟦Bella is a vampire⟧ = {w1, w2}
2. ⟦Jacob said that Bella is a vampire⟧ = {w1, w3}

Тогда при обновлении при помощи импозиции часть миров просто исчезнут, в частности w2, если обновление имело вид ⟦Jacob said that Bella is a vampire⟧ при условии, что обсуждаемый вопрос состоял в том, является ли Белла вампиром. В случае с ассерцией процесс носит аналогичный характер, однако прежде участники разговора должны согласиться на соответствующее обновление контекстуального множества. Такой формальный анализ позволяет переосмыслить поведение слифтов и вложенных конструкций с пропозициональной предикацией в терминах обновлений и исчезающих в связи с ними миров.

Было проведено экспериментальное исследование, подтверждающее полученные теоретические результаты. Общей темой многих современных исследований является тот факт, что коммитмент говорящего по отношению к семантически вложенной пропозиции выше в слифтах. В частности, в [228] эта гипотеза проверялась при помощи теста: после слифта (197) или стандартной вложенной конструкции (198) добавлялось продолжение фразы, которое свидетельствовало о коммитменте говорящего в пользу отрицания семантически вложенной пропозиции. Если имеет место полный коммитмент говорящего по отношению к семантически вложенной пропозиции, то последующее добавление должно создать противоречие.

1. Mary assures us that John can be trusted, but I don’t trust him.
2. John, Mary assures us, can be trusted, #but I don’t trust him. [228]

Как показывают эксперименты, коммитмент говорящего выше в случае со слифтами, что вместе с их поведением в рамках дискурса указывает на импозициональный характер осуществляемого обновления.

В проведенном исследовании использовались аналогичные материалы: подобные предложения испытуемых просили оценить по шкале от -3 (противоречивое предложение) до 3 (непротиворечивое предложение) с шагом в единицу. Сотня испытуемых с Amazon Mechanical Turk, идентифицирующих себя как носителей английского языка, оценивали 21 предложение –– 12 филлеров и 9 целевых (199, 200, 201) при помощи указанной шкалы.

1. Mary said that Sam is running for political office, but he isn’t.
2. Katherine is, Dean thinks, getting married, but she isn’t.
3. The hospital will hire a surgeon, Jeremy suspects, but it won’t.

Результаты подтвердили выдвинутую гипотезу: оценки предложений со стандартными вложенными конструкциями были значительно выше, чем со слифтами. При этом различие в оценках между слифтами в конце предложения и встроенными посередине статистически незначительно.

Еще одним подтверждением предложенной теоретической рамки анализа указанных конструкций стала сравнительно большая ограниченность предикатов пропозиционального отношения, способных быть частью слифта. Предикаты пропозиционального отношения, например, *думать, знать, бояться*, информируют слушающего о субъективном отношении говорящего или иного субъекта к пропозициональному аргументу данного предиката. Так, например, в предложении «Я счастлив, что Дин уехал из города» предикат «счастлив» выражает эмпоциональное отношение говорящему к вложенной пропозиции и в то же время свидетельствует о том, что говорящий считает ее истинной.

На основании их синтаксических и семантических свойств пропозициональные глаголы делятся на две группы –– ассертивные и неассертивные [236]. Класс ассертивных предикатов содержит нефактивные и импликативные предикаты, в то время как класс неассертивных предикатов включает в себя нефактивные и фактивные глаголы. Одно из синтаксических свойств ассертивных предикатов, которым не обладают неассертивные, –– это способность фигурировать в слифтах. Дж. Хупер определяет основное различие этих предикатов как семантическое, а именно: дело в уверенности в истинности соответствующей пропозиции [236: 95]. Опираясь на работу Дж. Хупер, предложены следующие обобщения. Для того, чтобы фигурировать в слифтах, пропозициональный предикат: 1) должен не быть фактивным, т. е. он не должен конвенционально подразумевать истинность пропозиционального аргумента; 2) не должен подразумевать, что говорящий не верит в истинность пропозиционального аргумента.

Таким образом, было представлено экспериментальное подтверждение того факта, что степень коммитмента говорящего по отношению к семантически вложенной пропозиции выше в случае со слифтами. Было описано поведение слифтов и обычных вложений в рамках дискурса в терминах импозиции и ассерции. Было установлено, что в то время как часть с пропозициональной предикацией может быть ассерцией или импозицией в случае со стандартными вложенными конструкциями, она может быть лишь импозицией в случае со слифтами. Обратное истинно для семантически вложенной пропозиции.

Основным теоретическим вкладом стало предложение анализировать дискурсы со слифтами и обычными вложенными конструкциями с пропозициональной предикацией при помощи двух видов обновлений контекста: импозиций [234] и ассерций [237]. Было показано, что понятия импозиции и ассерции обеспечивают связь между условиями истинности, поведением в рамках дискурса и намерениями говорящего, позволяя объяснить: 1) различия в коммитменте говорящего, выявленные в ходе экспериментального исследования; 2) различия в дискурсивном поведении слифтов и обычных вложенных конструкций, содержащих пропозициональные предикаты; 3) дистрибуцию в них самих предикатов пропозиционального отношения.

Исследование эвиденциальных конструкций на материале коми-зырянского языка содержит аргументацию в пользу их немодальной интерпретации. Также была предложена формальная семантика, опирающаяся на анализ эвиденциальных конструкций в терминах обновлений контекста, в частности импозиции и ассерции. Были получены результаты, уточняющие соотношение между различными видами обновления (импозиция/ассерция) в теоретической лингвистике и примерами конкретных эвиденциальных конструкций в речевой практике.

Эвиденциальные предложения имеют сложную структуру: они выражают две пропозиции. Одна –– основная –– пропозиция передает определенное количество информации, вторая –– эвиденциальная –– указывает на источник этой информации. При этом эвиденциальная система коми-зырянского языка чувствительна по отношению к тому времени, когда говорящий узнал содержание пропозиции, то есть совпало ли оно с протеканием события, а не к тому, был ли он его свидетелем. Эвиденциальное содержание передается в коми-зырянском языке с помощью морфем прошедшего времени. Есть две такие морфемы: так называемая морфема первого прошедшего времени -i (PST1) и так называемая морфема второго прошедшего времени -ema (PST2) [238–246]. Обе морфемы используются для описания событий, произошедших до момента говорения, и имеется хорошо известный эвиденциальный контраст между ними [240; 244; 245]:



|  |  |  |  |
| --- | --- | --- | --- |
| Van'a | kerka-yn | abu | uz'-ema |
| Vanja | house-ESS | NEG | sleep-PST2 |
| '(Кажется/Мне сказали, что) Ваня не спал дома.' | | | |



|  |  |  |  |
| --- | --- | --- | --- |
| Van'a | kerka-yn | iz | uz' |
| Vanja | house-ESS | NEG.PST1.3 | sleep |
| '(Обычно: как я сам засвидетельствовал,) Ваня не спал дома.' | | | |

В (202) говорящий использует второе прошедшее время, чтобы передать тот факт, что он сам не был свидетелем того, что Ваня не спал дома, ему сказали об этом или он вывел этот факт из имеющейся у него информации. Предложение типа (203) используется, когда говорящий лично наблюдал тот факт, о котором он говорит. Интерес для исследования представляют контексты, где могут быть использованы обе конструкции, то есть говорящий осуществляет выбор между ними. Различие первого и второго прошедшего времени в коми-зырянском языке сопоставляется с известным делением на прямые и косвенные источники информации, где последние также делятся на инференциальные и репортативные [247; 248].

Среди исследователей имеет место согласие по поводу наличия коммитмента говорящего по отношению к эвиденциальной пропозиции, но ведутся споры о степени коммитмента по отношению к основной пропозиции. Идея, стоящая за этим, сводится к следующему: если говорящий сам не был свидетелем того, что некоторый факт имел место, то не может иметь места полный коммитмент с его стороны по отношению к основной пропозиции. Некоторые утверждают, что коммитмент говорящего и вовсе не имеет места [249]; другие говорят о некоторой его степени [250–252]; есть варианты анализа эвиденцальных конструкций, которые подразумевают полный коммитмент говорящего [253–255].

Два соревнующихся подхода к коммитменту говорящего по отношению к основной пропозиции могут быть представлены следующим образом [255]:

1. Модальный подход: Evid(p) = <weak.modal(p); indirect.evid(speaker; p)>
2. Немодальный подход: Evid(p) = <p; indirect.evid(speaker; p)>

Оба подхода соглашаются в том, что имеет место эвиденциальная импликация, согласно которой говорящий обладает непрямым свидетельством в пользу основной пропозиции. Различие состоит в отношении к основной пропозиции. Согласно немодальному подходу, основная пропозиция фигурирует в эвиденциальной конструкции в немодифицированном виде, в то время как сторонники модального подхода утверждают, что основная пропозиция имеет связанную с ней модальную силу. Было показано, что вопрос о коммитменте говорящего по отношению к основной пропозиции отчасти сводится к вопросу о границе между семантикой и прагматикой. Сторонники модального подхода кодируют наблюдаемое ослабление при помощи семантики, в то время как сторонники немодального подхода объясняют его прагматически. В связи с этим основная стратегия аргументации в пользу немодального подхода –– показать отсутствие ослабления в некоторых контекстах, а также возможность подобной его вариативности средствами прагматики, что и делается на материале коми-зырянского языка.

Во-первых, говорящий не может напрямую отрицать свой коммитмент по отношению к основной пропозиции в последующем дискурсе. Так, использование второго прошедшего времени не может сопровождаться утверждениями типа “Я не верю, что p”:



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ponm-ys | pyšj-ema, | (\*no | me | syd' | og | dumajt) |
| dog.OBL-POSS3SG | run.away-PST2 | but | I | so | NEG.PRS.1 | think |
| 'Видимо, собака убежала, но я так не думаю.' | | | | | | |

Во-вторых, ослабление исчезает, когда имеется вложенная конструкция с глаголом отношения. В (207) в дополнении глагола стоит простая, не модифицированная пропозиция:



|  |  |  |  |
| --- | --- | --- | --- |
| Maša | vis'tal-is | pet'a | gәtras'-ema |
| Masha | say-PST1.3SG | Petja | marry-PST2 |
| a. ‘Masha said that Petja got married.’ b. ‘#Masha said that Petja might/must have gotten married.’ | | | |

В-третьих, степень коммитмента говорящего крайне зависима от контекста: она может отличаться в зависимости от надежности источника информации. Например, степень коммитмента в (207) может варьироваться в зависимости от авторитета Маши и ее отношения к говорящему. В некоторых случаях наличие эвиденциального показателя может восприниматься даже как усиление ассертивной силы.

В-четвертых, несмотря на вариативность степени коммитмента, носители языка в соответствие с максимой количества [256] предпочитают не говорить предложения, относительно которых у них нет достаточного количества свидетельств. Это утверждение подтверждается экспериментом, в рамках которого носителей языка просили сделать ставку на истинность основной пропозиции. Испытуемые ставили всю сумму.

И в-пятых, ослабление сложно описать при помощи одного модального оператора в связи с высокой степенью его контекстуальной зависимости, что ведет к нежелательным последствиям, в частности сторонникам модального подхода приходится расширять репертуар модальных операторов или же возможные употребления изначального оператора [250; 257; 258].

Что касается поведения в рамках дискурса, несмотря на то что обе пропозиции информативны, то есть вводят в дискурс новую информацию, они различаются в своих проективных свойствах и во взаимодействии с последующим дискурсом. Было выделено три основных различия. Что касается проективных свойств, то, например, в случае с отрицанием (208) основная, а не эвиденциальная пропозиция является целью отрицания:



|  |  |  |  |
| --- | --- | --- | --- |
| Van'a | kerka-yn | abu | uz'-ema |
| Vanja | house-ESS | NEG | sleep-PST2 |
| '(Как я могу вывести/Как мне сказали/Видимо,) Ваня не спал дома.' | | | |

Второе различие проявляется в связи с возможностью взаимодействия между пропозицией и последующим дискурсом: в отличие от основной пропозиции эвиденциальная пропозиция не взаимодействует напрямую с последующим дискурсом, например, она не может отрицаться так, как показано в (210):



|  |  |  |  |
| --- | --- | --- | --- |
| kod-ke | bura | okyšt-ema | tene |
| who-INDEF | good | kiss-PST2 | you.ACC |
| 'Кто-то хорошо тебя целовал.' | | | |



|  |  |  |
| --- | --- | --- |
| sy | abu | syd' |
| that | NEG | so |
| 'Это не так. (Меня никто не целовал. / #Ты не вывел, что меня целовали.)' | | |

Еще одно различие заключается в способности пропозиции отвечать на обсуждаемый вопрос (question under discussion) [231]. Предположение, лежащее в основе изучения дискурса, заключается в том, что дискурс организован в соответствии с некоторой целью. В любой момент разговора что-то обсуждается, оспаривается и т. д. Если думать об этом как об эксплицитином или имплицитном вопросе, на который пытаются ответить собеседники, то мы получим обсуждаемый вопрос (QUD). Анализ данных показал, что основная, а не эвиденциальная пропозиция отвечает на обсуждаемый вопрос:

* 1. QUD: Кто не спал прошлой ночью?

1. QUD: #Откуда ты знаешь, что Ваня не спал дома прошлой ночью?

|  |  |  |  |
| --- | --- | --- | --- |
| Van'a | kerka-yn | abu | uz'-ema |
| Vanja | house-ESS | NEG | sleep-PST2 |
| '(Как я могу вывести/Как мне сказали/Видимо) Ваня не спал дома.' | | | |

Иными словами, содержание эвиденицальной пропозиции оказывается менее важным –– в том смысле, что говорящий собирается обсуждать не её, а основную пропозицию. Был выдвинут тезис о том, что такая асимметрия может быть лучше всего описана при помощи разделения обновления на два вида: ассерцию и импозицию. В то время как ассерции –– это информативные пропозиции, чье содержание предлагается обсуждать и оспаривать в рамках дискурса, импозиции –– это пропозиции, чье отрицание не подразумевается, они также обладают проективными свойствами, отсутствующими в случае ассерции [259].

С опорой на аналогичный анализ для болгарского языка [255], была предложена следующая формальная семантика:



|  |  |
| --- | --- |
| ponm-ys | pyšj-ema |
| dog.OBL-POSS3SG | run.away-PST2 |

1. ПРЕДЛОЖЕНИЕ: основная пропозиция: the dog ran away:
2. ИМПОЗИЦИЯ: эвиденциальная пропозиция: говорящий позже узнал, что собака убежала:

* где *spu* –– говорящий, *τu* –– время произнесения, *p* –– основная пропозиция.”)

Предложение «Собака убежала» формализуется следующим образом: есть сущность *x*, и *x* –– это собака, и есть событие e, которое заключается в том, что x убегает во время *τ(e)*, которое предшествует времени говорения *τu*. Имеют место ассерция и импозиция. Импозиция формализуется так: есть другое событие *e1*, в ходе которого говорящий *spu* приобрел информацию, являющуюся содержанием основной пропозиции *p*, и время убегания собаки *τ(e)* предшествует времени приобретения информации *τ(e1)*.

Однако в контекстах, где первое прошедшее соревнуется со вторым прошедшим, оно используется лишь в случае наличия прямого свидетельства.



|  |  |
| --- | --- |
| ponm-ys | pyšj-is |
| dog.OBL-POSS3SG | run.away-PST1.3SG |

АССЕРЦИЯ: собака убежала:

Предлагается следующий анализ (215). Данная пропозиция анализируется так же, как и в случае с (212), однако здесь ассерция не сопровождается импозицией. Вывод о наличии прямого свидетельства оказывается речевой импликатурой, которая выводится, так как говорящий мог использовать второе прошедшее, но не стал.

Таким образом, поведение основной и эвиденциальной пропозиций в предложениях, где употребляется второе прошедшее время: было сопоставлено двум видам обновления –– ассерции и импозиции. Предложенный анализ позволяет предсказывать выбор первого или второго прошедшего времени в контекстах, где такой выбор возможен, а также классифицировать вывод о наличии прямого свидетельства в случае выбора первого прошедшего времени как речевую импликатуру. Была также проанализирована степень коммитмента говорящего по отношению к основной пропозиции: данные свидетельствуют в пользу полного коммитмента говорящего. Была предложена формальная семантика предложений, содержащих морфемы первого и второго прошедших времен коми-зырянского языка. Предложенный анализ связывает вместе эвиденциальное значение, темпоральное значение, коммитмент говорящего и поведение морфем в рамках дискурса.

# 6. Корпусная и компьютерная лингвистика

## 6.1 RCaucTile и PyCaucTile: плиточные карты для визуализации нахско-дагестанских языков

### 6.1.1 Введение

Плиточная карта –– это вид упрощённой картографической визуализации, на которой географические регионы изображаются соразмерными «плитками», чаще всего квадратами или шестиугольниками. Элементы располагаются на условной сетке координат, сохраняющей примерное взаимное расположение объектов. В настоящей работе предлагается адаптация методологии плиточных карт для визуализации языковых данных. Результатом исследования являются два программных пакета на R и Python, позволяющие создавать такие карты для восточнокавказских языков.

### 6.1.2 Лингвистическое картографирование

Лингвистическое картографирование широко используется в диалектологии и типологии, являясь устоявшимся инструментом для визуализации ареального распределения данных. C. R. Luebbering выделяет несколько видов лингвистических карт – точечные, ареальные, экстраполированные и гибридные, в зависимости от того, представлены ли языки отдельными точками, сплошными областями, плавными градиентами на основе нескольких точек или их комбинацией [260]. В типологических атласах, включая ставший стандартом Всемирный атлас языковых структур WALS [261], наиболее распространены точечные визуализации, в которых каждому языку соответствует одна координата, цветовое кодирование которой отражает значение признака.

В Типологическом атласе языков Дагестана [7], разработанном в Лаборатории, помимо точечных карт (см. ([рисунок 6.1](#fig-10))) также используются экстраполированные визуализации, где один язык представлен несколькими точками, соответствующими населенным пунктам, в которых он распространен (см. ([рисунок 6.2](#fig-11))). Оба варианта карт создаются при помощи программного пакета lingtypology, разработанного Г. А. Морозом [262].

Среди ограничений лингвистических карт отмечают условность и невозможность отражения реальной ситуации многоязычия при представлении языков как единичными точками, так и полигонами [263]. В работах R. K. Macaulay и C. R. Luebbering обсуждаются проблемы читаемости визуализаций: при избытке деталей и высокой языковой плотности карты становятся перегруженными, а распределение признака менее очевидным для восприятия [260; 264]. Эта проблема является наиболее актуальной для регионов с большим языковым разнообразием, к числу которых относится Восточный Кавказ, где на сравнительно компактной территории распространено более 50 языков из нахско-дагестанской, индоевропейской и тюркской языковых семей.

|  |
| --- |
| Рисунок 6.1 – Типы утренних приветствий (точечная карта) [265]. |

|  |
| --- |
| Рисунок 6.2 – Типы утренних приветствий (экстраполированная карта) [265]. |

#### 6.1.2.1 Плиточные карты

Плиточные карты рассматриваются в литературе в качестве альтернативы традиционным картографическим визуализациям в ситуациях, когда географическая детализация не является необходимой или даже затрудняет восприятие данных. Как отмечается в исследовании W. Meulemans и др., с возрастанием сложности данных карты, на которых сохраняется ландшафт и точное расположение объектов, становятся менее читаемыми из-за неизбежного совпадения элементов, что требует применения схематизации [266]. В регионах с высокой плотностью объектов, согласно M. Loecher и K. Ropkins, визуализации легко становятся перегруженными или страдают от наложения элементов, даже при работе с наборами данных умеренного размера [267].

В плиточных картограммах каждый географический регион представляется в виде образующих единую сетку единообразных плиток идентичного или пропорционального некоторому статистическому признаку размера [268]. Преимущество такого подхода заключается в том, что все регионы при корректном подборе размера сохраняют свою различимость: крупные регионы не преобладают в восприятии над малыми. Кроме того, строгая структура сетки систематизирует расположение элементов и облегчает сравнение данных между регионами [268], что делает плиточные карты наиболее подходящими для случаев, когда для визуализации важнее распределение признака среди объектов, чем их географическое положение.

Процесс создания плиточной карты предполагает трансформацию координат, при которой акцент смещается с точных расстояний между объектами на сохранение их смежности, взаимного расположения и глобальной позиции в рамках общей формы территории [269]. Несмотря на потерю исходной формы и ландшафта, данный тип визуализации сохраняет узнаваемость для зрителя за счет ментальной карты: как объясняется в работе G. McNeill и S. A. Hale, у зрителей в сознании обычно присутствует сложившийся ментальный образ географической зоны, что снижает необходимость в детализации карты и позволяет эффективно использовать схематизированное представление [268].

#### 6.1.2.2 Методология

Насколько нам известно, плиточные карты ранее не применялись для визуализации языковых данных, что требует адаптации традиционных подходов. В частности, представляется нерелевантным автоматическое определение координат на основе существующих алгоритмов построения плиточных карт, таких как recmap [270], в силу ограничений таких инструментов и специфики лингвистических данных. В отличие от административных зон с четкими границами, разделяющими территорию на непрерывные участки, языковые ареалы выделяются более условно и распределяются неравномерно, образуя плотные скопления и пустые участки.

Условные координаты на сетке были определены для языков вручную на основе информации о расположении населенных пунктов, соответствующих языкам, с использованием базы данных сел Дагестана и отредактированы с учетом особенностей распределения языков в регионе. При разработке соблюдались три основных принципа, сформулированных для плиточных карт в работе Meulemans и др.: сохранение смежности и взаимного расположения объектов, узнаваемость общей конфигурации и компактность размещения [266].

Разработанные пакеты включают 56 языков, распространенных на территории Восточного Кавказа, из классификации, предложенной в Типологическом атласе языков Дагестана [7]. Цветовое кодирование на основе генеалогии также заимствовано из Атласа для обеспечения преемственности визуализаций.

### 6.1.3 Результаты

#### 6.1.3.1 Техническая реализация

Отличительной особенностью разработанного инструмента является его доступность на двух языках программирования – R и Python, что расширяет возможности использования в различных исследовательских средах. Пакеты спроектированы единообразно, совпадая по структуре входных и выходных данных, названиям основных функций и логике работы. RCaucTile реализован на основе библиотеки ggplot2, являющейся стандартом визуализации данных в R. PyCaucTile использует plotnine – Python-реализацию ggplot2, позволяя создавать идентичные визуализации при помощи схожего синтаксиса.

#### 6.1.3.2 Структура данных

В основе пакетов лежит встроенный датасет с информацией о 56 языках из Типологического атласа языков Дагестана. Набор данных включает следующие столбцы, многие из которых совпадают с данными из Атласа:

* language –– название языка;
* branch –– ветвь языковой семьи;
* family –– семья;
* glottocode –– код языка из ресурса Glottolog;
* language\_color –– цветовой код для отдельного языка;
* branch\_color –– цветовой код для ветви;
* x, y –– условные координаты на сетке;
* abbreviation –– сокращенное название;
* примеры лингвистических признаков из Атласа (morning\_greetings, consonant\_inventory\_size).

Помимо доступной внутри пакетов информации по двум типологическим признакам, поддерживается работа с пользовательскими данными, что является одной из главных возможностей для исследователей. Для визуализации собственных признаков необходимо подготовить таблицу, содержащую столбец с названиями языков и столбец с соответствующими им значениями признака.

#### 6.1.3.3 Типы карт

Плиточные карты, создаваемые при помощи RCaucTile и PyCaucTile, разделяются на два основных типа в зависимости от визуализируемой информации. Цвет прямоугольника («плитки»), обозначающего язык, может отражать генеалогическую принадлежность языка или значение определенного типологического признака. Базовый шаблон карты, вызываемый ключевой функцией пакета без аргументов, включает языки на сетке координат с цветовым кодированием. Как можно заметить на ([рисунок 6.3](#fig-12)), цвета сгруппированы по языковым ветвям: нахские языки представлены коричневым цветом, андийские –– синим, лезгинская ветвь –– зеленым и так далее.

|  |
| --- |
| Рисунок 6.3 – Плиточная карта с цветовым кодированием на основе генеалогии. |

Данный шаблон задает распределение языков по умолчанию и служит основой для карт с отображением типологических признаков. В зависимости от характера признака а различаются два типа карт. Пакеты автоматически определяют характер входных данных и применяют соответствующую схему визуализации.

Категориальные карты используют дискретную цветовую схему для качественных признаков, где каждое значение кодируется отдельным цветом. На ([рисунок 6.4](#fig-13)) представлена плиточная карта на основе встроенного признака из исследования C. Naccarato и S. Verhees, классифицирующего языки по стратегиям утренних приветствий [265]. Карта демонстрирует ареальное распределение трех типов: вопросы о пробуждении, приветствия формата *доброе утро*, а также присутствие обеих стратегий. Языки, для которых данные по визуализируемому признаку отсутствуют, отображаются светло-серым цветом.

|  |
| --- |
| Рисунок 6.4 – Типы утренних приветствий (карта для категориального признака) [265]. |

В картах на основе числовых признаков применяется градиентная шкала для количественных данных, где интенсивность цвета связана с величиной значения.

В ([рисунок 6.5](#fig-14)) отображено распределение количества согласных фонем в инвентаре согласно данным из Типологического атласа языков Дагестана [271]. Градиентная визуализация позволяет быстро оценить ареальное распределение и выявляет языки с наиболее развитыми консонантными системами, представленные светлыми оттенками.

|  |
| --- |
| Рисунок 6.5 – Размер инвентаря согласных (карта для числового признака) [271]. |

Пакеты предоставляют возможность настройки визуализаций для адаптации под формат и цели исследования. Значения признаков могут отображаться непосредственно на плитках рядом с названиями языков, как показано на ([рисунок 6.5](#fig-14)), что особенно эффективно для количественных данных и позволяет быстро оценивать и сравнивать значения. Доступно изменение параметров цветовых палитр, заголовков и легенды, а также модификация языкового шаблона с исключением и сменой названий выбранных языков, что является актуальным при работе с другими генеалогическими классификациями.

### 6.1.4 Заключение

Плиточные карты рассматриваются как альтернативный подход к отображению лингвистических признаков, но не являются полноценной заменой традиционным типологическим картам. Такой метод смещает акцент с географической точности на распределение признака среди языков, может применяться для микротипологических исследований со сравнительным анализом. Плиточные карты представляются актуальными для территорий с высокой языковой плотностью, где традиционные карты сталкиваются с проблемой совпадения элементов и перегруженности. Это создает возможность адаптации разработанной методологии для картографирования других регионов с большим языковым разнообразием.

## 6.2 Сравнительный анализ особенностей выражения риторических и синтаксических структур, характерных ошибок и вариативности обозначения причинно-следственных связей в текстах обучающихся и экспертов

### 6.2.1 Вариативность в использовании маркеров причинно-следственных связей в текстах учащихся и экспертов

В качестве материала для исследования маркеров причинно-следственных связей использовались корпусы ERAC и PECLAP , собранные в научно-учебной лаборатории учебных корпусов НИУ ВШЭ. Первый корпус включает в себя научные статьи по шести дисциплинам, опубликованные в ведущих журналах в период с 2013 по 2020 гг., входящих в базу цитирования Scopus. Второй корпус представляет собой коллекцию проектов ВКР по тем же дисциплинам, написанных на английском языке студентами четвертого курса НИУ ВШЭ в 2015–2020 гг. Информация о размере корпусов представлена в [таблица 6.1](#tbl-corpora_lab_1).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.1 – Характеристики учебного и экспертного корпусов   | Дисциплина | Количество\_текстов | Количество\_токенов | empty | Средняя\_длина\_текста | | --- | --- | --- | --- | --- | | Экспертный корпус | | | |  | | Экономика | 56 | 528,542 |  | 9,438 | | Менеджмент | 61 | 639,772 |  | 10,488 | | Компьютерные науки | 77 | 654,252 |  | 8,497 | | Политология | 73 | 624,795 |  | 8,559 | | Юриспруденция | 91 | 688,763 |  | 7,569 | | История | 65 | 604,378 |  | 9,298 | | Всего | 423 | 3,740,502 |  | 8,843 | | Экономика | 68 | 142,182 |  | 2,091 | | Менеджмент | 58 | 123,259 |  | 2,125 | | Компьютерные науки | 59 | 103,888 |  | 1,761 | | Политология | 29 | 52,726 |  | 1,818 | | Юриспруденция | 77 | 169,174 |  | 2,197 | | История | 43 | 85,072 |  | 1,978 | | Всего | 334 | 676,301 |  | 2,025 | |

Языковые единицы, отобранные для анализа причинно-следственных связей, были взяты из Louvain EAP Dictionary (LEAD), веб-словаря по английскому языку для академических целей для неносителей английского языка [272–274], см. [таблица 6.2](#tbl-corpora_lab_2).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.2 – Маркеры в учебном и экспертном корпусах   | Category | Существительные | Союзы | Глаголы | Предлоги | | --- | --- | --- | --- | --- | | Причина | *cause* | *because* | *contribute* | *because of* | |  | *factor* | *given that* |  | *due to* | | *ground* | *since* |  | *following* | | *origin* |  |  | *given* | | *reason* |  |  | *in view of* | | *root* |  |  | *on account of* | | *source* |  |  | *owing to* | | Следствие | *consequence* | *so* | *result* | *accordingly* | |  | *effect* | *so that* | *lead* | *consequently* | | *implication* |  |  | *hence* | | *outcome* |  |  | *naturally* | | *result* |  |  | *thereby* | |  |  |  | *therefore* | |  |  |  | *thus* | |

Проведенное исследование выявило системные различия в использовании маркеров причинно-следственных связей в академических текстах учащихся и профессиональных авторов. Статистический анализ с применением критерия хи-квадрат показал, что в целом маркеры причины значимо чаще употребляются студентами, причем наиболее существенные различия наблюдаются в экономике. Однако за этим количественным преобладанием скрывается качественное различие: учащиеся демонстрируют склонность к ограниченному, шаблонному использованию узкого набора лексем (например, существительного «factor» и предлога «due to» в составе клишированной конструкции «due to the fact that»), в то время как эксперты пользуются более широким репертуаром синонимичных средств. Что касается маркеров следствия, то их общая частотность оказалась выше в экспертном корпусе, хотя данная тенденция варьировалась в зависимости от дисциплины. Качественный анализ выявил и функциональные различия: эксперты часто используют существительные следствия во введениях и обзорах литературы для формулирования целей и обобщения существующих исследований, тогда как студенты склонны применять их преимущественно в выводах для описания результатов. Кроме того, учащиеся злоупотребляют базовыми союзами, такими как «so», в ущерб более конвенциональным вариантам, что указывает на недостаточное владение жанровыми и стилистическими нормами.

Для комплексной оценки вклада различных лингвистических признаков была построена биномиальная логистическая регрессионная модель, предсказывающая принадлежность текста к учебному или экспертному корпусу. Финальная модель, прошедшая оптимизацию, включила пять наиболее значимых предикторов и продемонстрировала высокую дискриминативную способность. Анализ выявил устойчивую обратную связь: более высокая частотность маркеров следствия (наречий, глаголов, существительных и союзов) и союзов причины достоверно ассоциирована с профессиональными текстами. Ранжирование переменных методом случайного леса подтвердило, что наибольший вклад в различение групп вносят наречия и глаголы следствия (см. [рисунок 6.6](#fig-corp1)).

|  |
| --- |
| Рисунок 6.6 – Коэффициенты при линейной регрессии для частей речи при предсказании уровня текста |

Полученные результаты подчеркивают необходимость целенаправленного обучения вариативному и стилистически адекватному использованию маркеров причинно-следственных связей в курсах академического письма

### 6.2.2 Вариативность в согласовании подлежащего и сказуемого в академических текстах носителей и неносителей английского языка по различным дисциплинам

Настоящее исследование посвящено анализу вариативности в согласовании по числу между подлежащим и сказуемым в профессиональных академических текстах на английском языке. Фокус работы сосредоточен на четырех проблемных классах существительных, для которых характерно рассогласование между грамматической и понятийной числовыми характеристиками: собирательные существительные (group, team), неисчисляемые существительные, обозначающие абстрактные понятия (research), исчисляемые существительные с формой множественного числа, лишенной стандартной морфологической маркировки (data, media), и существительные в единственном числе, оканчивающиеся на -*s* (series, dynamics). Как отмечают Байбер и др. [275] и Сельсе-Мурсия и Ларсен-Фриман [276], подобные существительные демонстрируют нестабильные паттерны согласования, что может быть обусловлено синтаксической сложностью, семантической множественностью и отсутствием четких перцептивных референтов [277; 278].

Актуальность работы определяется тем, что, несмотря на существующие исследования вариативности согласования в речи изучающих язык и в различных вариантах английского языка [279; 280], данный феномен остается малоизученным в контексте профессионального академического письма, где соблюдение дисциплинарных и жанровых конвенций имеет paramount importance. Более того, предыдущие работы фрагментированы и не дают целостной картины влияния лингвистического бэкграунда автора и дисциплинарной принадлежности текста. Таким образом, данное исследование ставило перед собой цель восполнить этот пробел, отвечая на три ключевых вопроса: во-первых, существуют ли различия в паттернах согласования у носителей и неносителей английского языка; во-вторых, наблюдается ли вариативность согласования между точными и гуманитарными дисциплинами; и в-третьих, какие синтаксические факторы определяют вариативность в выборе числа глагола.

Для достижения поставленных целей сотрудниками пермской части НУЛ учебных корпусов был составлен специализированный корпус научных статей общим объемом 4,4 миллиона слов, опубликованных в рецензируемых журналах за период с 2022 по 2024 год. Корпус был стратифицирован по двум основным параметрам: дисциплинарная семья и лингвистический бэкграунд автора. Авторы классифицировались как носители или неносители английского языка на основе анализа их имен, аффилиаций и образовательной истории, что является общепринятой практикой в подобных корпусных исследованиях [281; 282]. Для исключения смешанного влияния, из анализа были исключены статьи с соавторством, где авторы принадлежали к разным группам. Каждая дисциплинарная ячейка содержала равное количество текстов от носителей и неносителей, что обеспечило сбалансированность дизайна. Информация о размере корпуса представлена в [таблица 6.3](#tbl-corpora_lab_3).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.3 – Состав корпусов по отраслям знаний   | Disciplinary\_family | Field | Discipline | Number\_of\_texts\_NS | Number\_of\_texts\_NNS | Number\_of\_words\_NS | Number\_of\_words\_NNS | | --- | --- | --- | --- | --- | --- | --- | | Soft sciences | Arts and humanities | History | 24 | 24 | 236,244 | 203,209 | | Philosophy | 24 | 24 | 208,178 | 204,579 | | Literature | 24 | 24 | 183,320 | 179,303 | | *Total* | 72 | 72 | 627,742 | 587,091 | | Social sciences | Linguistics | 24 | 24 | 223,294 | 225,798 | | Law | 24 | 24 | 359,812 | 290,629 | | Political science | 24 | 24 | 206,931 | 200,570 | | *Total* | 72 | 72 | 790,037 | 716,997 | |  | ***Soft total*** | 144 | 144 | 1,417,779 | 1,304,088 | | Hard sciences | Life sciences | Chemistry | 24 | 24 | 144,373 | 143,022 | | Biology | 24 | 24 | 158,863 | 154,966 | | Medicine | 24 | 24 | 113,725 | 111,624 | | *Total* | 72 | 72 | 416,961 | 409,612 | | Physical sciences | Astronomy | 24 | 24 | 135,594 | 121,985 | | Geology | 24 | 24 | 172,028 | 149,607 | | Physics | 24 | 24 | 176,049 | 127,601 | | *Total* | 72 | 72 | 483,671 | 399,193 | |  | ***Hard total*** | 144 | 144 | 900,632 | 808,805 | |  | ***Corpus total*** | 576 |  | 4,431,304 |  | |

Отобранные для анализа существительные соответствовали критериям частотности в академическом дискурсе и потенциальной вариативности в согласовании. Корпус был обработан с помощью программного обеспечения AntConc для первоначального поиска контекстов, после чего был проведен тщательный ручной анализ для идентификации случаев согласования и аннотирования синтаксических признаков. Для каждого случая фиксировались число глагола, а также синтаксические переменные: наличие детерминатива, количество препозитивных модификаторов, наличие of-фразы и дистанция между подлежащим и сказуемым.

Для статистической обработки данных использовалась комбинация методов: для анализа бинарных предикторов применялся расчет φ-коэффициента с интерпретацией эффекта по Коэну [283], для непрерывных переменных использовался U-критерий Манна-Уитни, а для многомерного анализа влияния нескольких факторов на согласование с существительным data была применена биномиальная логистическая регрессия с пошаговым исключением незначимых переменных в R.

Анализ выявил системные, хотя и не всегда статистически значимые, различия между группами авторов. Носители языка продемонстрировали бóльшую вариативность, особенно в «мягких» науках, где в их текстах по литературе было зафиксировано 63% случаев pluralia tantum с существительным group. Это согласуется с выводами Фернандес-Пены [279] о большей гибкости носителей в использовании понятийного согласования. Однако φ-коэффициент не показал значимой связи, что указывает на сложность данного фактора. Для существительных media и data была обнаружена контрастная картина: носители преимущественно использовали media с единственным числом, в то время как неносители — с множественным. Этот результат противоречит данным Байбера и др. [275], но подтверждает тезис Моралло [284] о реклассификации media как singularia tantum носителями языка. Напротив, для data неносители чаще, чем носители, использовали форму единственного числа. Паттерны для существительных на -s (series, dynamics) оказались разнонаправленными. Носители чаще использовали series с глаголом в единственном числе, тогда как неносители демонстрировали обратную тенденцию. Для dynamics, напротив, неносители чаще выбирали единственное число. Это противоречит исследованию Аль-Шаера [285], где носители последовательнее трактовали подобные существительные как сингулярные. Таким образом, носители языка действительно демонстрируют более вариативное и, вероятно, более идиосинкразическое использование согласования, что может объясняться опорой на семантические и контекстуальные факторы, в то время как неносители чаще следуют формальным, иногда упрощенным, грамматическим правилам [277]. Однако отсутствие статистической значимости в большинстве случаев указывает на то, что фактор лингвистического бэкграунда не является определяющим сам по себе и его влияние опосредовано другими переменными.

Дисциплинарный фактор оказался статистически значимым и последовательным для трех из четырех классов существительных. Гуманитарные науки значимо чаще использовали множественное число с коллективными существительными по сравнению с точными. Аналогичная, хотя и менее выраженная, тенденция была выявлена для существительного data. Наиболее сильная дисциплинарная ассоциация была обнаружена для существительных на -s: гуманитарные науки последовательно предпочитали множественное число.

Результаты показали, что синтаксический контекст оказывает избирательное, но важное влияние на выбор числа глагола. Наиболее сильные эффекты были зафиксированы для коллективных существительных. Примеры с множественным числом глагола имели значительно большую дистанцию между подлежащим и сказуемым. Это подтверждает гипотезу Левина [286] о том, что увеличение синтаксической сложности и расстояния усиливает влияние семантической множественности коллективных существительных. Наличие of-фразы продемонстрировало сильную положительную связь с множественным числом. Конструкции типа a group of researchers семантически смещают фокус на множественность элементов, что и отражается в выборе глагола. Для существительного data слабая, но значимая связь показала, что большее число препозитивных модификаторов ассоциируется с единственным числом. Это можно объяснить тем, что развернутая номинативная группа чаще воспринимается как единый, специфицированный блок информации. Наличие детерминатива было значимо связано с множественным числом. Возможно, определенный артикль the указывает на конкретный, ограниченный набор данных, который легче концептуализировать как совокупность дискретных единиц. Ни один из анализируемых синтаксических факторов не показал статистически значимой связи с согласованием для существительных на -s.

Критическим результатом стала попытка построить комплексную логистическую регрессионную модель для предсказания согласования с data, которая показала низкую прогностическую силу. Анализ методом случайных лесов также подтвердил незначительный вклад включенных переменных. Это ключевой вывод, который указывает на то, что синтаксические факторы сами по себе не могут адекватно объяснить вариативность. Авторы справедливо предполагают, что для понимания механизмов согласования необходимым является учет семантических и дискурсивных факторов, которые не были включены в текущую модель.

Проведенное исследование демонстрирует, что вариативность в согласовании подлежащего и сказуемого в академическом дискурсе является сложным, многогранным явлением, определяемым взаимодействием нескольких факторов. Лингвистический бэкграунд автора создает определенные тенденции, но не является решающим детерминантом, что подчеркивает важность профессиональной академической социализации для неносителей языка. Дисциплинарная принадлежность оказывается более устойчивым и значимым предиктором, отражая глубинные эпистемологические и риторические различия между точными и гуманитарными науками. Синтаксический контекст, особенно дистанция между подлежащим и сказуемым и наличие of-фразы для коллективных существительных, играет важную, но ограниченную роль, что особенно ярко видно на примере неудачной регрессионной модели для data.

Практическая значимость исследования заключается в том, что его результаты могут быть использованы для совершенствования методики обучения академическому письму, обеспечивая более тонкое и контекстуально обусловленное объяснение правил согласования для изучающих английский язык как иностранный.

### 6.2.3 Новые корпусы и авторская позиция в текстах, сгенерированных ИИ (Туляков Д.С.)

#### 6.2.3.1 Корпус устных защит проектов защиты ВКР на английском студентов-лингвистов НИУ ВШЭ

Для дальнейшего исследования авторской позиции в и особенностей L2 академической речи был создан корпус устной речи защит проектов ВКР на английском языке студентов НИУ ВШЭ-Пермь образовательной программы “Иностранные языки и межкультурная коммуникация в бизнесе”. Защиты были записаны и траскрибированы, а затем проверены стажерами-исследователями НУЛ. Корпус полностью анонимизирован и содержит метаданные об оценке, полученной студентом за презентацию. Корпус насчитывает 61 текст и приблизительно 32 тыс. слов.

Корпус был создан с учетом действующих правовых норм, и в данный момент он на итоговой стадии оформления как РИД (номер заявки 0000-2726; номер РИД в СУРИД 8.0146-2025).

В данный момент стажеры-исследователи продолжают работать над корпусом и вручную аннотируют тексты корпуса по риторическим шагам (например, Orientation, Rationale, Framework). Корпус также апробируется в исследовании стажера-исследователя Федоровой А.Д. “Анализ маркеров дискурса в защитах проектов ВКР студентов”, представленном на корпусном симпозиуме “Корпусная лингвистика и дискурс-анализ”:

“Изучение маркеров дискурса привлекает многих исследователей. Однако небольшое количество работ посвящено исследованию этого явления в устной академической речи. Данная работа представляет собой анализ использования маркеров дискурса в корпусе устных презентаций проектов выпускных квалификационных работ студентов четвертого курса образовательной программы «Иностранные языки и межкультурная коммуникация в бизнесе». В рамках данной работы предполагается представить результаты анализа частотности маркеров дискурса и их функции. Анализ частотности был проведен с помощью корпусного менеджера AntConc. Также была использована функциональная классификация Лоретты Фанг и Рональда Картера (2007). Группа маркеров, указывающая на причинно-следственные связи в предложении или тексте, стала самой многочисленной и составила 70%. Второй по величине группой стали маркеры, отвечающие за управление структурой текста, — 23,8%. Число когнитивных маркеров (сигнализирующих о процессе мышления говорящего или поясняющих его мысли), составило 6 %, самой малочисленной категорией стала группа маркеров, указывающих на процесс взаимодействия говорящего и слушающего — 0,2%. Самыми высокочастотными маркерами стали also (148), but (57) и now (51). Данные результаты показывают, что студенты в своих защитах больше используют маркеры, относящиеся к неакадемическому дискурсу. Академические маркеры дискурса, напротив, использовались реже. Так маркер however использовался тридцать два раза, finally — двадцать два раза, moreover — девять. Таким образом можно сделать вывод о том, что академическая речь студентов обладает характерными чертами разговорного дискурса.”

#### 6.2.3.2 Корпус сгенерированных ИИ аннотаций исследовательских статей и сопоставительное исследование маркеров значимости в аннотациях, написанных человеком и сгенерированных ИИ

**Введение.** Появление и стремительное развитие инструментов на базе генеративного искусственного интеллекта (ИИ) существенно повлияло на практики написания текстов как студентов, так и исследователей. Согласно опросу, проведённому в 2024 году, 86% студентов по всему миру используют ИИ-инструменты для выполнения различных задач, включая проверку грамматики, резюмирование и перефразирование документов, а также создание черновиков научных работ (Digital Education Council, 2024); более свежие отчёты свидетельствуют о том, что использование ИИ среди студентов становится ещё более распространённым. Исследователи также всё чаще применяют ИИ для повышения качества научного письма, увеличения эффективности и поддержки саморегулируемого обучения, а всё большее число научных журналов допускает заявленное использование ИИ в своих правилах подачи материалов (например, Elsevier и APA). Этот сдвиг вызвал всплеск публикаций, посвящённых этическим аспектам написания текстов с помощью ИИ, включая вопросы авторства, научной добросовестности и прозрачности. Учитывая темпы внедрения, можно предположить, что использование генеративных ИИ-инструментов в академической среде, особенно в научном письме, будет продолжать расти.

В этом контексте важно понять, какие задачи академического письма генеративный ИИ может поддерживать без ущерба для целостности и качества научных текстов. Один из способов получить такое понимание — изучить особенности ИИ-сгенерированных текстов путём их сравнения с аналогичными текстами, написанными без помощи ИИ. Поскольку большие языковые модели (LLMs), лежащие в основе генеративных ИИ-инструментов, обучаются на наборах данных, включающих разнообразные текстовые источники, можно ожидать, что тексты, созданные ИИ (например, части научной статьи), будут демонстрировать характерные черты релевантных текстов из обучающего корпуса. Однако конкретные способы, с помощью которых ИИ-сгенерированные академические тексты сходны или отличаются от текстов, написанных людьми в рамках определённого жанра, заслуживают внимательного изучения, поскольку это позволяет лучше понять сильные и слабые стороны ИИ как помощника в научном письме.

Среди множества характеристик академического письма, которые могут быть затронуты использованием ИИ, особенно важным является промоциональный язык. Под промоциональным языком понимаются лингвистические выборы и риторические стратегии, направленные на формирование положительного восприятия научного вклада со стороны аудитории [287]. Недавние исследования выявили тенденцию к увеличению объёма промоционального языка, используемого исследователями в статьях и грантовых заявках. В качестве причин указываются высокая конкуренция за внимание читателей в условиях информационной перегрузки, а также борьба за финансирование и возможность публикации в престижных журналах. Исследования показывают, что с течением времени частота использования «раздувающих» лексических единиц (например, таких прилагательных, как critical, important или original) увеличивается как внутри отдельных дисциплин — медицины [287], прикладной лингвистики [288; 289], биомедицинских наук [290], — так и в академии в целом [291; 292]. Эта тенденция особенно заметна в аннотациях к научным статьям и грантовым заявкам, где авторы не только резюмируют свои исследования, но и активно их продвигают [293]. Такой подход оказывается эффективным: статьи, в аннотациях которых используется больше промоционального языка, получают больше внимания и чаще цитируются.

Хотя предыдущие исследования изучали промоциональный язык в аннотациях, написанных людьми, мало известно о том, как в этом отношении выглядят аннотации, созданные ИИ. Цель настоящего исследования — выяснить, подчеркивают ли ИИ-сгенерированные аннотации к академическим статьям важность исследования в той же степени, что и аннотации, написанные людьми. Поскольку резюмирование является одной из самых распространённых задач, выполняемых с помощью генеративного ИИ, можно предположить, что многие исследователи, использующие ИИ в качестве помощника, применяют функцию резюмирования при работе над аннотациями. Если внедрение ИИ в академическую среду будет продолжаться, как прогнозируется, всё большее число аннотаций будет хотя бы частично сгенерировано. И авторам, использующим ИИ для написания аннотаций, и читателям этих аннотаций будет полезно знать, следует ли ожидать от ИИ дальнейшего преувеличения значимости исследования (которая, как показывают исследования, уже может быть завышена).

**Сравнительные исследования академических текстов, написанных людьми и ИИ.** Большинство сравнительных работ сосредоточены на студенческих эссе. Исследования показывают, что рецензенты часто не могут отличить тексты, написанные студентами, от тех, что созданы ИИ, а иногда даже оценивают ИИ-эссе выше. Это свидетельствует о поверхностном сходстве.

Тем не менее ИИ-эссе отличаются ограниченным набором стратегий вовлечения, меньшим количеством выражений позиции и риторической гибкости. Они демонстрируют слабую авторскую позицию, низкое разнообразие лексики и упрощённую структуру.

В области научных текстов различия также заметны. Например, эксперты в медицине легко распознают ИИ-сгенерированные статьи и оценивают их ниже, тогда как лингвисты демонстрируют низкую точность.

ИИ-аннотации следуют стандартной структуре, но отличаются шаблонностью, меньшей вариативностью и бедностью терминологии. Они проще, короче и менее точны, хотя читаются легче.

Интересно, что по некоторым параметрам ИИ-аннотации приближаются к человеческим , но используют меньше абстрактного и убедительного языка, что противоречит другим данным.

В целом, ИИ способен имитировать академические нормы, но уступает в нюансах, точности и адаптивности. Настоящее исследование дополняет литературу, анализируя, насколько ИИ-аннотации отражают акцент на важности, характерный для текстов, написанных людьми.

Промоциональный язык и маркеры важности в академическом письме. Несмотря на распространённое мнение об объективности научного письма, исследования показывают, что академический дискурс часто носит риторический и промоциональный характер. Учёные [294; 295] подчёркивают влияние институциональных и дисциплинарных факторов, а также стремление убедить читателя в ценности исследования, особенно в условиях конкуренции за финансирование и публикации.

Аннотации — жанр, где промоциональный язык особенно заметен. Они не только резюмируют содержание, но и продвигают работу. Исследования фиксируют рост использования таких выражений в аннотациях, особенно в грантовых заявках и медицинских статьях, что отражает тенденции «экономики внимания».

Анализ промоционального языка ведётся как риторически, так и корпусно. Риторические исследования изучают оценочные выражения, маркеры позиции и самопрезентации [288], часто опираясь на жанровую модель Swales. Корпусные исследования фокусируются на частотности лексем типа important, novel, critical и их распределении по дисциплинам [290; 292].

Особую роль играют маркеры важности — лексико-грамматические средства, подчёркивающие значимость информации. Они активно изучаются в устных академических жанрах и аннотациях, где входят в категорию «hyping». Их частотность растёт, особенно в гуманитарных и социальных науках.

Таким образом, промоциональный язык и маркеры важности — ключевые элементы академического письма, и важно понять, как ИИ воспроизводит или трансформирует эти практики.

Теоретическая модель «позиции» (stance). В данном исследовании маркеры важности рассматриваются как проявления авторской позиции. Согласно [275], позиция — это выражение чувств, оценок и установок автора. В академическом дискурсе такие выражения играют важную роль в аргументации, выстраивании доверия и взаимодействии с читателем.

В модели [296] маркеры важности входят в группу оценочных маркеров, наряду с хеджами, усилителями и самоупоминаниями. Они выражают аффективную составляющую позиции, демонстрируя значимость исследуемого объекта.

Модель [275] классифицирует маркеры важности как выражения оценочной позиции, реализуемые через наречия, глаголы, прилагательные, существительные с дополнениями и модальные глаголы. Эти маркеры — лексические единицы из семантической области «важности», хорошо подходящие для корпусного анализа по частотности и контексту употребления.

**Данные.** Для данного исследования были собраны два корпуса: корпус аннотаций, написанных людьми (HUMabs) и корпус соответствующих аннотаций, сгенерированных ИИ (AIabs). Чтобы обеспечить разнообразие, HUMabs включает 218 аннотаций к статьям, опубликованным в 32 ведущих журналах по 8 дисциплинам (4 естественнонаучные и 4 гуманитарные) в период с 2016 по 2020 год — до широкого распространения генеративного ИИ.

Корпус AIabs был создан с помощью DeepSeek. Все статьи были предварительно обработаны: сохранены только заголовок и основное содержание, без авторов, аннотаций, ссылок и дополнительных разделов. Каждая статья загружалась в отдельный чат с запросом: «Напиши аннотацию для этой статьи для подачи в [название журнала]». Все аннотации AIabs были сгенерированы в августе 2025 года содержаться в [таблица 6.4](#tbl-corpora_lab_4):

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.4 – Характеристики HUMabs and AIabs   | Discipline | HUMabs | | AIabs | | | | --- | --- | --- | --- | --- | --- | | **Discipline (hard/soft)** | **No files** | **Median words** | **Total words** | **Median words** | **Total words** | | Chemistry (h) | 34 | 144.0 | 4,850 | 150.5 | 4,906 | | Physics (h) | 34 | 120.0 | 4,377 | 119.0 | 4,299 | | Mathematics (h) | 28 | 106.5 | 3,101 | 115.5 | 3,722 | | Engineering (h) | 34 | 176.5 | 6,006 | 181.0 | 5,912 | | Business (s) | 20 | 143.0 | 3,086 | 146.5 | 2,991 | | Linguistics (s) | 22 | 165.0 | 3,885 | 158.5 | 3,582 | | History (s) | 21 | 152.0 | 3,572 | 150.0 | 3,520 | | Political Science (s) | 25 | 152.0 | 3,903 | 167.0 | 4,193 | | **All corpus** | **218** | **147.0** | **32,780** | **150.5** | **33,125** | |

**Отбор маркеров важности.** Список маркеров важности, использованных в исследовании изучавших их употребление в корпусе из 120 презентаций «Three Minute Thesis». Как и в их исследовании, маркеры были классифицированы по частям речи, что соответствует нашему подходу к ним как к лексическим проявлениям авторской позиции. Это также позволяет анализировать типичные грамматические конструкции с маркерами (например, существительное + предлог), в соответствии с моделью [275].

Для расширения списка были учтены дополнительные исследования [296], а также данные из тезаурусов Merriam-Webster и Dictionary.com. В список включались только те единицы, которые выражают именно важность, исключая маркеры новизны или полезности. После исключения лексем, не встречающихся в корпусах, итоговый список составил 72 маркера: 10 глаголов, 20 существительных, 32 прилагательных и 10 наречий.

**Аналитическая процедура.** Для анализа был использован Python-скрипт, который проводил лемматизацию текстов и рассчитывал абсолютные и нормализованные частоты всех четырёх категорий маркеров важности. Далее частоты сравнивались для выявления статистически значимых различий — как в целом, так и по категориям и отдельным леммам. В качестве метрик использовались логарифмическое правдоподобие (LL) и процентное отклонение (%DIFF — сравнение AIabs с HUMabs).

Дополнительно анализ проводился отдельно для естественнонаучных и гуманитарных дисциплин, чтобы выявить, в каких областях различия в частоте маркеров выражены сильнее.

**Результаты: Сравнение маркеров важности в AIabs и HUMabs по категориям и типам дисциплин.** Сравнение частотности маркеров важности в аннотациях, созданных ИИ (AIabs), и написанных людьми (HUMabs), как в целом, так и по категориям, показало, что в AIabs их значительно больше. Особенно выражено это в категориях прилагательных и наречий, тогда как различия в глаголах и существительных статистически незначимы. Наиболее заметное отличие — прилагательные, которые встречаются в AIabs более чем в три раза чаще, см. [таблица 6.5](#tbl-corpora_lab_5a).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.5 – Нормализованные частоты (слов на миллион) маркеров важности по категориям во всех дисциплинах   | **importance marker** | **AIabs** | **HUMabs** | **%DIFF** | **LL** | **p** | | --- | --- | --- | --- | --- | --- | | adjectives | 8935.85 | 5369.128 | 66.43 | 29.60 | < 0.001 | | adverbs | 1962.26 | 610.12 | 221.62 | 25.62 | < 0.001 | | nouns | 4769.81 | 4331.91 | 10.11 | 0.4 | n.s. | | verbs | 1388.68 | 1494.814 | -7.1 | 0.13 | n.s | | all | 17056.6 | 11805.98 | 44.47 | 31.64 | < 0.001 | |

При рассмотрении дисциплин отдельно (естественные и гуманитарные науки) наблюдается аналогичная тенденция, но с различиями в масштабах, см. [таблица 6.6](#tbl-corpora_lab_5b):

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.6 – Нормализованные частоты (слов на миллион) маркеров важности по категориям в естественных и гуманитарных дисциплинах   | importance marker | AIabs | HUMabs | %DIFF | LL | p | | --- | --- | --- | --- | --- | --- | | in hard disciplines |  |  |  |  |  | | adjectives | 8,068.37 | 4,636.10 | 74.03 | 30.92 | < 0.001 | | adverbs | 1,645.52 | 218.17 | 654.23 | 40.73 | < 0.001 | | nouns | 3,397.21 | 3,163.52 | 7.39 | 0.27 | n.s. | | verbs | 1,751.69 | 1,799.96 | -2.68 | 0.02 | n.s. | | all | 14,862.78 | 9,817.83 | 51.39 | 34.21 | < 0.001 | | in soft disciplines |  |  |  |  |  | | adjectives | 10,079.80 | 6,299.32 | 60.01 | 29.00 | < 0.001 | | adverbs | 2,379.95 | 1,107.57 | 114.88 | 15.65 | < 0.001 | | nouns | 6,579.87 | 5,814.76 | 13.16 | 1.50 | n.s. | | verbs | 909.98 | 1,107.57 | -17.84 | 0.60 | n.s. | | all | 19,879.60 | 14,329.23 | 38.73 | 29.80 | < 0.001 | |

В обеих группах дисциплин статистически значимые различия между AIabs и HUMabs наблюдаются только в категориях наречий, прилагательных и общего количества маркеров. Однако степень различий варьируется. Это поднимает вопрос о том, насколько точно ИИ отражает дисциплинарные особенности использования маркеров важности.

Так, в HUMabs гуманитарных дисциплин наречия используются значительно чаще, чем в естественных (1108 против 218 слов на миллион; %DIFF = +408%; LL = 10.86; p < 0.001). В AIabs это различие сглаживается и становится статистически незначимым (2380 против 1646; %DIFF = +45%; LL = 2.26; p = 0.13).

Аналогичная, но менее выраженная картина наблюдается для прилагательных: в HUMabs гуманитарных дисциплин они встречаются чаще (6299 против 4636; %DIFF = +36%; LL = 4.44; p = 0.035), но в AIabs разница снижается до +25% и становится лишь погранично значимой (10080 против 8068; LL = 3.9; p = 0.048). Это свидетельствует о том, что ИИ снижает дисциплинарную вариативность в использовании маркеров важности.

Если рассматривать все маркеры вместе, различия сохраняются: в HUMabs гуманитарных дисциплин их частота выше (14329 против 9818; %DIFF = +46%; LL = 14.38; p < 0.001), а в AIabs — тоже значимо выше, но с меньшим разрывом (19880 против 14863; %DIFF = +34%; LL = 10.44; p = 0.001). Таким образом, ИИ частично нивелирует дисциплинарные различия, но не устраняет их полностью.

**Результаты: Распределение маркеров важности в AIabs и HUMabs.** Поскольку предыдущие результаты указывают на «гомогенизирующий» эффект ИИ в использовании маркеров важности, в этом разделе рассматривается, проявляется ли он также в распределении маркеров по отдельным аннотациям, см. [таблица 6.7](#tbl-corpora_lab_6).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.7 – Распределение и разнообразие маркеров важности в AIabs и HUMabs   |  | AIabs | HUMabs | AIabs(soft) | HUMabs(soft) | AIabs(hard) | HUMabs(hard) | | --- | --- | --- | --- | --- | --- | --- | | mean | 2.59 | 1.78 | 3.24 | 2.35 | 2.15 | 1.38 | | st. dev. | 2.2 | 1.92 | 2.21 | 2.02 | 2.08 | 1.74 | | CV | 0.8474 | 1.08 | 0.68 | 0.86 | 0.97 | 1.26 | | ZMA | 43/218 | 64/218 | 9/88 | 16/88 | 34/130 | 48/130 | | ZMA, % | 19.7 | 29.4 | 10.2 | 18.2 | 26.2 | 36.9 | |

Данные показывают, что коэффициент вариации в AIabs ниже, чем в HUMabs, что означает более равномерное распределение маркеров в аннотациях, созданных ИИ. Это подтверждается тем, что доля аннотаций без маркеров важности в AIabs всегда ниже.

Для оценки разнообразия маркеров был рассчитан коэффициент типа к токену (TTR), основанный на количестве уникальных лемм и общем числе их употреблений, см. [таблица 6.8](#tbl-corpora_lab_7):

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.8 – Коэффициент типа к токену (TTR) маркеров важности в AIabs и HUMabs   |  | AIabs | HUMabs | AIabs(soft) | HUMabs(soft) | AIabs(hard) | HUMabs(hard) | | --- | --- | --- | --- | --- | --- | --- | | type | 64 | 63 | 50 | 52 | 51 | 42 | | token | 565 | 387 | 285 | 207 | 280 | 180 | | TTR | 0.11 | 0.16 | 0.18 | 0.25 | 0.18 | 0.23 | |

TTR оказался выше в HUMabs, что свидетельствует о большем лексическом разнообразии: люди используют более широкий набор маркеров важности, чем ИИ.

**Промежуточные выводы.** Анализ частотности маркеров важности показал, что аннотации, созданные ИИ, содержат их значительно больше, чем тексты, написанные людьми — особенно в категориях прилагательных и наречий. Это свидетельствует о склонности ИИ усиливать риторическую значимость исследования. При этом различия между дисциплинами (гуманитарные и естественные науки) в ИИ-аннотациях выражены слабее, чем в аннотациях, написанных людьми, что указывает на тенденцию к нивелированию стилевых особенностей разных научных областей.

Дополнительный анализ показал, что ИИ не только распределяет маркеры важности более равномерно по аннотациям, но и использует их менее разнообразно. В то время как авторы-человеки демонстрируют более широкий лексический репертуар, ИИ опирается на ограниченный набор выражений. Это может свидетельствовать о формализованном подходе ИИ к академическому стилю, что важно учитывать при использовании генеративных моделей в научной коммуникации.

### 6.2.4 Разбор языка фейковых новостей на английском языке: От методов убеждения к прагматическим механизмам

#### 6.2.4.1 Постановка задачи исследования фейковых новостей

Данное исследование вносит вклад в изучение языка фейковых новостей, сосредоточившись на пропаганде как типе фейковых новостей, который поддерживает и продвигает определённую сторону или точку зрения путём намеренного упущения важной контекстуальной информации [297] фейковые новости типа III в их типологии). С точки зрения языка, пропаганда обладает интересными убедительными качествами: упущение истины идёт рука об руку со стратегическими лингвистическими ходами, призванными привлечь внимание читателя к предпочитаемым автором нарративам и политическим идеологиям. Зачастую это делается неявно, а не явно, путём использования базовых прагматических механизмов и творческого расширения смыслового потенциала языковых средств для манипулирования аудиторией. Цель нашего исследования — точно определить, как выглядят эти лингвистические стратегии и как они способствуют убедительному дискурсу в пропаганде. В частности, наше исследование направляется двумя исследовательскими вопросами.

1. Как используются и распространены методы убеждения в англоязычных пропагандистских текстах?
2. Каковы основные лингвистические и прагматические механизмы каждого из методов?

Для достижения наших целей мы возьмем за отправную точку существующую модель убеждения, разработанную Пискорским [298], которая была успешно применена в работе Да Сан Мартино и др. [299] для детального исследования убеждения в пропаганде. Хотя эта модель в первую очередь предназначена для применения в компьютерных и социальных науках, здесь мы используем ее в качестве первого шага к целенаправленному лингвистическому анализу методов убеждения, где характер этих методов (например, «Апелляция к авторитету», «Выдвижение сомнений», «Размахивание флагом») поможет нам определить конкретные ресурсы, на которых следует сосредоточиться (например, ресурсы для атрибуции в «Апелляции к авторитету»), одновременно позволяя дополнить модель дальнейшими исследованиями в области лингвистики и прагматики. Таким образом, мы можем добавить к нашей системе ещё один аналитический уровень, более серьёзно относящийся к языку, способствуя тем самым усилиям по автоматическому выявлению фейковых новостей и повышению уровня медиаграмотности, где язык играет ключевую роль. Данные для исследования взяты из корпуса Fakespeak-ENG, а именно из тщательно отобранной выборки пропагандистских текстов, представляющих два различных жанра: социальные сети, с одной стороны, и новостные статьи и записи в блогах, с другой.

#### 6.2.4.2 Пропаганда и убеждение

По мнению Джоуэтт и О’Доннелл [300: 7], «пропаганда — это преднамеренная, систематическая попытка формировать восприятие, манипулировать когнитивными процессами и направлять поведение для достижения реакции, способствующей достижению желаемых целей пропагандиста», — определение, которое соответствует нашему собственному пониманию пропаганды, данному во введении. В этом смысле она схожа с убеждением — фактически, эти два термина использовались взаимозаменяемо в литературе и в повседневной речи — за одним важным исключением: цель [300: 2]. В то время как пропаганда направлена на распространение идеологии среди отдельного человека или группы людей, убеждение — это динамичный межличностный процесс [301], призванный удовлетворить потребности как убеждающего, так и убеждаемого [302: 10]. В частности, убеждающий может стремиться либо изменить точку зрения собеседника, либо укрепить существующие убеждения [301], используя языковые и коммуникативные ресурсы [303: 28]. Исходя из этого, мы можем разработать иерархическую структуру отношений, где пропаганда — это форма убеждения, неизбежно наследующая от последнего межличностное измерение, где язык играет ключевую роль, сохраняя при этом свою манипулятивную цель. В лингвистике механизмы и функции убеждения традиционно изучались в рамках критического дискурс-анализа с точки зрения социальной власти и привилегированного положения институтов по отношению к рядовым гражданам (например, [304], динамика власти значительно изменилась в результате более широкого доступа к таким информационным каналам, как платформы социальных сетей, блоги и сайты новостей (ср. [302: 4]). Признавая контекстно-зависимую и динамичную природу власти, Партингтон и Тейлор переносят фокус на язык как на «основное средство достижения и осуществления власти», особенно в современных либеральных демократиях, которые и рассматриваются в настоящем исследовании. Таким образом, они предлагают полезную схему, позволяющую проводить систематический анализ языка убеждения с точки зрения пяти основных методов и стратегий: авторитета, сравнения и контраста, проблемы-решения, модели «гипотеза-доказательство-объяснение» и ассоциации, которые имеют некоторое сходство с схемой, используемой в данном исследовании. Влияние этих методов и стратегий на коммуникацию носит перлокутивный характер [305], стремясь изменить убеждения и точку зрения собеседника, что делает убеждение весьма привлекательным явлением для прагматических исследований.

Следует, однако, отметить, что конкретные языковые средства, используемые для выражения убеждения, весьма разнообразны и не поддаются какой-либо строгой классификации. Тем не менее предыдущие исследования выявили ряд поверхностных форм, которые с достаточной вероятностью могут использоваться для убеждения, среди которых — усилительные наречия, модальные глаголы возможности и предсказания, личные местоимения, числа, побудительные наклонения, слова, выражающие эмоции, дейксис и вопросы (например, [306]; [307]). Ни одна из них не связана однозначно с убеждением, но все они требуют «цепочки выводов всё более абстрактных понятий, ведущей от языковой формы к функции убеждения» [301: 4], где, например, усилители (например, very) могут представлять автора как знающего поставщика информации. Выявление методов убеждения в данном исследовании гарантирует, что обнаруженные нами лингвистические особенности действительно связаны с функцией убеждения, в дополнение к рассмотрению изначально убедительной природы пропаганды, описанной выше.

Одним из аспектов убеждения, не кодифицированным в языке и, следовательно, требующим ещё более глубокого вывода, является имплицитность, реализуемая в коммуникации, например, посредством разговорной импликатуры, пресуппозиции и образных выражений. Импликатура, в частности, является полезным типом имплицитной коммуникации, когда передаваемое не совсем соответствует действительности. В своём экспериментальном исследовании Маццарелла и др. обнаружили, что, поскольку импликатура менее обязывает, человек, который подразумевает, а не говорит или предполагает, воспринимается как менее заслуживающий порицания и, таким образом, может сохранить часть своей репутации надёжного источника, если утверждение окажется не совсем верным. Действительно, именно импликатура и пресуппозиция выделяются как средства манипулятивного дискурса, например, в кликбейтных заголовках.

Аналогичным образом, [308: 2] исследовали различные триггеры пресуппозиций, такие как глаголы изменения состояния (например, *dobbiamo costruire il nostro futuro* «мы должны построить наше будущее»), как манипулятивные стратегии в итальянской и французской политической коммуникации, как созданные человеком, так и искусственным интеллектом. Потенциальная опасность образных выражений, таких как метафора и гипербола (вспомним лавины иммигрантов, где задействованы оба), а также сарказма [309], заключается в их огромной убедительной и эмоциональной силе в (политическом) дискурсе [310]. Более того, по мнению Клариджа [311: 370], и метафора, и гипербола представляют собой «близость к обману или потенциальную возможность обмана» в силу своей небуквальной природы, «не давая прямого или просто декодируемого представления о положении дел» [311: 370]. В настоящее время нам не известны какие-либо комплексные исследования, посвященные разнообразию явных или неявных языковых стратегий, используемых в целях убеждения и манипуляции в пропаганде. Да Сан Мартино и др. [299: 5637] указывают на «ошибочные рассуждения и/или эмоциональные призывы» как на отличительные черты пропаганды; в настоящем исследовании та же концепция убеждения используется для описания всего спектра лингвистических и прагматических механизмов пропаганды, ориентированной на англоязычную аудиторию.

#### 6.2.4.3 Данные и методы

Данные взяты из корпуса Fakespeak-ENG, обширной коллекции фейковых и подлинных новостей, созданных в период с 2019 по 2024 год. Исходный корпус содержит 644 809 слов, хранящихся в 4112 текстах, каждый из которых относится к одному из трех жанров: социальные сети, новости и блоги, пресс-релизы. Для выявления пропагандистских текстов среди множества типов фейковых новостей в корпусе мы использовали метки, присвоенные фактчекерами. Метка, наиболее точно соответствовавшая нашему определению пропаганды и, следовательно, послужившая отправной точкой для построения нашей выборки, была «Полуправда». Согласно PolitiFact, утверждения, подпадающие под эту метку, «частично точны», но не содержат «важных деталей» или «вырывают информацию из контекста». Другими словами, они в целом соответствуют ключевому аспекту пропаганды в нашем исследовании — упущению.

Поскольку не все случаи использования метки соответствуют нашему точному определению пропаганды, потребовалась дополнительная фильтрация. Для этого мы подвергли каждый текст тщательному ручному исследованию на основе контрольного списка критериев. Текст считался пропагандистским, если он отвечал следующим критериям: (i) утверждение было в целом правдивым, (ii) он продвигал определенную сторону или точку зрения (но не обязательно точку зрения политической организации) и (iii) его целью было принести пользу определенному лицу или группе.

Окончательная выборка в итоге состояла из 171 текста (16 996 слов), из которых 156 текстов (7950 слов) были постами в социальных сетях, а 15 текстов (9046 слов) — новостными статьями и постами в блогах (пресс-релизы были исключены из-за малого количества), что является достаточно большой выборкой для исследовательского анализа убеждения в двух основных жанрах, посредством которых обычно распространяется пропаганда.

Мы аннотировали пропагандистские тексты с целью убеждения, следуя рекомендациям Пискорского. Из наших результатов видно, что существует семь основных категорий методов убеждения: «Атака на репутацию», «Оправдание», «Отвлечение», «Упрощение», «Призыв», «Манипулятивная формулировка» и «Другие».

Аннотации были выполнены разметчиками, которые следовали первоначальным рекомендациям и используя либо MAXQDA (VERBI Software 2023), либо UAM CorpusTool. Все внесенные в процессе корректировки обсуждались совместно для обеспечения единообразия аннотаций. Третий автор выступал в роли посредника, сравнивая аннотации друг с другом и принимая решающее решение в случае расхождений. В новостных статьях и блогах мы аннотировали только основной текст, следуя работе [297].

После выявления (под)приёмов убеждения мы подвергли примеры дальнейшему прагматическому анализу, чтобы выявить конкретные языковые средства, с помощью которых убеждение выражается в пропаганде. В рамках прагматической теории мы опирались главным образом на понятия разговорной импликатуры и пресуппозиции [256], уделяя особое внимание любым потенциально ложным выводам, сделанным читателем на основе знаний о мире, или если предполагаемое содержание фактически не соответствовало общей позиции мейнстримной западной политики. В рамках когнитивно-функциональной лингвистики мы уделили особое внимание роли образного языка, основываясь на наших знаниях о его убедительной и эмоциональной силе в коммуникации, касающейся политики и социальных проблем. Кроме того, учитывая характер жанров, включенных в исследование, особенно постов в социальных сетях, мы в значительной степени опирались на роль оценки и эмоций, опираясь на теории оценки и позиции [275], как на вербальном, так и на типографическом уровнях коммуникации. При этом точные средства выражения убеждения в каждом случае во многом зависели от конкретных (суб)приемов (например, атрибуции в «Апелляции к авторитету», дейксиса в «Апелляции ко времени» и гиперболы в «Преувеличении»).

#### 6.2.4.4 Результаты и обсуждение

В нашей выборке пропагандистских текстов обнаружено 1018 случаев убеждения. Манипулятивная формулировка является наиболее распространенным методом, с 349 случаями, за ним следуют «Оправдание» (210), «Атака на репутацию» (158), «Другое» (112), «Призыв» (103), «Упрощение» (52) и «Отвлечение» (34). Относительно большая доля случаев, которые не удалось классифицировать в соответствии с первоначальными рекомендациями (т. е. категория «Другое»), вызывает вопросы относительно эффективности и охвата рекомендаций. При 248 случаях «Влиятельный язык» встречается почти в три раза чаще, чем следующий по частоте приём — «Апелляция к авторитету» (96). Однако оба они полезны для массового убеждения, характерного для пропаганды, апеллируя либо к эмоциям людей (влиятельный язык), либо к их доверию посредством цитирования источников (апелляция к авторитету). Кроме того, заслуживают упоминания «Апелляция ко времени» (76), «Вызывание сомнений» (64) и «Апелляция к страху и предубеждениям» (52). В следующих подразделах наш анализ прагматики убеждения будет сосредоточен на этих пяти приёмах. Сравнение двух жанров исследования показывает, что в сообщениях в социальных сетях встречается больше примеров убеждения, чем в новостных статьях и блогах (585 против 433), хотя первые составили несколько меньшую долю данных. Ещё одно интересное отличие заключается в том, что в сообщениях в социальных сетях большее предпочтение отдаётся атакам на репутацию и случаям категории «Другое». В частности, пользователи социальных сетей, по-видимому, чаще атакуют конкретных людей, а именно их качества (вызывание сомнений) и непоследовательность действий (призыв к лицемерию); более частая встречаемость категории «Другое» может объясняться более неформальным, субъективным стилем данного жанра социальных сетей.

Наиболее распространёнными подтехниками в нашем корпусе являются «Нагруженный язык», «Апелляция к авторитету», «Апелляция ко времени», «Вызывание сомнения» и «Апелляция к страху, предубеждения».

Язык с выраженной лексикой: Согласно Пискорски и др. [298], язык с выраженной лексикой несёт в себе сильные эмоциональные подтексты, которые могут быть как положительными, так и отрицательными. По нашим данным, эти подтексты преимущественно негативные, что подтверждает предыдущие результаты исследований по выявлению дезинформации. Лишь очень небольшая доля слов в наших данных используется с положительной оценкой.

Обращение к авторитету: этот метод связан с источником информации. Наиболее важным для лингвистического анализа является то, как источник связан с атрибутируемым материалом посредством набора языковых выборов, которые со временем стали довольно конвенциональными.

Обращение ко времени: Обращение ко времени используется, когда автор подчёркивает своевременность идеи, чтобы побудить к (сомнительному) действию. При ближайшем рассмотрении выяснилось, что большое количество примеров в наших данных (36%) взято из одного текста, что затрудняет обобщение; однако в остальных текстах прослеживаются некоторые закономерности. Само слово «ВРЕМЯ» упоминается всего четыре раза в составе устойчивых выражений, таких как «it’s time» и «anytime», которые едва ли передают ощущение срочности. Вместо этого в пропагандистских текстах эту функцию, по-видимому, выполняют различные дейктические отсылки ко времени, стремящиеся минимизировать временное расстояние между точкой отсчёта и действием, обладая способностью к неограниченному расширению, то есть охватывая «широкий временной диапазон» [312: 404].

Кроме того, мы наблюдаем ряд лексических ресурсов для выражения срочности, включая прилагательные (срочно), наречия (отчаянно) и существительные (приоритет), хотя ни один из них не используется систематически в наших данных.

Вызывание сомнений: этот метод использует широкий спектр языковых ресурсов, которые подвергают сомнению общую достоверность, качество или конкретные действия человека. В наших данных они распределены довольно равномерно между более или менее явными и неявными ресурсами.

Используя триггер пресуппозиции, автор фактически заставляет читателя принять реальность, которая не только не поддаётся проверке и субъективна, но и крайне тенденциозна и даже ложна [308: 4], а именно, что основные средства массовой информации намеренно нагнетают страх перед COVID-19.

Апелляция к страху, предубеждениям: В наших пропагандистских данных отвращение к идее наиболее ярко выражено в отношении большого количества, частоты или количества чего-либо, например, времени, людей, вероятности и т. д.

По мнению Трнавац и Табоады, условные отношения могут способствовать выражению оценки либо посредством усиления, либо посредством ослабления (см. также ниже уступительные отношения).

Оставшиеся подтехники рассматриваются в контексте общих категорий, к которым они относятся. В связи с низкой частотой встречаемости и/или отсутствием чётких закономерностей мы не будем рассматривать подтехники «Оправдание» и «Призыв» (но см. примеры в Дополнительных материалах).

Манипулятивная формулировка: В этой категории все оставшиеся подтехники демонстрируют интересные лингвистические и прагматические особенности. Во-первых, среди «Преувеличения» и «Минимизации» мы в основном встречаем преувеличение в наших пропагандистских текстах. Как и ожидалось, преувеличение наиболее ярко проявляется в прилагательных, которые характеризуются свойством градуируемости и, следовательно, могут сочетаться или не сочетаться с модификаторами степени.

Наконец, «запутывание, намеренная неопределенность, путаница» — это общий термин для преднамеренной неясности аргументации. Типичным лексическим ресурсом является, предположительно, косвенный доказательный маркер, используемый для избежания ответственности при распространении потенциально ложной информации.

Атака на репутацию: Среди оставшихся подприемов в этой категории мы уделяем особое внимание «Апелляции к лицемерию» из-за его убедительного сочетания явных и неявных пропагандистских стратегий. С одной стороны, когда авторы обвиняют кого-либо в лицемерии или непоследовательности, они, как правило, делают это, создавая явный контраст между двумя суждениями, между «реальным» и «идеальным» описаниями событий.

Упрощение: единственный подприём в этой категории, допускающий некоторое обобщение, — это каузальное упрощение, когда автор предполагает единую причину или основание для сложной проблемы. По нашим данным, это достигается в основном посредством каузальных (p, потому что q) или временных (до p, q) дискурсивных отношений.

Отвлечение: Здесь мы сосредоточимся на искажении чьей-либо позиции — побочном приёме, при котором реальный субъект заменяется ложным для искажения аргумента. Таким образом, он изобилует ложными предпосылкой и предположениями о мыслях, желаниях, намерениях, отношениях и действиях третьего лица.

Как показано выше, мы выявили значительное количество случаев (112), которые, предположительно, выполняли функцию убеждения в контексте их использования, но не подходили ни к одной из категорий убеждения, установленных Пискорским и др. [298]. Мы воздержались от их строгой классификации и вместо этого отметили некоторые общие тенденции. Наше резюме ниже основано на сравнении с методами убеждения, выявленными Партингтоном и Тейлором, поскольку их подход был более лингвистически ориентированным и, следовательно, может заполнить пробелы, оставленные Пискорским и др. [298].

Мы обнаружили, что подавляющее большинство примеров в этой категории можно объяснить, ссылаясь на системы оценки и позиции, на понятия установки, диалогического взаимодействия, а также на другие маркеры разговорного, вовлеченного стиля. Распространяя пропагандистские нарративы, авторы часто выражают свои личные убеждения и точки зрения, в том числе для того, чтобы вселить надежду на конкретный результат, успех которого часто зависит от самих авторов.

Следовательно, именно интерактивное и межсубъективное измерение отсутствовало в первоначальных рекомендациях. И это несмотря на то, что оценка и позиция играют важную роль в новостных жанрах [275], а эмоциональное и межличностное убеждение подробно описано в работе Партингтона и Тейлора. Их концепция также охватывает некоторые менее часто используемые в наших данных техники (например, противопоставление «они» и «мы»), но не охватывает большинство других субтехник, рассматриваемых в данном исследовании. Именно благодаря этому разнообразию техник и стратегий, а также предыдущему применению к пропагандистским текстам в работе Да Сан Мартино и др. [299], концепция убеждения, разработанная Пискорским и др. [298], оказалась столь полезной для нашего лингвистического исследования и его распространения на прагматические исследования.

### 6.2.5 Представление климатического кризиса в хорватских интернет-новостных СМИ (Трнавац Р.)

На протяжении десятилетий под изменением климата Земли подразумевались как антропогенные, так и естественные сдвиги, вызывающие потепление атмосферы, океанов и суши, нарушающие жизнеобеспечивающие экосистемы и наносящие вред здоровью человека. Хотя термины «изменение климата» и «глобальное потепление» используются с 1970-х годов, первый доминирует с 1990-х годов. Однако ни один из этих терминов не несёт экзистенциальной угрозы: «изменение» звучит безобидно и расплывчато, подразумевая любые изменения. Такая трактовка позволяет политическим деятелям сомневаться в существовании изменений или отвергать их как естественный процесс, не зависящий от человека.

С 2021 года, под влиянием КС-26 и суровых предупреждений МГЭИК об антропогенном потеплении, климатические проблемы стали центром общественного дискурса и получили широкое освещение в СМИ. Исследования показывают, что в этом освещении используются разнообразные фреймы, и подчёркивается необходимость изучения неанглоязычных национальных контекстов. Поэтому мы рассмотрим, как хорватские онлайн-новостные СМИ представляют климатические проблемы в этом менее изученном европейском контексте.

Исследования в области коммуникации по вопросам изменения климата обширны. Приведённый ниже обзор по необходимости краток и поэтому сосредоточен на вопросах, наиболее важных для нашего анализа: агентность, новостные ценности, (мультимодальная) метафора и метонимия, визуальные репрезентации и освещение изменения климата в хорватских СМИ. Шефер и Шлихтинг изучили 133 эмпирических исследования, посвящённых дискурсу об изменении климата: 92 из них основывались на данных США, в то время как Австралия, Германия и Великобритания были представлены менее чем в 20 исследованиях, а все остальные (в основном европейские) страны — менее чем в 10. Исследования агентности показывают, что распределение агентских ролей имеет значение: персонификация климата размывает ответственность, тогда как указание конкретных субъектов проясняет пути к действию. Воспринимаемая агентность также различается в зависимости от издания: элитные газеты делают акцент на государственных и корпоративных субъектах, в то время как платформы социальных сетей выдвигают на первый план низовой активизм и обвинения. Связь между деятельностью и решениями ещё больше повышает воспринимаемую эффективность: дискурс, сочетающий атрибуцию с конкретными мерами или образовательными призывами, побуждает аудиторию действовать, а не просто наблюдать. ДеМейер призывает коммуникаторов перейти от повествования, основанного на проблемах, к повествованию, основанному на действиях, помещая людей, принимающих позитивные меры по борьбе с изменением климата, в центр каждого повествования для создания коллективной деятельности.

Исследования мультимодальности, метафоры и метонимии в контексте окружающей среды сосредоточены главным образом на рекламе, кампаниях и протестных визуальных материалах и показывают, что многие очевидные метафоры основаны на метонимических сигналах. Эта научная работа остаётся англоцентричной, а образные репрезентации в онлайн- и традиционных новостных СМИ до сих пор остаются в значительной степени неизученными, что ставит под сомнение применимость типологии визуальных и мультимодальных метафор Форсвилля к новостным визуальным материалам. Исследования климатических изображений проверяют, какие визуальные эффекты стимулируют активность в отношении мер по борьбе с изменением климата, но результаты остаются неубедительными относительно того, какие изображения более эффективны — изображения, ориентированные на решение или воздействие. O’Neill et al обнаружили, что визуальные эффекты формируют вовлеченность двумя различными способами: некоторые усиливают заметность (воспринимаемую важность изменения климата), в то время как другие повышают эффективность (уверенность в возможности значимых действий). Фотографии, ориентированные на воздействие, повышают заметность, но также создают психологическую дистанцию и отчуждение; портреты узнаваемых людей могут даже снизить воспринимаемую важность.

Исследования репрезентации изменения климата в хорватских СМИ ограничены. Врселья анализирует реакцию аудитории на сообщения, основанные на страхе; Калайжи документируют активизм знаменитостей и нарративы отрицания; а Башич, используя критический дискурс-анализ, показывают, что хорватские СМИ представляют изменение климата как отдалённую научную проблему, предлагая мало точек зрения с точки зрения человеческих интересов и слабо приписывая ответственность по сравнению с англоязычными СМИ.

Мы рассматриваем, как изменение климата представлено в хорватских онлайн-новостных СМИ, рассматривая два исследовательских вопроса, связанных с агентностью и дискурсивным воздействием.

1. В какой степени вербальные и визуальные репрезентации климатического кризиса совпадают в их изображении агентности? Мы исследуем это, анализируя новостную ценность, визуальное оформление изменения климата и использование метафор, связывающих агентность с более широкими вопросами ответственности и возможными способами действия.
2. В свете (1), как хорватские онлайн-СМИ формируют более широкий дискурс об изменении климата и климатическом кризисе?

Следуя концепции межсемиотического параллелизма Лю и О’Холлорана (2009), наш визуальный анализ выдвигает на первый план те же элементы, что и в вербальном компоненте, и включает дополнительные фреймы визуального климата. Чтобы оценить взаимодействие агентности с метафорическим языком, новостной ценностью и визуальным оформлением, мы используем четыре теоретические модели: модель транзитивности Халлидея и Маттиссена; теорию концептуальной метафоры Лакоффа и Джонсона и последующие расширения Форсвилля и Харта; анализ новостных ценностей; и визуальное оформление изменения климата.

Чтобы определить формы агентности, изображённые в новостях о климатическом кризисе, мы применили модель транзитивности Сиднейской школы к текстовым и визуальным данным. Впервые предложенная Халлидеем и уточнённая Халлидеем и Маттиссеном, эта модель описывает диапазон вариантов, доступных говорящим «для выражения своего восприятия участников, процессов и обстоятельств внутри и за пределами своего внутреннего мира».

Сущность в изображениях охватывает более широкий спектр семантических ролей, чем в вербальном корпусе, поскольку визуальные образы обычно изображают конкретные сущности. В нашей классификации эти роли выражаются через термины для таких сущностей. Более того, изображения допускают более широкую интерпретацию, в то время как лингвистический контекст проводит зрителя через «означаемые изображения».

Мы также исследуем агентность, используя анализ дискурсивных новостных ценностей (DNVA); классифицируя как вербальные микроконтексты, так и изображения. DNVA, метод критического дискурсивного анализа, раскрывает идеологический аспект отбора новостей, показывая, как новостные ценности «акцентируются, встречаются редко или отсутствуют в репортажах по определённым темам и событиям». Он позволяет проанализировать, как вербальные и визуальные элементы подкрепляют, усиливают или противоречат друг другу, и помогает определить, какие практики в мультимодальных новостных материалах успешны или неудачны.

Мы дополнили наш анализ изображений фреймворком О’Нила, который ориентирован на визуальное содержание изображений, связанных с изменением климата, в то время как модель новостных ценностей Беднарека и Кейпла выделяет идеологические аспекты. Наша переработанная версия типологии О’Нила выделяет следующие типы изображений:

1. узнаваемые люди — изображения с узнаваемыми лицами (например, политиками, учеными, руководителями бизнеса, знаменитостями и т. д.);
2. воздействие изменения климата — изображения, демонстрирующие воздействие или последствия изменения климата на окружающую среду и сообщества (например, таяние льдов, экстремальные погодные явления, виды, находящиеся под угрозой исчезновения). Эти изображения часто содержат «визуальные синекдохи, отражающие отдаленные климатические риски»;
3. причины изменения климата — визуальные изображения, демонстрирующие производство энергии, выбросы и загрязнение (например, дымовые трубы, транспорт); (d) Решения (например, ветряные турбины);
4. изображения протестных демонстраций и активизма, например, активистов и митингов;
5. научные изображения — графика и визуальные материалы из области климатологии (например, диаграммы, графики, карты).

Работа О’Нила исследует, какие фреймы доминируют в коммуникации СМИ, освещающих климатические вопросы, и как современное визуальное оформление влияет на более широкое участие общественности. Каждый фрейм способствует определенным формам взаимодействия, отодвигая на второй план другие: например, когда изменение климата изображается как спорное, политизированное или оторванное от повседневной жизни явление, это редко повышает у людей чувство собственной эффективности.

В нашем наборе данных не было карикатур на тему изменения климата — категория, встречающаяся в данных Великобритании и США, — но он включал несколько категорий, выходящих за рамки схемы О’Нила. Мы сгруппировали их под дополнительной меткой «Другое», включающей неидентифицируемых людей, позитивно и негативно оформленные пейзажи, животных, книги и художественные изображения, последние из которых встречались особенно часто.

В совокупности вышеперечисленные структуры обеспечивают прочную основу для изучения дискурса хорватских онлайн-СМИ о климатическом кризисе.

Материал был собран в августе 2024 года с помощью расширенного поиска Google. Домен был установлен на «.hr», язык — на хорватском, а временные рамки — на 2020–2024 годы. Используя термин klimatska kriza («климатический кризис»), мы провели отдельные поиски для всех грамматических случаев этого словосочетания. Для каждого поиска первые семь страниц с результатами (≈350 результатов) были просмотрены вручную; последующие страницы в основном повторяли предыдущие результаты. Были выявлены медиаматериалы и удалены дубликаты, что позволило получить 93 уникальных текста из национальных, региональных и местных онлайн-газет и новостных порталов. Хотя состав результатов Google не контролируется исследователем, мы считаем, что эта выборка типична для хорватского онлайн-новостного дискурса за этот период.

Термин klimatska kriza («климатический кризис») встречается 229 раз в 93 текстах. Около 20% этих случаев занимают заметные позиции (заголовок, подзаголовок или лид). Примерно в трети текстов этот термин встречается как минимум в двух из четырёх структурных позиций — заголовке/лиде/введении, середине, заключении и теге — и может повторяться несколько раз в одной позиции.

93 текста содержали 242 изображения — примерно два-три на статью — а в некоторых галереях их было гораздо больше. Если текст включал видео, мы анализировали его начальный кадр. Изображения сначала собирались автоматически: код Python в Jupyter Notebook использовал fitz и PdfReader для извлечения из PDF-файлов и, при необходимости, изменения размера.

Затем результаты проверялись и дополнялись вручную. Каждое изображение кодировалось дважды: (1) с адаптированной версией климат-визуальной типологии О’Нила и (2) с помощью схемы новостных ценностей Беднарека и Кейпла. Поскольку одно изображение может отражать несколько категорий, при необходимости присваивалось несколько кодов.

В качестве единицы анализа вербального подкорпуса мы рассматривали контекст предложения с этим термином (расширяемый до соседних предложений при необходимости уточнения). Визуальная интерпретация начиналась с заголовка, сопоставлялась с ведущим изображением и затем связывалась с текстом, в котором упоминались климатический кризис или изменение климата, а также с ведущими абзацами. Аннотация включала четыре этапа: (1) определение исходных доменов метафор (если они присутствовали) и преднамеренных метонимий, (2) указание агентства, (3) кодирование новостных значений и (4) назначение климатически-визуального типа в соответствии с классификацией О’Нила.

Наше внимание сосредоточено на том, как представлены человеческие субъекты, их ответственность и способность к действию. Преобладают неметафорические фреймы: климатический кризис и изменение климата в основном представлены как проблемы или вызовы (проблема, izazov; ključno društveno-političko pitanje ‘ключевой социально-политический вопрос’) — и часто как темы для обсуждения или обмена информацией (например, kompleksna i apstraktna tema ‘сложная и абстрактная тема’). В рамках этого буквального фрейма «проблемы» распространёнными словосочетаниями являются riješiti klimatsku krizu ‘решить климатический кризис’ и odgovoriti na klimatsku krizu ‘реагировать на климатический кризис’. Некоторые фреймы подчёркивают их фактологический статус (например, objektivna znanstveno dokazana činjenica ‘объективно научно доказанный факт’), тем самым междискурсивно связываясь с дискурсами отрицания, которые подвергают сомнению реальность изменения климата. В большинстве примеров климатический кризис и изменение климата рассматриваются как активные, злонамеренные Агенты, выполняющие семантическую роль Субъекта. Субъект климатического кризиса и изменения климата представлен как в нейтральных, так и в отрицательных терминах. К нейтральным глаголам относятся uzrokuje («вызывает»), utječe («воздействует»), čini («делает») и motivira («мотивирует»). К глаголам с отрицательным оттенком относятся razara («разрушает»), ubija («убивает») и prijeti («угрожает»). Они сочетаются с каузативными глаголами и объектами, указывающими на пагубные последствия — ugrožava, izaziva negativne posljedice («ставит под угрозу, вызывает негативные последствия»). Иногда субъект актера находится на заднем плане, в то время как пострадавшие территории или объекты (например, zemlje «страны») вынесены на передний план.

Цель — вторая по частоте семантическая роль, приписываемая изменению климата в наших данных. Обычно она обозначается как сущность, на которую негативно влияет процесс или обстоятельство.

Выделенная роль редко приписывается климатическому кризису или изменению климата, тогда как оба термина часто используются в качестве носителей в различных микроконтекстах. Они неизменно трактуются негативно — как проблемы или вызовы.

Негативность соответствует всем ключевым семантическим ролям и связанным с ними типам агентств. Новостная ценность своевременности, которая подчёркивает неотложность решения проблемы климатического кризиса, позволяет различать конкретные примеры негативности.

Метафоры играют ключевую роль в коммуникации по вопросам изменения климата, поскольку они формируют общественное понимание. Исследования метафорического фрейминга выявляют доминирующие метафоры, такие как война и раса, и изучают, как эффективные метафоры могут побуждать к действию, устанавливать приоритеты и влиять на распределение ресурсов.

Метафорические репрезентации встречаются в наших данных лишь изредка и в основном носят условный характер. Мы выделили четыре повторяющихся репрезентации, наиболее значимая из которых представляет климатический кризис и изменение климата как ЗЛОГО АГЕНТА или ВРАГА в метафорической ВОЙНЕ. Приписывание роли субъекта ситуации или переломного момента — само по себе метафора– отодвигает на задний план действия людей и действия, усугубившие кризис.

В нашем наборе данных ответственность человека лишь изредка выдвигается на первый план: кризис редко описывается как явно антропогенный. Чаще же человеческое участие отодвигается на второй план либо с помощью пассивных конструкций, либо просто подразумевается. В целом, освещение событий изображает людей и места их обитания как жертв, находящихся под угрозой климатических явлений, с которыми необходимо бороться, противостоять и адаптироваться. Помимо доминирующего фрейма войны, три другие повторяющиеся концептуализации изображают климатический кризис/изменение как: (1) ЧЕЛОВЕК — например, [empatij] nam je hitno potrebna kako bismo se zajedno suočili s klimatskom krizom («Нам срочно нужно сочувствие, чтобы вместе противостоять климатическому кризису»); (2) ОБЪЕКТ, который измерим, подвижен или видим — например, veličina klimatske krize («масштабы климатического кризиса»); и (3) ПРОСТРАНСТВО — например, svijet koji klizi u klimatsku krizu («мир, скатывающийся в климатический кризис»).

Метафора ПРОСТРАНСТВА предполагает, что можно войти в изменение климата или климатический кризис и выйти из него. Выбранный глагол указывает на природу этого движения и доступную движущим силам (обычно людям): например, kliziti «скольжение» подразумевает неконтролируемое движение и ограниченную свободу действий. Как метафорические ОБЪЕКТЫ, климатические явления кажутся измеримыми и управляемыми. Однако такая интерпретация вводит в заблуждение, поскольку подразумевает, что их можно устранить или устранить простыми действиями — подразумевание, искажающее сложность и устойчивость климатического кризиса и изменения климата.

Дополнительные исходные домены встречаются лишь в отдельных примерах — например, климатический кризис представлен как СПОРТИВНЫЙ СОПЕРНИК, ЖИВОЙ ОРГАНИЗМ, ПРИРОДНАЯ СИЛА или даже ПИЩА. В метафорическом контексте роль климатического кризиса и изменения климата реализуется главным образом через роль Цели (Goal), поскольку лейтмотивом соответствующих примеров выступает «борьба с катастрофическими последствиями изменения климата». Это выдвигает на первый план, прежде всего, новостные ценности Негативности (Negativity) и Воздействия (Impact).

Два доминирующих визуальных типа — «Люди» и «Воздействие изменения климата». Изображения, связанные с причинами изменения климата, решениями, протестными изображениями, произведениями искусства и научными материалами, встречаются примерно с одинаковой частотой, но гораздо реже. На некоторых изображениях изображены животные, связанные с последствиями, причинами или решениями. Пейзажи в негативном оформлении, часто перекрывающие визуальные образы воздействия, встречаются редко, а пейзажи в позитивном оформлении встречаются ещё реже. На многих изображениях преобладают вербальные элементы, особенно лозунги, чаще всего в протестных изображениях, которые освещают причины изменения климата, предлагают решения и подчёркивают неотложность. В некоторых случаях текст (например, заголовок) встроен непосредственно в изображение. На некоторых изображениях также изображены книги.

В категории «Люди» большинство изображений изображают неопознанных людей. Среди опознаваемых персонажей чаще всего и в равной степени встречаются учёные и активисты, реже — представители СМИ и политики. На некоторых изображениях представлены другие специалисты (например, педагоги). Наши визуальные материалы редко политизируют изменение климата: политики появляются редко, в то время как неопознанные люди — «люди вроде нас» — встречаются часто, что делает проблему более актуальной. Многие изображения фокусируются на последствиях, которые могут побудить к конструктивным действиям, однако частое изображение учёных и активистов может дистанцировать тему, ассоциируя её со специализированными группами. Несмотря на недавние призывы к использованию изображений, более ориентированных на решение проблем, такие визуальные материалы редко встречаются в хорватских онлайн-новостях.

Наиболее часто встречающиеся категории новостных ценностей в классификации Беднарека и Кейпла пересекаются с наиболее значимыми категориями в системе О’Нилса. Влияние и негатив доминируют в освещении ущерба от изменения климата, в то время как персонализация, эстетическая привлекательность и элитарность встречаются реже. Категория «Люди» О’Нила совпадает с двумя категориями Беднарека и Кейпла: персонализация (Personalization) (обычные люди, активисты, дети) и элитарность (Elitness) (ученые, писатели, политики, представители СМИ). Эстетическая привлекательность проявляется главным образом в произведениях изобразительного искусства — орнаментальных композициях, инсталляциях, живописных пейзажах, — которые представляют спокойную природу на фоне климатических катастроф, часто через метонимические образы флоры или деревьев, символизирующих идеальную среду. Позитив встречается реже и обычно сочетается с эстетической привлекательностью. Позитив и негатив могут сосуществовать в одном изображении, отражая двойственность добра и зла в природе.

Визуальные образы представляют климатический кризис преимущественно в глобальном масштабе. Созвучие (Consonance) и близость (Proximity) встречаются редко: лишь несколько примеров показывают явления, характерные для Хорватии, и лишь немногие отражают географию Хорватии. Превосходство (Superlativity) — события, представленные как исключительно интенсивные или масштабные, — также встречается нечасто. В целом, изображения подчёркивают глобальные негативные последствия, а не прямую катастрофу, однако они подразумевают, что осознанные действия и активизм могут привести к позитивным изменениям. Классификация изображений по новостным ценностям дополняет нашу адаптированную модель О’Нила, однако ни одна из них не включает в себя категорию метонимических или метафорических признаков, которые мы поэтому ввели в нашу собственную климато-визуальную типологию.

Роль «Цель» — сущности, затронутые процессами изменения климата, — доминирует во многих изображениях. Эти визуальные образы делятся на три основные группы: (1) артефакты — примеры включают поврежденную Статую Свободы, разрушенные дома и сады, электромобили, оскверненные активистами-экологами, произведения искусства из переработанных материалов, подчеркивающие кризис, затопленные улицы и графику, символизирующую климатические потрясения в сфере финансов или интернета; (2) природные явления — чаще всего изображаются Земля, потрескавшаяся почва, морские пейзажи, затопленные ландшафты, бурное море, разрушенные наводнениями парки, Средиземное море, леса, короткий световой день, киты, крупный рогатый скот, выжженная местность, воздух, Европа, пингвины и рыбные запасы; (3) группы людей — чаще всего в зоне риска находятся обычные граждане, рабочие и пожарные, а также человечество в целом, находящееся под угрозой из-за супербактерий.

Агентивность также представлена категориями «Атрибут» и «Идентификация». Мы объединили «Атрибут» и «Идентификацию» в одну категорию для визуального анализа, поскольку оба они описывают демонстрируемые объекты или явления. В отличие от вербальных данных, где изменение климата и климатический кризис выступают в роли носителей атрибутов, мы не обнаружили визуальных примеров роли носителей. Вместо этого изображения, часто подкрепленные текстом, либо изображают последствия изменения климата, либо представляют идеализированный образ, к которому следует стремиться.

Ещё одна важная категория визуального корпуса — Sayer, чьи темы во многом пересекаются с теми, кто представлен в категории «Актёр» (Actor): эксперты, активисты, учёные, авторы, публикации, политические деятели, представители СМИ, сотрудники служб гражданской обороны и даже символический глобус, изображённый на плакате. Словесные выражения, отсутствующие в письменном корпусе, но заметные в визуальных материалах, представлены во многих формах: книги и романы, письменные предупреждения, предупреждающие знаки об экстремальной жаре, символические изображения резкого повышения температуры и надвигающихся пандемий, диаграммы, отражающие уязвимые отрасли, пострадавшие от сокращения рабочих мест, карты, показывающие эксплуатируемые районы северной Адриатики, письменные заявления климатических активистов, рекламные материалы для художественных инсталляций, а также фрески и рисунки, пропагандирующие сохранение климата. Итак, нарратив об изменении климата в хорватских онлайн-новостях несколько противоречив. Он в основном вызывает тревогу, подкреплённую новостными ценностями «Негатив» и «Влияние», а также неоднократным представлением человечества как цели как в тексте, так и в визуальных образах. Напротив, позитивизму, освещающему усилия по спасению планеты, уделяется гораздо меньше внимания. В текстовом плане климатический кризис и изменение климата чаще всего выступают в роли действующих лиц, представляя сам кризис как причину серьёзного вреда человечеству.

Метафорическое оформление в вербальной части онлайн-новостей следует этой схеме: климатический кризис и изменение климата — абстрактные сущности — представляются злонамеренными агентами в доминирующем военном сценарии, в то время как люди предстают жертвами, противостоящими более сильному врагу. Такое представление отодвигает на второй план действия человека и ответственность за возникновение кризиса. Другие метафоры (например, представление кризиса как объекта манипуляции) также искажают способы его решения. Изменение климата и климатический кризис лишь изредка прямо описываются как антропогенные; вместо этого в репортажах люди и места их обитания в основном изображаются как пассивные жертвы, которым угрожают климатические явления.

В большинстве статей изменение климата рассматривается как глобальная проблема и иллюстрируется образами мест, удалённых от Хорватии. Поскольку новостные ценности «Близость» и «Созвучие» используются лишь изредка, угроза представляется отстранённо.

Адриатическое море или Хорватия представлены крайне редко, и лишь в нескольких статьях Средиземноморье, частью которого является Хорватия, описывается как очаг изменения климата. Ограниченное освещение историй, персонализированных местным сообществом, свидетельствует о том, что позитивные действия и активизм обычных людей получают лишь второстепенное внимание. В нашем наборе данных редко описывались конкретные, повседневные действия обычных местных жителей; новостные сюжеты редко освещали инициативы, которые могли бы способствовать формированию чувства самостоятельности. Как показали предыдущие исследования других национальных контекстов, в наших данных очень мало текстов, возлагающих вину за климатическую катастрофу на конкретных субъектов, таких как правительства или корпорации. В вербальной информации преобладают голоса элиты — политической, научной и медийной, — которые сосредоточены преимущественно на угрозах. Поскольку новостная ценность персонализации невелика, повествование кажется отстранённым и, следовательно, недостаточно релевантным для обычного читателя.

Визуальные материалы редко политизируют проблему: политики показаны редко, в то время как учёные — часто, что может создать впечатление, что проблема далека от реальности и чрезмерно техническая. Неопознанные лица часто встречаются, что подразумевает, что «люди, подобные нам», затронуты, однако узнаваемый местный контекст встречается редко. Многие изображения изображают последствия изменения климата, которые могли бы побудить к конструктивным действиям, но описания практических решений — зачастую более убедительных и жизненно важных для долгосрочного смягчения последствий — остаются редкими. Метафорические репрезентации климатического кризиса/изменения встречаются относительно редко в вербальных данных, и большинство из них носят традиционный характер. В визуальных материалах такие репрезентации встречаются ещё реже. Немногочисленные метафорические образы встречаются преимущественно в образовательном контексте.

### 6.2.6 Повышение точности орфокоррекции за счёт двухэтапного применения нейронных сетей

Системы автоматической проверки и оценки письменных текстов являются хорошим подспорьем как для учителей, так и для учащихся. Тем не менее существующие системы обладают весомыми недостатками: они плохо интегрируются в учебный процесс; не предоставляют учащемуся возможности самостоятельно исправлять допущенные ошибки или дополнительно упражняться на схожих контекстах [313; 314: 2; 315].

В попытке избавиться от этих недостатков мы создали собственную систему, которая способна искать ошибки в текстах на иностранном языке и исправлять их. Следуя методике, предложенной в работе [316], мы провели дообучение модели-детектора ошибок на основе архитектуры ELECTRA [317] и использовали ее вывод для дообучения модели-корректора T5 [318]. Мы рассмотрели как базовую ее версию, так и версию, специально предобученную для задачи коррекции грамматических ошибок [319]. Для дообучения мы использовали два варианта корпуса англоязычных учебных текстов REALEC [320]. Первый вариант опирался на ручную разметку, уже существующую в корпусе: она включала по большей части исправление самых бесспорных и простых для исправления ошибок (грамматических форм, артиклей). Второй вариант включал в себя предложения из корпуса REALEC, автоматически исправленные с помощью модели Llama-3.3 [321]. Предложения, вошедшие в этот вариант корпуса, содержали более глубокие и сложные, но во многом излишние исправления (выбор более уместного вводного слова, исправление стилистических недочетов, изменение синтаксической структуры предложения при условии грамматической правильности существующей).

Сравнивая различные варианты дообученной нами на разных данных нейросети T5, мы делаем вывод о влиянии на конечный результат как тренировочных данных, так и вывода модели-детектора ошибок. Базовая T5, дообученная нами на корпусе с более простой разметкой, вносит в написанный студентом текст точечные исправления, при этом игнорируя ошибки, исправление которых требует более глубокого вмешательства в написанный студентом текст (синтаксически неверное построение фразы, неверный порядок слов). Но эта же архитектура, дообученная на корпусе с нейросетевой разметкой, начинает вносить в поданный ей для анализа текст более сложные исправления, которые часто оказываются излишними. Результаты, полученные после дообучения T5, предобученной на Lang8, демонстрируют, что вне зависимости от корпуса дотренировки эта модель тяготеет к внесению более простых исправлений. Но использование базовой T5, дообученной на переразмеченных данных с учетом вывода модели-детектора позволяет достичь «золотой середины»: данная модель способна как вносить в текст сложные исправления, так и воздерживаться от исправлений излишних. Данный результат свидетельствует о важности учета вывода модели-детектора: мы показываем, что он позволяет стабилизировать модель, сделав ее исправления более направленными и точечными, не нанося при этом ущерба сложности разметки. Также данное исследование позволяет конкретизировать полученный ранее результат Yuan et al. [316], определив, в чем именно заключается отмеченный авторами прирост эффективности.

Появление такой «промежуточной» модели важно для разработанной нами системы автоматической оценки учебных текстов, потому что, в отличие от существующих сегодня ресурсов, она способна работать с достаточно широким спектром ошибок, но в то же время не рискует преждевременно познакомить студента с языковыми структурами, к изучению которых он может быть не готов.

Еще одним результатом данной работы является разработанное онлайн-приложение, с помощью которого учителя и студенты могут воспользоваться моделью. Функционал данного приложения предоставляет ученику возможность работать с нейросетью как самостоятельно, так и в паре с учителем: в таком случае учитель тоже получает доступ как к написанному учеником тексту, так и к внесенным нейросетью правкам. Также система способна предлагать студенту для самостоятельной коррекции предложения из корпуса данных REALEC с ошибками тех же типов, что и допущенные им ошибки: в отличие от предыдущих систем упражнений, разработанных на материале корпуса REALEC, например, [322], более «активный» подход к исправлению своих ошибок, возможно, помогает студенту лучше усвоить информацию.

Ограничением проведенного исследования является тот факт, что большинство авторов текстов REALEC — носители русского языка, что делает затруднительным использование полученной нейросети при работе с носителями других языков. Перспективой данного исследования является дальнейшее тестирование разработанной системы и дальнейшее ее масштабирование. Сейчас ведется подготовка к экспериментальной фазе, во время которой будет доказана или опровергнута эффективность данной системы при подготовке студентов к Независимому экзамену.

### 6.2.7 Оценка свободы синтаксического ветвления в различных языках

Нами был проведен анализ статистики употребления синтаксических связей. Предлагается собрать статистику для пар и троек слов, причём слова будут представлены только частью речи, а для синтаксической связи будет учитываться её тип. Для троек предлагается учитывать как связи между двумя потомками одной вершины, так и связи для трёх разных уровней. Для выбранной синтаксически размеченной коллекции предлагается рассчитать статистику появления связей между всеми видами пар и троек слов. Далее выбираются десять наиболее частотных комбинаций.

Для сравнения нескольких коллекций следует объединить списки наиболее частотных комбинаций и для каждой из коллекций взять статистику для объединённого списка. Степень сходства синтаксиса коллекций определяется с использованием ранговой корреляции.

Свобода синтаксического ветвления определяется при помощи оценки соотношения частот правого и левого синтаксического ветвления. При этом более значимыми считаются более частотные синтаксические связи. Высокая вариабельность положения зависимого слова показывает, что в языке существуют какие-то правила, определяющие его позицию (синтаксические или семантические).

Нами показывается, что сходство синтаксических конструкций существенно зависит от принципов разметки корпуса, а также его стиля. Предложенный метод позволяет найти конкретные отличия в синтаксисе двух корпусов, после чего может быть проведён их более глубокий лингвистический анализ. Так, например, в главе показано, что в текстах беллетристики с большим количеством диалогов существенно выше доля местоимений в роли субъекта, тогда как в формальных и аналитических текстах значительно больше герундиальных оборотов при субъекте.

Применение предложенных методов на материале корпусов из состава Universal Dependencies показало, что языки одной языковой группы больше похожи друг на друга, однако корректное филогенетическое дерево языков не может быть построено, так как сходство между некоторыми языками внутри группы оказывается слишком малым. Это может быть связано, например, с существенными отличиями стилей текстов внутри корпусов. Так, например, для древних текстов показано, что во многих корпусах, основой которых являются переводы Библии, одну из верхних строк по вариабельности занимают союзы, использующиеся одновременно для соединения слов и эффекта парцелляции.

Анализ вариативности синтаксического ветвления позволил провести ранжирование языков по степени строгости их синтаксического ветвления. Общая картина соответствует языковой интуиции. Конкретное ранжирование соседних языков может вызывать вопросы в связи с отличиями в стилях текстов внутри корпусов. К десятке языков с наиболее строгим синтаксисом отнесены японский, классический китайский, хинди, урду, корейский, уйгурский, ирландский, тайский, фарси, китайский, шотландский гэльский, арабский, турецкий, уэльский, фарерский. К десятке языков с наиболее свободным синтаксисом отнесены (в порядке убывания строгости) финский, русский, армянский, баскский, польский, словацкий, готский, литовский, латвийский, западный армянский, старый восточнославянский, церковнославянский, эстонский, латынь, древнегреческий.

Сравнение результатов с полученными в проекте Typometrics показало сходство, однако в нашем случае оценивается не преимущественное ветвление в языке в целом, а именно интегральную строгость следования превалирующему направлению для каждой отдельной связи (см. [рисунок 6.7](#fig-corp2)).

|  |
| --- |
| Рисунок 6.7 – Сравнение результатов определения свободы/строгости синтаксического ветвления в языке в целом (ось X) в сравнении с долей левого ветвления в языке, полученного в работе [323, рис. 8] (данные взяты из https://github.com/typometrics/djangotypometrics/blob/master/sud-treebanks-v2.11-analysis/head\_initiality.tsv) |

На примере количественных данных показано, что подход, использовавшийся ранее для определения преимущественного ветвления в языке устарел и требует замены на количественные методы. Оценка доли десяти синтаксических конструкций с максимальным значением параметра важности, от есть проявляющих свободу синтаксического ветвления, из всех конструкций, а также значение суммы параметра важности по всем конструкциям позволяет нам выделить пять групп языков: языки с регулярным синтаксисом, в которых практически отсутствуют симметричные конструкции (например, уйгурский и японский); языки с регулярным синтаксисом, в которых имеется небольшое количество симметричных конструкций (например, классический китайский); языки с небольшим количеством частотных конструкций, обладающих высокой долей синтаксической симметрии (готский, дренегреческий, древние славянские языки), при соблюдении строгости ветвления для большинства конструкций; языки, которым синтаксическая симметрия присуща в принципе, на любом уровне частотности конструкции (русский, украинский, белорусский); прочие языки.

Также мы приходим к выводу, что русский язык, благодаря присущей ему относительной простоте определения части речи и синтаксических связей, будет выбран для материала для дальнейшей работы.

## 6.3 Конвергентные процессы в синтетической и естественной речи

### 6.3.1 Специфика синтетических текстов и текстов, порожденных естественно в разных жанровых условиях

#### 6.3.1.1 Описание эстетически нагруженных вербальных стимулов (картин) мультимодальными языковыми моделями и информантами

В отчетном периоде продолжалась работа по созданию датасета описаний картин из коллекции Государственного Эрмитажа [324]. На данный момент коллекция включает в себя описания 2000 картин из экспозиционных залов, имеющих наибольшую посещаемость: Французская живопись XIX-XX веков, Итальянская живопись XIII–XVIII веков, Английская живопись XVI–XX веков, Французская живопись XV–XVIII веков, Русская живопись XVIII — начала XX века, Фламандская живопись XVII-XVIII веков, Немецкая живопись XVI-XVIII веков, Немецкая живопись XIX-XX веков, Голландская живопись XV-XVI веков, Испанская живопись XV-XX веков, Бельгийская живопись XIX-XX веков, Живопись различных европейских школ, Иконы. Каждая картина описывалась информантами и Visual Language Model (VLM), разработанной компанией Яндекс и находящейся в закрытом доступе[[3]](#footnote-315). VLM представляют собой класс мультимодальных моделей на основе технологии Transformers [325]. VLM натренированы на задачу Image to Text: распознают изображения, а затем дают комплексное вербальное описание, сопоставляя эмбеддинги частей изображения и слов.

Для создания датасета мы провели эксперименты с моделью и с 100 информантами (mean age=21,2; 74 женщины, 26 мужчин, студенты вузов г.Петербурга). Информантам в удаленном асинхронном режиме предлагалась таблица, где в одной колонке располагались картины (от 76 до 100), а во второй колонке предлагалось оставить описание. Информантам предлагалась следующая инструкция: «Опишите содержание картины своими словами. Вы можете начать с таких фраз, как “На картине изображено…”. Предоставьте подробные описания, состоящие из нескольких предложений. Пожалуйста, воздержитесь от поиска дополнительной информации о произведении искусства — опишите только то, что вы видите. Если вы узнаете историческую личность, пожалуйста, укажите её». Участники работали без ограничений по времени, и в среднем на заполнение одной таблицы уходило три недели. В среднем каждый респондент описал 176 картин. Участники предоставляли письменные описания картин, вводя свои тексты в предназначенные для этого ячейки общего файла электронной таблицы, который они затем возвращали кураторам проекта. В итоге полученный набор данных достиг разного уровня охвата: для каждого произведения искусства было получено от 3 (17,5% от общей выборки), 4 (53,9%), 5 (17,4%) или 12 (11,2%) отдельных описаний от разных информантов.

Метаданные для каждой из картин были сохранены в виде JSON-файла, названного по уникальному идентификатору картины в базе проекта.

По сравнению с прошлым годом объем датасета был увеличен на 50%. Полученный датасет позволил провести ряд экспериментов.

Во-первых, было проведено количественное сравнение характеристик текстов описаний картин от ИИ и от информантов. Статистический анализ текстов выявил различия в их количественных характеристиках: описания от моделей характеризовались средней длиной предложения 13,4 слова (стандартное отклонение 2,23), средним числом слов в тексте 65,1 (стандартное отклонение 20,1), с минимумом 18 и максимумом 156 слов; описания от информантов — средней длиной предложения 10,3 слова (стандартное отклонение 13,9), средним числом слов 58,7 (стандартное отклонение 40,3), с минимумом 3 и максимумом 324 слова. Таким образом, описания, сгенерированные моделями, отличаются большей стабильностью объема и длины предложений, а также меньшей вариативностью, что указывает на их более унифицированный и предсказуемый характер по сравнению с человеческими текстами, демонстрирующими значительную индивидуальную разнородность.

Кроме того, мы выдвинули гипотезу о том, что 1) лексическое и семантическое разнообразие описаний информантов зависит от типа описываемого изображения (люди/ натурфакты/абстрактные изображения — например, информанты дают более похожие друг на друга описания изображений природы, но более разнообразные, когда нужно описать людей или абстракции); 2) тип описываемого изображения предопределяет также степень отличности описаний информантов от описаний модели (например, описания абстрактных изображений максимально отличаются у информантов и моделей, а описания монофигурных композиций — минимально).

Для проверки гипотез мы использовали метод векторизации текстовых описаний с помощью модели *sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2.*

Использовалась подвыборка из 1215 картин. Для каждой картины были получены по три описания от мультимодальных моделей (Qwen2-VL-7B-Instruct, Qwen2.5-VL-7B-Instruct, Gemma-3-27b-it), что в сумме составило 3 645 описаний, и по 3 описания, полученных в рамках анкетирования от 24 информантов (средний возраст — 21 год, среднее количество картин, описанных одним информантом — 176). Итоговое количество исследуемых описаний “от человека” — 4 721.

Картины были дополнительно размечены по шести композиционным категориям: один человек (портрет или фигура в пейзаже), два человека (взаимодействующие или нет), три человека, четыре человека и более (групповая сцена), природа или прочие объекты (пейзаж, натюрморт, архитектура без людей или с мелкими фигурами), абстракция (без узнаваемых объектов или с неидентифицируемыми формами). Количество описанных картин по категориям составило соответственно 423 (1 269 описаний), 337 (1 011), 264 (792), 112 (336), 64 (192) и 16 (48).

Анализ включал векторизацию описаний с помощью модели-трансформера и вычисление косинусной близости в трех парадигмах: между описаниями от информантов, между описаниями от моделей и между парами описаний от информантов и моделей. Результаты показали, что описания от моделей в целом демонстрируют большую семантическую однородность по сравнению с описаниями от информантов независимо от категории. Медианные значения косинусной близости для описаний от информантов варьировались от 0,72 (три человека и абстракция) до 0,79 (один человек), с наибольшим разбросом в категориях с большим числом фигур; для описаний от моделей — от 0,77 (два и три человека) до 0,84 (природа и прочие объекты).

Анализ по категориям выявил закономерности в вербализации визуальной информации VLM-моделями и информантами. Наиболее согласованными оказались описания одиночных фигур и объектов природы: модели достигали медианной близости значений косинусного расстояния между описаниями 0,83 и 0,84 соответственно, информанты — 0,79 и 0,78, а межгрупповая близость составляла 0,78. По мере увеличения числа людей на изображении семантическое разнообразие возрастало: для сцен с четырьмя и более фигурами медианная близость значений косинусного расстояния между описаниями падала до 0,73 у информантов и 0,78 у моделей, с межгрупповой близостью 0,71. Абстрактные композиции демонстрировали наименьшую межгрупповую близость 0,69, что частично объясняется малым объемом выборки, но также указывает на различия в интерпретации неконкретных форм. На [рисунок 6.8](#fig-spb_1) представлено сравнение медианных значений косинусной близости описаний по категориям картин для информантов, моделей и межгрупповой выборки.

|  |
| --- |
| Рисунок 6.8 – Медианное значение косинусной близости описаний по категориям картин |

Исследование выявило заметные различия в том, как люди и визуально-языковые модели семантически обрабатывают изображения: модели дают более согласованные описания, а расхождения с человеческими текстами увеличиваются при усложнении сцены или переходе к абстрактным картинам.

В данный момент ведется исследование вербализации цвета в описаниях от информантов и от моделей. Проводится экспертная разметка описаний, которая в дальнейшем будет проанализирована на предмет различий в выражении семантики цвета. Первичные результаты указывают на то, что информанты описывают цвета в большем объеме, а также более разнообразными способами, чем модели.

#### 6.3.1.2 Оценка качества суммаризации художественных текстов большими языковыми моделями

В отчетном периоде сотрудники лаборатории, продолжая одну из магистральных тем “Цифровые методы в изучении литературы”, начали разработку задачи оценки качества автоматической суммаризации художественных текстов русском языке. В первую очередь оценивается успешность и применимость алгоритмов суммаризации на материале Корпуса русского рассказа, также предпринимается попытка определить лексические и структурные особенности художественного текста, которые могут влиять на результат автоматической суммаризации (лексическое разнообразие, нарушение фабулы, ненадежный рассказчик, смена лица повествования, вставки флешбеков и фрагментов других жанров).

За основу дизайна данного исследования был взят эксперимент ученых Колумбийского университета [326]. Эксперимент включал в себя несколько этапов:

1. формирование сбалансированной выборки из Корпуса русского рассказа;
2. суммаризация избранных документов;
3. автоматическая оценка результатов моделей;
4. экспертная оценка результатов моделей и структурной сложности исходных текстов;
5. сравнение полученных данных и выявление особенностей, повлиявших на результат выдачи.

Выборка аннотированных текстов из Корпуса русского рассказа была сбалансирована по параметру лица повествования. В результате в нее вошли 100 рассказов: 50 рассказов с повествованием от первого лица; 50 рассказов — от третьего лица.

Суммарный объем корпуса, взятого в работу, составил 330 129 словоупотреблений. Средняя длина рассказа —– 3 301,29 (sd = 2781,14).

В ходе эксперимента было обработано 100 рассказов и получено 300 саммари —– по 100 от каждой модели (GigaChat, T5 и sumy).

Нарратив в произведении — один из инструментов передачи авторской позиции, так как от расположения событий в тексте, уделенного им художественного времени, зависит их важность для рассказываемой истории. События в тексте могут располагаться не только нелинейно, но и сжиматься и растягиваться [327]. Предварительная оценка результатов показала, что алгоритмы автоматической суммаризации не всегда справляются с авторской передачей времени. Кроме того, эмпирический анализ показывает, что автоматическая суммаризация не всегда справляется с обработкой произведений с ненадежным рассказчиком. Модели плохо обрабатывают неожиданные развязки, не включают их в текст пересказа и таким образом создают неправильное впечатление о тексте, упуская ключевые моменты сюжета. Так, в рассказе «В плену» речь идет о двух разведчиках, которым нужно было найти «языка». Солдаты встретили поляка Яна Балицкого, готового сбежать в русскую армию добровольно, но попали в плен, а он успел убежать. Оказалось, что поляк дошел до русской армии и стал служить им переводчиком, однако рассказчик до последнего это скрывает. Ни одна из моделей не включает в пересказ развязку, поэтому создается ошибочное впечатление о том, что один из героев — предатель.

В качестве основных метрик автоматической оценки результатов суммаризации были избраны классические ROUGE [328], BLEU [329] и BERTScore [330].

Автоматическая оценка саммари показала, что наиболее качественные тексты создает GigaChat. Чуть менее успешной оказалась модель t5 и наименее успешно с задачей справилась модель sumy, что особенно видно по метрикам, подсчитывающим совпадающие n-граммы. Нейросетевые же метрики показали, что все модели справились с задачей сравнительно хорошо, однако абстрактивные модели — чуть лучше.

Стоит отметить, что все модели показали сравнительно низкие оценки ROUGE и BLEU, в то время как показатели BERTScore в целом высокие. Это говорит о том, что статистические подходы не совсем подходят для задачи суммаризации из-за отсутствия перефразирования, а большие языковые модели, напротив, подбирают синонимичные конструкции слишком вольно.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.9 – Результаты автоматической оценки работы моделей   | Модель | ROUGE-1 | ROUGE-2 | ROUGE-L | BLEU | BERT-Precision | BERT-Recall | BERT-F1 | | --- | --- | --- | --- | --- | --- | --- | --- | | GigaChat | 0,3405 | 0,0693 | 0,2028 | 0,0178 | 0,6709 | 0,6864 | 0,6782 | | Т5 | 0,3225 | 0,0612 | 0,1930 | 0,0145 | 0,6754 | 0,6771 | 0,6757 | | sumy | 0,2442 | 0,0348 | 0,1432 | 0,0063 | 0,6247 | 0,6573 | 0,6400 | |

В общей сложности 300 кратких содержаний от трех разных моделей прошли экспертную оценку по следующим критериям:

* охват (coverage) — упоминание важных сюжетных точек, метрика оценивает смысловую составляющую;
* достоверность (faithfulness) — наличие несуществующих в исходном тексте деталей или искажение истории;
* связность (coherence) текста;
* анализ — наличие верной интерпретации авторской позиции или тем рассказа.

По каждому из четырех пунктов эксперт выставляет оценку по шкале Лайкерта, где 1 — худший результат, а 4 — лучший. Также экспертам было предложено самостоятельно выбрать фрагменты суммаризации, где допущена ошибка и отнести ее к одному из четырех критериев, приведенных выше.

Из [таблица 6.10](#tbl-2) видно, что саммари, созданные человеком, несколько уступают большой языковой модели GigaChat в удобочитаемости текста и его аналитичности, однако имеют преимущество в передаче основных моментов сюжета и высокой достоверности. Краткие содержания модели GigaChat показывают самые высокие результаты в сравнении с другими моделями, но несколько уступают эталонным суммаризациям в охвате и достоверности, что не позволяет рассматривать этот инструмент как то, что может заменить пересказ человека, однако их преимущество заключается в высокой связности генерируемого текста и включения анализа в пересказ. Тексты модели T5, несмотря на внешнюю схожесть с пересказами эталона, о чем говорят высокие показатели автоматической оценки, значительно уступают другим моделям, совершая фактические ошибки в 74 рассказах из 100. Краткие содержания экстрактивной модели sumy отличаются сравнительно высокой достоверностью, однако обладают наименьшей связностью, аналитичностью и плохо передают основные события из текста оригинала, включая в повествование незначительные детали и упуская важные моменты для понимания.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.10 – Экспертная оценка кратких содержаний   | Модель | Охват |  | Достоверность |  | Связность |  | Анализ |  | Итог |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  | mean | sd | mean | sd | mean | sd | mean | sd | mean | sd | | GigaChat | 3,35 | 0,66 | 3,05 | 1,01 | 3,91 | 0,35 | 2,49 | 1,14 | 12,71 | 2,40 | | T5 | 1,94 | 0,93 | 1,48 | 0,85 | 2,71 | 1,11 | 1,32 | 0,53 | 7,45 | 2,48 | | Sumy | 1,62 | 0,72 | 3,39 | 0,92 | 1,68 | 0,68 | 1,09 | 0,41 | 7,7 | 2,06 | | Human | 3,55 | 0,69 | 3,69 | 0,76 | 3,73 | 0,49 | 1,75 | 1,03 | 12,72 | 1,99 | |

Чтобы выделить особенности художественного текста, влияющие на результат суммаризации, были рассмотрены такие показатели как коэффициент лексического многообразия, нарушение фабулы, наличие ненадежного рассказчика в тексте, смена лица повествователя (грамматическая и смысловая), включение флэшбеков и повествования о снах, а также вставка фрагментов других жанров — песен, стихотворений, анекдотов и др. Корреляционный анализ проводился путем подсчета коэффициента корреляции Пирсона.

Оценка структурной сложности рассказа складывается из пяти критериев и составляет от 0 до 5 баллов. Эксперты оценивали текст с опорой на 5 критериев. По каждому из критериев текст мог получить оценку 1 или 0. Так, в общей сложности нарративная структура произведения оценивалась максимум в 5 баллов и минимум в 0, где 5 — наиболее усложненное повествование, 0 — наиболее простое. В [таблица 6.11](#tbl-3) представлены результаты экспертной разметки структры рассказов на нарративную сложность.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.11 – Оценка нарративной сложности текстов   | Критерии | К1. Фабула | К2. Рассказчик | К3. Лицо | К4. Флешбеки | К5. Вставки других жанров | | --- | --- | --- | --- | --- | --- | | Кол-во рассказов | 51 | 38 | 16 | 53 | 18 | |

Корреляционный анализ позволяет сделать следующие выводы:

* TTR: повышение коэффициента TTR скорее положительно влияет на лексическую схожесть текстов модели и эталона в случае генеративных моделей (T5, GigaChat), однако несколько отрицательно — в случае экстрактивной модели (sumy). На передачу основной информации текста коэффициент TTR влияет положительно для модели T5, несколько улучшает качество текстов экстрактивной суммаризации и практически не имеет связи с результатом большой языковой модели GigaChat;
* лицо повествования (грамматический показатель): параметр практически не влияет на результат суммаризации, однако гипотеза о том, что перволичное повествование будет несколько затруднять пересказ модели частично подтвердилась, но только относительно лексической схожести сгенерированного и эталонного текстов. На содержательные качества саммари, такие как фактическая точность, полнота передачи сюжета, связность и наличие анализа перволичное повествование не влияет в случае модели GigaChat и слабо положительно влияет относительно моделей T5 и sumy;
* нарушение фабулы: нелинейное повествование слабо негативно влияет на качество суммаризации GigaChat, T5, особенно на фактическую точность текстов, но влияет слабо положительно на достоверность и логическую связность экстрактивной модели sumy. Этот параметр не влияет на лексическую схожесть текстов саммари и эталона;
* ненадежный рассказчик: этот параметр коррелирует с перволичным повествованием, так как одним из критериев определения ненадежного рассказчика является первое лицо повествования. В целом ненадежный рассказчик практически не влияет на результат модели GigaChat, однако несколько негативно влияет на лексическую схожесть саммари и эталона и связность предложений (r = -0,16; r = -0,16). Этот параметр, как и первое лицо, негативно влияет на лексическую схожесть саммари T5 (r = -0,14), в особенности sumy и эталона (r = -0,19), однако скорее положительно влияет на полноту передачи основной информации, связность текста и анализ в текстах этих моделей;
* смена лица рассказчика или повествования: этот параметр практически не влияет на результат работы моделей, однако имеет слабую отрицательную связь с автоматическими и экспертными оценками генеративных моделей GigaChat и T5;
* флешбеки: повествования о прошлом практически не влияют на результаты абстрактивных моделей GigaChat и T5, однако если и влияют, то скорее слабо негативно. В то время как для экстрактивной суммаризации вставки флешбеков, напротив, являются положительным параметром, хорошо влияющим на фактическую точность текста;
* вставки других жанров: наличие в тексте фрагментов других жанров практически не влияет на результат работы большой языковой модели, однако слабо негативно влияет на лексическую схожесть текстов T5, sumy с текстами эталона и слабо положительно — на полноту передачи основных событий текста и связность краткого содержания модели sumy.

Таким образом, получены данные о том, что существующие алгоритмы автоматической суммаризации не всегда справляются с задачей краткого пересказа художественного текста. Расхождения во влиянии разных параметров на результат суммаризации указывают на то, что они влияют на результат не так сильно, как обучающая выборка каждой конкретной модели. В перспективе исследования разработка нового инструмента суммаризации путем тонкой настройки моделей с использованием улучшенных версий эталонных кратких содержаний в качестве обучающей выборки.

## 6.4 Проблемные зоны и технологические решения в автоматической обработке стандартных и нестандартных речевых данных

### 6.4.1 Диаризация: оценка качества на материалах корпуса, обработка новых данных спонтанной устной речи КУРС

В 2025 сотрудники лаборатории продолжили эксперименты по оценке возможностей и качества инструментов автоматической диаризации, т.е. процесса выделения говорящих в потоке речи. В дальнейшем модель и соответствующие утилиты становились частью общего цикла обработки звуковых данных корпуса. Анализ включал в себя сравнение распознанного текста, автоматически определенных временных меток и дикторов с данными транскрипций корпуса КУРС (Корпус устной речи студентов), полученными вручную экспертами. В эталонную выборку вошли 195 эпизодов, аудиозаписей с повседневной речью, собранных по методике «речевого дня». Записи сформированы как наборы аудиофайлов, охватывающие весь день информанта — от пробуждения до сна — либо отдельные разговоры.

Экспертная разметка эталонной выборки представляла собой таблицу, где каждой реплике соответствуют уникальный идентификатор говорящего, краткое описание эпизода, название аудиофайла, время начала и продолжительность речи. Текст реплик содержал маркеры синтагматического и фразового членения, а также обозначения пауз и паралингвистических явлений [331]. Накладывающаяся речь записана в одну реплику, список одновременно говорящих разделен символом “@” или “#”. Тестовые автоматические данные в свою очередь включали в себя обозначение найденного говорящего, временную метку начала и окончания речи. Обозначение дикторов, формат таймкодов и пересечения реплик в речи отличают экспертную разметку от тестовой: в таблице от диаризатора идентификаторы говорящих повторяются для каждого из аудиофайлов, предлагаются временные метки начала и окончания фрагмента, одновременная речь разнесена по разным репликам с пересечением таймкодов.

Для тестирования качества форматы разметки были приведены к единому виду, чтобы устранить различия, не влияющие на смысловую нагрузку, но способные исказить метрики при сравнении гипотезы и референса. Так, была полностью удалена пунктуация, приведён регистр всех слов к нижнему, заменены «ё» на «е», числительные, как арабские, так и прописью, были заменены на маркер «числ». При этом при конвертации меток спикеров существовал риск ошибок, в результате которых диагностические параметры могли быть рассчитаны неверно. Сами пересечения реплик в экспертной разметке не подлежат преобразованию, т.к. извлечь таймкоды для отдельных говорящих в этом случае невозможно. Помимо прочего, встречаются редкие орфографические ошибки в эталонной разметке. Риски минимизировались дополнительными тестами для алгоритмов и ручной проверкой выровненных пар слов из экспертной и тестовой разметок.

Процесс обработки аудиоданных для тестирования состоял из четырех этапов: нормализации, диаризации и разделения аудио на фрагменты, распознавания и подсчета метрик. Вначале аудиофайлы были приведены к единому формату и необходимой для ASR частоте дискретизации, 16 кГц, была выровнена громкость. Далее проводилась диаризация аудио определение и сегментация фрагментов по говорящим. Полученные сегменты сохранялись в отдельные файлы и передавались модели распознавания. В конце по полученным расшифровкам подсчитывались диагностические параметры качества диаризации и распознавания.

В качестве метрик использовались стандартные параметры качества диаризации и распознанного текста [332]: DER (Diarization Error Rate), DWER (Diarization Word Error Rate), SAWER (Speech Attributed Word Error Rate) и собственно WER (Word Error Rate). Все четыре основываются на подсчете количества замен, пропусков и добавлений слов, DWER дополнительно указывает, какой процент слов от общего количества слов был определен неверно, т.е. отнесен не к тому спикеру. SAWER — количество неправильно распознанных слов в речи каждого из говорящих.

Для распознавания речи взята модель ASR Whisper Large v3 (ранее замеры качества и сравнение с НТР проводилось в [333]), для диаризации выбрана pyannote [334]: фреймворк с открытым исходным кодом и модель, демонстрирующая DER 34,4 и 24 на данных датасетов DIHARD и ETAPE.

Результат: Результаты расчётов метрики DER по референсной и тестовой выборке показывают значительное варьирование качества диаризации. Минимальное значение DER составило 10,3, в то время как максимальные значения стремятся к 100. При этом для 10% аудиозаписей ошибка диаризации не превышает 27,9. Среднее значение DER по всей выборке оказалось также достаточно высоким — 68,3, при этом медиана составила 52,6. Наилучшие результаты и относительная стабильность алгоритма достигаются в аудиозаписях с простыми условиями: нет лишних шумов, количество говорящих не превышает одного-двух, речь громка и отчетлива.

Средние значения параметров WER, DWER и SAWER находятся в диапазоне 46,1 до 64, медиана — от 46,2 до 62,9. Менее чем у 10% аудиозаписей WER составляет 27,4, DWER — 24,0, SAWER — 30,6. 90-й процентиль достигает для WER 62,6, DWER — 81,1, SAWER — 96,8. Максимальные значения достигают 100. Несмотря на то что минимальные значения (от 8,4 до 13,3) говорят о наличии немногих высококачественных распознаваний, в целом система, как и при диаризации, демонстрирует заметное варьирование качества.

К настоящему моменту с помощью моделей диаризации и ASR обработано 869 макроэпизодов КУРС.

### 6.4.2 Автоматическое распознавание атипичной речи на русском языке (на материале речи пациентов с афазией) : разработка бенчмарка

Одной из задач проекта стало сравнение инклюзивности и оценка качества распознавания атипичной (афазической) речи у моделей автоматического распознавания речи (АРР), которые доступны на российском рынке.

В качества материала исследования выступил датасет RuAphasiaBank, собранный в отчетном году сотрудниками лаборатории. Датасет RuAphasiaBank представляет собой корпус речевых записей, собранный для исследования афазической речи. Записи были получены в Центре нейрореабилитации ФСНКЦ ФМБА России (г. Красноярск). Все пациенты (или их доверенные лица), включенные в исследование, подписывали информированное согласие на анонимизированную запись голоса. Для проведения исследования было получено официальное разрешение этического комитета НИУ ВШЭ. Запись осуществлялась во время занятий с логопедом в стационаре. Расстояние между пациентом и логопедом составляло около 50 см, а между пациентом и звукозаписывающим устройством не более 25 см. Использовались стандартные диктофоны, встроенные в телефоны марки iPhone. Для соблюдения принципов конфиденциальности во время записи устройства были в офлайн-режиме. После записи передавались исследовательской группе на физических носителях, где проходили процедуру анонимизации: каждая запись прослушивалась, и любая информация личного характера (ФИО, дата и место рождения, место проживания) исключалась из записи с помощью специальных инструментов обработки звука. Каждой записи присваивался уникальный идентификационный номер, который представляет собой сокращенное описание метаданных записи (номер фонограммы в рамках датасета, номер спикера в рамках датасета, пол, возраст, диагноз, степень тяжести).

Как можно заметить в [таблица 6.12](#tbl-4) датасет включает 188 аудиозаписей, из которых 164 относятся к атипичной речи, а 24 — к речи нейротипичных информантов. Всего в выборке представлено 70 пациентов с различными формами афазии и 20 нейротипичных спикеров. Временной объем датасета составляет 9,6 часов звучания, при этом средняя продолжительность речевого материала, приходящаяся на одного пациента, составляет около 0,14 часа.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.12 – Основные характеристики датасета RuAphasiaBank   | Тип распределения/ Признак | По пациентам | По типу афазии | По степени нарушения | По типу речи | По типу текста | По нейротипичным информантам | | --- | --- | --- | --- | --- | --- | --- | | **Общее количество** | *Пациентов/ записей : 70 / 164* | *Пациентов/ записей :*КМА – 58/140ЭфА – 2/6Сем – 2/3Сенс – 5/10АМ – 1/1Т – 2/4 | *Пациентов/ записей:*Легкая – 5/12Средняя – 30/91Тяжелая – 35/61 | *Записей:*Монолог – 19Чтение – 27Пересказ – 18Слоги – 13 | Записей :Дидакт – 17Худож – 29Упрощ – 21 | *Информантов/ записей : 20/24* | | **Гендер (М и Ж)** | *Пациентов/ записей (кол-во):* 54/129 (М)16/35 (Ж) | *Пациентов M/Ж (кол-во):*КМА – 44/14ЭфА – 2/0Сем – 2/0Сенс - 3/2АМ – 1/0Т – 2/0 | *Пациентов M/Ж (кол-во):*Легкая – 3/2Средняя – 25/5Тяжелая – 27/8 | *Записей (кол-во) M/Ж :*Диалог – 39/12Монолог – 12/3Чтение – 10/4Пересказ – 8/3Слоги – 11/0 | *Записей (кол-во) M/Ж:*Дидакт – 14/3Худож – 21/8Упрощ – 9/12 | *M/Ж (кол-во):* 16/4 | | *Средний возраст пациентов/ информантов* | 53,4 М 61,3 Ж | КМА – 52,4ЭфА – 49,5Сем – 63Сенс – 65,4АМ – 62Т – 48 | Легкая – неизв.Средняя – 51,4Тяжелая – 57,7 | Диалог – 56,5Монолог – 50,8Чтение – 46,2Пересказ – 45,2Слоги – 53,7 | Дидакт – 45,8Худож – 44,4Упрощ – неизв. | 58,5 | | *Терапевт* | 8 | - | - | - | - | - | | *Время звучания (в час)* | Общее – 9,6 часСреднее на 1 пациента 0,14 час | КМА – 8,3 часЭфА – 0,31 часСем – 0,15 часСенс – 0,77 часАМ – 0,09 часТ – 0,20 час | Легкая – 0,55 часСредняя – 5,4 часТяжелая – 3,9 час | Диалог – 5,6 часМонолог – 1 часЧтение – 1,5 часПересказ – 0,85 часЧтение слогов – 0,62 час | Дидакт – 0,86 часХудож – 1,6 часУпрощ – 1,3 час | 4,3 час | |

В рамках данного исследования собранные данные использовались для проведения тестирования систем автоматического распознавания речи (АРР). Актуальность данного вопроса объясняется тем, что люди, страдающие афазией, остаются участниками речевой коммуникации с банками, медицинскими учреждениями, розничной торговлей и другими институциями, в повседневную жизнь которых уже вошли речевые технологии. Собранный датасет позволил оценить качество распознавания и инклюзивность движков АРР на русском языке, доступных как коммерческие сервисы от ведущих российских разработчиков. На собранных данных тестировались общедоступные демо-версии облачных сервисов АРР: Сервисы, оцененные в данном исследовании, включают: Yandex SpeechKit [335], SaluteSpeech [336] (приложение для компьютера), Shopot [337], T-bank [338]. Эти движки уже вошли в состав голосовых помощников, платформ для анализа речи и клиентского опыта и стали универсальными решениями преобразования звучащей речи в текст.

Для оценки качества работы систем АРР на собранных данных рассчитывалась WER (пословная ошибка распознавания) и CER (посимвольная ошибка распознавания). Для расчета метрик были вручную размечены эталонные тексты: 44 аудиозаписи пациентов 28-76 лет (31 мужчина и 8 женщин) с легкой и умеренной, преимущественно, комплексно-моторной афазией. Для разметки были отобраны три типа речи: монолог (11 записей), чтение вслух (17 записей) и пересказ (16 записей). В качестве контрольного подмножества данных вручную затранскрибировали 15 записей монологической речи 13 нейротипичных респондентов в возрасте 55-65 лет. Подробные расчеты WER и CER для каждой модели представлены в [таблица 6.13](#tbl-5).

В ходе экспериментов была выявлена неоднородность качества работы систем автоматического распознавания речи (АРР) на контрольном подмножестве и на различных типах речи. Наибольшая точность при распознавании монологической речи была зафиксирована у сервиса Yandex SpeechKit (медиана WER = 0,16 – 0,55). В то же время, система SaluteSpeech показала на данном типе речи наименьшие результаты (медиана WER = 0,34 – 0,72). Следует отметить, что для сервисов Yandex SpeechKit и SaluteSpeech была характерна тенденция к двукратному повышению точности распознавания монологов респондентов с афазией по сравнению с речью нейротипичных респондентов. Системы Shopot и T-Bank продемонстрировали эквивалентное качество АРР для монологов пациентов и контрольной группы. При этом сервис T-Bank показал ухудшение метрик для речи пациентов с афазией по сравнению с нейротипичной речью, что указывает на его пониженную устойчивость к речевым нарушениям, характерным для данного нарушения.

Модель АРР Yandex SpeechKit демонстрирует наиболее стабильное качество распознавания на монологах и записях контрольной группы, что подтверждается относительно низкими и плотно кластеризованными значениями WER и CER. Вместе с тем, на фонограммах с чтением и пересказом пациентов с афазией наблюдается значительный рост количества ошибок распознавания и их высокая вариабельность. Наибольшие затруднения у модели вызывает задача на чтение: максимальные значения WER превышают 100%, а показатель CER характеризуется широким разбросом.

Модель АРР SaluteSpeech демонстрирует более высокую точность обработки монологической речи лиц с афазией по сравнению с нейротипичными респондентами. При выполнении задания на чтение модель показала значения WER, сопоставимые с контрольной группой, однако распределение ошибок в контрольной группе было значительно более плотным. Наибольшие трудности у системы вызвало подмножество пересказа, где медиана WER достигла 72%. Следует отметить, что значительно более низкий уровень посимвольной ошибки (CER) в речи лиц с афазией по сравнению с контрольной группой указывает на снижение качества работы модели при обработке спонтанной, быстрой и лексически разнообразной нейротипичной речи. Выраженное расхождение между WER и CER в записях пациентов с афазией свидетельствует о среднем качестве распознавания на фонемном уровне (низкий CER), но о нарушенном лексико-семантическом сопоставлении (высокий WER). В противоположность этому, сбалансированное соотношение WER ≈ CER в контрольных записях, вероятно, отражает влияние нелингвистических факторов и сложности, связанные с распознаванием относительно быстрой спонтанной речи.

Модель АРР Shopot демонстрирует относительно низкий уровень ошибок при распознавании монологов (медиана WER = 29%). При этом монологи пациентов с афазией (WER = 29 – 59%) распознаются с сопоставимой или более высокой точностью по сравнению с речью контрольной группы (WER = 31 – 88%), что может свидетельствовать о специфической адаптации модели к медленной речи. Результаты контрольной группы характеризуются большей стабильностью (WER: 31–41% до 90-го процентиля). Вместе с тем, модель Shopot показывает более низкое качество при распознавании пересказа (медиана WER = 36%) и чтения (медиана WER = 38%). Наибольшие трудности система испытывает в заданиях на пересказ, где точность распознавания ниже, чем в заданиях на чтение, что указывает на проблемы при обработке несвязной речи пациентов. Стабильно более низкие значения CER по сравнению с WER на всех наборах данных позволяют сделать вывод о том, что модель обеспечивает точное распознавание на фонемном уровне, однако испытывает значительные трудности на уровне лексико-семантической интерпретации.

Модель АРР T-Bank демонстрирует оптимальные результаты при обработке монологов (WER = 30–52%) и речи в контрольной группе (WER = 28–38%). Несмотря на сопоставимую базовую точность для обеих подгрупп, записи пациентов характеризуются значительно большим разбросом ошибок (до 52% против 38%). Следует отметить, что контрольная группа сохраняет исключительную стабильность как по показателям WER (28–38%), так и CER (15–23%). Наихудшие результаты система T-Bank показывает при распознавании чтения (WER = 52–105%) и пересказа (WER = 43–98%). Наиболее проблемным сценарием является чтение: значение WER достигает 105% на 99-м процентиле, что свидетельствует о неприспособленности модели к обработке речевых данных подобной сложности. При этом показатель CER остается относительно низким (25–59%). Полученные данные позволяют предположить, что основная проблема модели заключается в распознавании целых слов, а не в интерпретации на уровне фонем. Точность распознавания пересказа (медиана WER = 43%) является средней, превосходя результаты по чтению, но уступая показателям для монологов. В качестве итога представлены лучшие результаты по типам речи: в [таблица 6.13](#tbl-5) наименьшие проценты ошибок для каждой категории выделены жирным шрифтом.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.13 – Метрики качества российских движков АРР, полученные на датасете RuAphasiaBank (% — процентиль)   | Движок АРР | Монолог (50/99%) | Чтение (50/99%) | Пересказ (50/99%) | Контрольная группа, монологи (50/99%) | | --- | --- | --- | --- | --- | | Yandex SpeechKit | WER: 0.16/0.49CER: 0.1/0.31 | WER: 0.55/1.11CER: 0.28/0.81 | WER: 0.49/0.98CER: 0.24/0.55 | WER: 0.42/0.48CER: 0.26/0.3 | | SaluteSpeech | WER: 0.34/0.59CER: 0.19/0.41 | WER: 0.60/1.00CER: 0.23/0.51 | WER: 0.72/1.00CER: 0.20/0.47 | WER: 0.64/0.71CER: 0.26/0.69 | | Shopot | WER: 0.29/0.59CER: 0.19/0.53 | WER: 0.38/0.80CER: 0.2/0.81 | WER: 0.36/0.88CER: 0.22/0.58 | WER: 0.31/0.88CER: 0.18/0.74 | | T-bank | WER: 0.30/0.52CER: 0.15/0.31 | WER: 0.52/1.05CER: 0.25/0.59 | WER: 0.43/0.98CER: 0.26/0.74 | WER: 0.28/0.38CER: 0.15/0.23 | |

Анализ метрик, представленных в [таблица 6.13](#tbl-5), позволяет выявить характерные особенности каждого движка АРР. Yandex SpeechKit демонстрирует наибольшую точность при обработке монологов пациентов с афазией. Модель Shopot показывает самый низкий уровень ошибок по метрике WER в наиболее сложных для распознавания сценариях — чтении и пересказе. В то же время SaluteSpeech лидирует в этих же сценариях по метрике CER, причем для записей пациентов с афазией у данной модели наблюдается выраженное расхождение между значениями WER и CER. T-Bank превосходит остальные системы в распознавании спонтанной разговорной речи нейротипичных информантов, а также в целом демонстрирует стабильное качество при обработке различных типов речи.

Все рассмотренные системы АРР демонстрируют значительное снижение качества распознавания в заданиях на чтение и пересказ, что подтверждается ростом показателя WER при стабильно низких значениях CER. Данная картина указывает на системные проблемы, связанные скорее с лексической интерпретацией, чем с ограничениями акустического распознавания. При обработке спонтанной речи проявляется общая закономерность: качество распознавания речи контрольной группы не всегда превышает результаты для атипичной речи пациентов с афазией. В частности, все системы АРР, за исключением T-Bank, показывают более высокий процент ошибок при распознавании речи в контрольной выборке, состоящей из записей монологов нейротипичных информантов, выполненных в идентичных технических условиях. Наблюдаемый тренд позволяет выдвинуть предположение о том, что современные системы АРР на российском рынке могут быть непреднамеренно переобучены на гиперартикулированной речи, что затрудняет корректную обработку естественной разговорной речи даже в отсутствие речевых патологий.

Собранный и размеченный датасет RuAphasiaBank является уникальным ресурсом, необходимым для мониторинга качества систем АРР для русского языка, поскольку в известном датасете AphasiaBank [339] русский язык не представлен, а сущесвующие датасеты для русского языка меньше по объему и более специфичны по выборке [340]. Его использование позволило оценить системы АРР при распознавании разных видов атипичной речи, а также естественной разговорной монологической речи пользователей возрастного диапазона 55 – 66 лет. В результате проведенного тестирования мы получили два, на наш взгляд, интересных наблюдения.

Во-первых, в записях пациентов с афазией такие виды речи, как пересказ и чтение, несмотря на свой репродуктивный, а не продуктивный характер, оказались более трудными для АРР, чем монолог. Это связано с тем, что большинство использованных записей чтения и пересказов принадлежит пациентам с комплексной моторной афазией, которые испытывают трудности с планированием высказывания на основе нового содержания, и с артикуляцией, у них затруднен выбор артикулем, переход от одной моторной программы к другой. В монологе те же пациенты часто прибегают к использованию автоматизированных речевых рядов (*первое, второе, третье и компот*), устойчивых речений (*дела как сажа бела*), клише (*меня зовут*) произнесение которых для пациента облегчено речевой привычкой.

Во-вторых, неожиданным оказалось то, что три из четырех протестированных систем показывают на атипичных монологах даже лучшие результаты, чем на типичной разговорной монологической речи информантов старше 50 лет. Данный факт связан, по-видимому, с тем, что в разговорной речи в естественных условиях у информантов наблюдается стремление к экономии речевых усилий, приводящее к разнообразным фонетическим редукциям и усечениям форм, и эта тенденция с возрастом усиливается. Модели же обучены, в основном, на речи дикторов, актеров, имеющих профессиональную привычку к четкой, даже утрированной, артикуляции. У пациентов с афазией как раз наблюдается стремление к утрированной артикуляции, хотя при этом, довольно часто, снижена сила голоса.

Следует отметить, что некоторые из рассмотренных систем дают для отдельных видов речи относительно высокие метрики качества, но при изменении типа речи их показатели резко ухудшаются. Иными словами, говорить об инклюзивности моделей, попавших в тестовую выборку, пока сложно даже применительно к речи нейротипичных пользователей.

В перспективе созданный датасет RuAphasiaBank позволит осуществлять регулярный мониторинг инструментов АРР, используемых в России коммерческими и государственными организациями, формируя репрезентативные бенчмарки для оценки качества распознавания афазической речи.

РИД: База данных «RuAphasiaBank, база данных фонограмм речи пациентов с афазией». Свидетельство о государственной регистрации базы данных: № 2025623465.

Вид РИД: база данных. Авторский коллектив:

* Колмогорова Анастасия Владимировна (заведующий Лабораторией языковой конвергенции);
* Явшиц Екатерина Валерьевна (младший научный сотрудник);
* Сугян Анна Хачатуровна (стажер-исследователь);
* Сергеева Мария Олеговна (стажер-исследователь).

База данных представляет собой структурированную коллекцию аудиозаписей речи и сопутствующих метаданных. Состав РИД включает:

* Набор аудиофайлов в формате .wav с записями речи;
* Текстовый файл в формате .csv, содержащий подробные метаданные для каждого аудиофайла.

### 6.4.3 Объектный сентимент-анализ отзывов на культурные институции с помощью больших языковых моделей (на материале датасета отзывов посетителей Владимиро-суздальского музея-заповедника)

В рамках проекта решалась задача проверки эффективности больших языковых моделей (LLM) как инструмента для аспектно-ориентированного анализа тональности (ABSA) в условиях отсутствия предопределенного набора целевых аспектов и присущей данным политематичности.

Впервые в качестве данных для сентимент-анализа выступили тексты отзывов на культурные институции. Традиционным объектом применения методов сентимент-анализа являются отзывы на товары и услуги, однако постепенно формируется относительно новый тип отзывов – отзывы посетителей учреждений науки и культуры. У подобных текстов существует своя специфика, в значительной мере отличающая их от других оценочных субжанров. В частности, подобные тексты совмещают в себе черты не только оценочных жанров, но и рефлексивов, нарративов о жизни и т.д.

Исследование было подчинено задачам, поставленным заказчиком – Владимиро-Суздальским музейным комплексом. Необходимо было собрать датасет отзывов с существующих цифровых платформ и автоматически выявить, что нравится, а что не нравится посетителям музейного комплекса. При этом требовалось сфокусироваться на динамике, т.е. сохранить тренды по годам (с 2020 по 2025).

В качестве материала исследования использовались 12,187 отзывов из 9 публично доступных онлайн-ресурсов на 15 объектов Владимиро-Суздальского музея-заповедника в период 2020-2025 гг. Отзывы были получены путем веб-скрейпинга (автоматического сбора текстовых данных из интернета). Характеристики корпуса отзывов указаны в [таблица 6.14](#tbl-6).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.14 – Распределение отзывов по категориям   | Музейным объектам | N | По платформам-источникам | N | По годам | N | | --- | --- | --- | --- | --- | --- | | Суздальский Кремль | 2527 | Yandex Maps | 6305 | 2020 | 2400 | | Спасо-Евфимиев монастырь | 2218 | Google Maps | 5090 | 2021 | 1643 | | Музей деревянного зодчества | 1888 | Tripadvisor | 523 | 2022 | 2373 | | Музей хрусталя | 781 | Otzovik | 93 | 2023 | 2013 | | Успенский собор | 690 | 2gis | 85 | 2024 | 3467 | | Музей Мальцовых | 673 | Fooby | 40 | 2025 | 291 | | Исторический музей | 663 | Autotravel | 25 |  |  | | Димитриевский собор | 620 | Irecommend | 23 |  |  | | Палаты | 505 | Tonkosti | 3 |  |  | | Церковь Бориса и Глеба | 583 |  |  |  |  | | Золотые ворота | 403 |  |  |  |  | | Музей природы | 344 |  |  |  |  | | Дом-музей Столетовых | 271 |  |  |  |  | | Музей «Старый Владимир» | 40 |  |  |  |  | | Музей-усадьба В. Храповицкого | 21 |  |  |  |  | |

Для выполнения задачи использовались следующие методы:

1. Веб-скрейпинг
2. Автоматическая предобработка данных (формирование таблиц, очистка от пустых отзывов, сортировка по годам)
3. Аспектно-ориентированный сентимент-анализ
4. Применение больших языковых моделей (запуск моделей через llama-cpp-python локально)
5. Стратегический промпт-инжиниринг

Разработан пайплайн обработки отзывов для автоматического формирования отчетов по позитивным и негативным категориям объекта по годам. Пайплайн представлен в [рисунок 6.9](#fig-spb_2). Модель вызывается 2 раза: сначала ей дается промт с инструкцией извлечь негативные и позитивные ключевые слова из отзывов по каждому объекту и по каждому году; затем модель получает инструкцию разделить все полученные на первом этапе ключевые слова на предзаданные обязательные и факультативные категории. Среди обязательных категорий: экспозиция, персонал, месторасположение, еда и туалеты, цены, внешний вид объектов и территории; среди факультативных: посещение с детьми, условия для людей с ограниченными физическими возможностями, эмоции и атмосфера, общее впечатление, знания и образование, развлечения и шоппинг, доступность (как добраться до места), история и патриотическое воспитание. Под ключевым словом в данной работе подразумевался минимальный фрагмент текста, имеющий предикацию – это могло быть адъективно субстантивное сочетание (*увлекательная экскурсия*), а могло быть простое предложение с глагольной группой (*учитель прошел бесплатно*).

|  |
| --- |
| Рисунок 6.9 – Пайплайн обработки отзывов. |

В качестве используемой большой языковой модели была выбрана модель YandexGPT5-Lite-8B-instruct-Q8\_0[[4]](#footnote-336), показавшая наилучший результат для выбранной задачи. Характеристики модели представлены в [таблица 6.15](#tbl-7).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.15 – Параметры модели YandexGPT   | Параметр | Значение | | --- | --- | | Количество параметров | 8 миллиардов | | Базовая архитектура | Llama | | Квантизация | Q8\_0 (8 bit) in GGUF format | | Размер модели | 8.54 GB | | Максимальный размер контекста | 32K | | Совместимость | Совместима с llama.cpp | |

В результате обработки были сформированы 15 таблиц с позитивными и негативными ключевыми словами по каждому объекту, а также 15 текстовых отчетов, где ключевые слова были распределены по позитивным и негативным категориям с помощью большой языковой модели. Отчеты были проверены экспертно, количество галлюцинаций модели (неправильное отнесение к категориям, логические, синтаксические ошибки, несуществующие ключевые слова) составили не более 7% от общего числа ключевых слов. Пример итогового результата можно увидеть в [рисунок 6.10](#fig-spb_3).

|  |
| --- |
| Рисунок 6.10 – Пример текстового отчета по категориям. |

Категориями с самым большим числом позитивных ключевых слов стали Экспозиция, Общее впечатление, Эмоции и атмосфера ([рисунок 6.11](#fig-spb_4)). Категориями с самым большим числом негативных ключевых слов: Экспозиция, Цены, Общее впечатление ([рисунок 6.12](#fig-spb_5)).

|  |
| --- |
| Рисунок 6.11 – Тематические категории с наибольшим числом позитивных ключевых слов. |

|  |
| --- |
| Рисунок 6.12 – Тематические категории с наибольшим числом негативных ключевых слов. |

Результат: разработан и апробирован эффективный пайплайн для аспектного сентимент-анализа отзывов на культурные институции с помощью больших языковых моделей. Заказчику предоставлены данные в динамике по 15 объектам за требуемый период времени. На основании полученных результатов ведется работа по описанию коллективного портрета современного посетителя музея.Разработанный метод может быть экстраполирован на другие культурные институции.

### 6.4.4 Конвергентные речевые практики русскоязычного сегмента Интернет (на материале корпусных данных)

Задача данной части проекта – исследование цифровыми методами распространенных речевых и дискурсивных практик, используемых носителями русского языка в социальных сетях русскоязычного сегмента сети Интернет. В фокусе работы в отчетном году — дискурсивные практики зообъявлений и гороскопов.

Согласно Дж. Берджеру, первой древний союз человека и животного пошатнула декартовская идея l’animal-machine, а окончательно он был разрушен в эпоху позднего капитализма [341]. Нечеловеческие животные оказались вытеснены из центра человеческого опыта на периферию культуры: в книги, мультфильмы, иллюстрации, детские игрушки и зоопарки. Квартира так же, как и зоопарк, лишает животного автономии и ограничивает его естественные потребности. Домашнее животное служит антропоцентричным желаниям хозяина, наряду с умной бытовой техникой [342]. Оно должно быть/стать милым, удобным, неприхотливым и полезным, даже если это противоречит его естественным потребностям.

Целью нашего исследования является выявление схем репрезентации «идеального» питомца на «рынках животных» в социальных сетях. С помощью корпусного анализа предпринимается попытка определить особенности речевого жанра «зоообъявление», который, несмотря на свою распространенность, ранее не подвергался подобному количественному описанию.

В качестве материала используются тексты из групп «ВКонтакте», которые занимаются публикацией зоообъявлений, направленных на мотивацию подписчика принять некоторое животное «в добрые руки». Обращение к Интернет-сообществам именно помощи животным обуславливается стремление проследить лексические особенности, которые подобные тексты, несмотря на их бытование в функционально немаркетинговом информационном поле, наследуют из рекламного дискурса [343]. Кроме того, интерес представляют дискурсивные паттерны, которые используются при представлении животных разных видов для потенциальных хозяев. В соответствии с этой задачей были собраны посты из двух групп: одна специализируется на публикации объявлений о котах/кошках, вторая — о собаках. Объем каждого из корпусов составил 22291 пост (2301189 словоупотреблений) и 22783 поста (2255550 словоупотреблений) соответственно.

Частотный, коллокационный и анализ эмоций позволили выделить устойчивые схемы репрезентации животных, использующиеся в зоообъявлениях для создания привлекательного образа потенциального питомца.

Наиболее характерной для конструирования образа как кошки, так и собаки оказывается лексико-тематическая группа «принятие в дом/семью» (*дом*, *человек*, *искать*, *семья*, *хозяин*). Однако отметим, что частотной в cat\_c оказывается лемма «*дом*» (9838[[5]](#footnote-352), ср. dog\_c: 8168), в то время как в dog\_c обнаруживается высокое ранговое положение другой леммы — «*семья*» (dog\_c: 5807, cat\_c: 3504). На обнаруживаемое речевое выражение ментального стереотипа «собака — друг человека» указывает также и то, что лемма «*компаньон*» используется при описании собаки практически в два раза чаще (532), чем при описании кота/кошки (275).

В объявлениях о собаках также оказывается важным указание на породу собаки (например, *хаски* — 2122, *овчарка* — 1660, *метис* — 1283). В этом проявляется внимание некоторых людей к вопросу «чистоты» породы: породистые собаки находятся выше в иерархии, чем метисы и беспородные.

Кошка же, по всей видимости, рассматривается как атрибут домашнего уюта, подчеркивается ее домашность (лемма «*домашний*» в cat\_c и dog\_c соответственно: 1647 vs. 780) и ласковость (лемма «*ласковый*» в cat\_c и dog\_c соответственно: 4869 vs. 2360). Кроме того, кошек чаще, чем собак, именуют «*малыш*» (cat\_c: 1799 vs. dog\_c: 738), «*малышка*» (1494 vs. 524). В отличие от собаки, образ кошки в данном дискурсе отличается меньшей автономностью.

Тема чистоты является актуальной для реализации образа как собаки, так и кошки, однако частотно значимый разрыв обнаруживается именно для последних (ср., лемма «*паразит*» встречается в cat\_c и в dog\_c соответственно: 1212 и 557). Вероятно, это связано с преимущественно «indoors» условиями содержания котов, в отличие от собак. Активное и даже неизбежное взаимодействие собак с внешней средой смягчает требования к их гигиене (см. [рисунок 6.13](#fig-spb_6) и [рисунок 6.14](#fig-spb_7)).

|  |
| --- |
| Рисунок 6.13 – Коллокаты леммы «кошка» |

|  |
| --- |
| Рисунок 6.14 – Коллокаты леммы «собака» |

Обратимся к результатам автоматического детектирования эмоций, для которого использовалась модель rubert-tiny2-russian-emotion-detection[[6]](#footnote-361). Так, в корпусах выделяются эвокативные структуры, характеризуемые следующими эмоциями: *любовь* (cat\_c: 43,35%[[7]](#footnote-363), dog\_c: 32,84%), *печаль* (cat\_c: 10,21%, dog\_c: 10,11%), *восхищение* (cat\_c: 9,86%, dog\_c: 12,49%), *любопытство* (cat\_c: 4,45%, dog\_c: 3,9%), *благодарность* (cat\_c: 3,88%, dog\_c: 2,1%), *забота* (cat\_c: 2,91%, dog\_c: 6,19%), *радость* (cat\_c: 1,88%, dog\_c: 2,93%), *одобрение* (cat\_c: 0,1%, dog\_c: 1,42%).

Наиболее выраженной оказалась эмоция *любовь* (cat\_c: 43,35%[[8]](#footnote-364), dog\_c: 32,84%): так, в текстах конструируется эмоциональное состояние, которое должна вызывать ситуация адопции животного. Интересно, что эмоция *восхищение* в большей мере ассоциирована с зоообъявлениями о собаках (cat\_c: 9,86%, dog\_c: 12,49%). Действительно, в этих постах выделяются характеристики животного, которые призваны восхитить потенциального хозяина, в первую очередь связанные со спортивными достижениями породистых собак (*имеет первые медали*) или чертами внешности (*яркий красавец*, *классный*, *просто шикарный*).

Астрологический прогноз — другой особый жанр, получивший широкое распространение в медийном пространстве постсоветской действительности Такие особенности гороскопического текста, как функциональная направленность, рекомендательный характер, а также специфика выражения модальности и оценки, вызывают лингвистический интерес и описываются в ряде исследований [344; 345]. Неслучайно характерные признаки названных текстов изучаются и как черты медиажанров [346]: гороскопы встречаются в журналах, газетах, распространяются по радио, публикуются на форумах в сети Интернет.

Цель следующего этапа исследования — проследить особенности лексического состава текстов, в том числе проверить, проявляются ли архетипические черты в астрологических прогнозах и представляют ли гороскопы отдельный речевой жанр. Под гороскопами будут пониматься ежедневные предсказания, то есть тексты, которые предписывают то или иное поведение или пророчат некоторые события в определенный день согласно знаку зодиака человека.

Материалом исследования выступили гороскопы, размещенные в открытых группах «ВКонтакте», которые занимаются регулярной публикацией ежедневных астрологических прогнозов для каждого из 12 знаков зодиака с 2014 по 2025 гг. (включительно). Пример выгрузки, за исключением некоторых не используемых в рамках этого исследования метаданных (количество лайков, комментариев, репостов, просмотров, оставленных пользователями), представлен в [таблица 6.16](#tbl-8). В «сыром» виде полученный датасет насчитывал 90642 строки (около 17 тыс. строк на каждый знак зодиака). Далее осуществлялась очистка данных в несколько этапов. Так, были удалены тексты, не являющиеся гороскопами вида «ежедневное предсказание»: рекламные и поздравительные посты, гороскопы на неделю, год и иные временные промежутки, а также гороскопы, составленные не в соответствии с зодиакальной системой (например, кельтские гороскопы). Затем автоматически были «развернуты» посты, которые в одном тексте включали несколько видов, — общий, бизнес- и любовный гороскопы.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.16 – Пример выгрузки (фрагмент датасета)   | ID | Заголовок | Текст | Год публикации | Знак зодиака | | --- | --- | --- | --- | --- | | 1436 | Любовный гороскоп на 16 марта 2025 года | Спокойствие, только спокойствие — так проще достичь желаемого! | 2025 | Стрелец | | 4911 | Гороскоп на 30 мая 2024 года | Нынче у вас должна обостриться интуиция. Ей и следуйте. | 2024 | Лев | | 31540 | Гороскоп на 25 апреля 2018 | Наиболее трудные дела пока отложите. | 2018 | Рак | | 33931 | Бизнес-гороскоп на 01 октября 2017 | Будьте осторожны в финансовых делах. Возможны происки неожиданных конкурентов. | 2017 | Телец | | 46142 | Общий гороскоп на завтра, 9 сентября, 2016 | Вы прекрасный исполнитель и способны многого добиться. Сегодня Вам предстоит работать с общественностью. | 2016 | Козерог | |

После удаления «мусорных» постов и гороскопов-дублей (без учета знака зодиака), токенизации и лемматизации, выполненных с помощью MyStem[[9]](#footnote-366), а также сегментации на предложения с использованием библиотеки NLTK [347], объем корпуса составил 28149 строк или 1185425 словоупотреблений (15578 лексических типов). Средний размер группы гороскопов, приходящихся на каждый знак зодиака, равен 2346 текстам со средним суммарным объемом 99427,42 токена. Текст гороскопа как единица анализа имеет следующее лингвостатистическое описание (в средних значениях): длина в словах — 42,39 (σ — 13,97), длина в предложениях — 3,42 (σ — 1,18).

В исследовании используются частотный, коллокационный и сентимент-анализ.

Рассмотрим лексические особенности гороскопических текстов. На примере наиболее релевантных сочетаний слов прослеживаются прагматико-дискурсивные черты подобного текста: предсказание событий на завтра (*день*, *завтра*), ориентация на положительный исход дня (*благоприятный*, *день*), а также межличностные отношения и рекомендации по их организации, в частности через открытость к новым встречам и знакомствам (*человек*, *новый*, *знакомство*, *отношение*).

Далее обратимся к [рисунок 6.15](#fig-spb_8), на котором для каждой «зодиакальной» выборки визуализированы по 50 наиболее частотных содержательных и не входящих во множество общих для всего корпуса лексем.

|  |
| --- |
| Рисунок 6.15 – Облака частотных слов для каждого подкорпуса (исключая стоп-слова и общие для всего корпуса) |

Например, в облаке слов, характеризующем Овна заметен фокус на карьеру (*шеф*, *торговля*, *совещание*) и советы «заботиться» (возможно, из-за стереотипов недостатка созидательности у представителей этого знака зодиака). Значительное место занимает лексика, связанная со знаниями, их получением и применением (*образование*, *изучение*, *философский*). В отношении Тельца заметна «предостерегающая» лексика (*угроза*, *агрессивный*), а также связанная с работой — примечательно, что появляется фигура подчиненного (*работник*, *офис*, *совещание*), а также материальный аспект (*вложение*). Вновь появляется указание на сферу взаимоотношений (*привязанность*, *заботиться*). Тема работы наблюдается и в частотных распределениях гороскопов, адресованных Козерогу (*перегрузка*, *масштаб*, *нервничать*), что согласуется с образом «трудяги», через который часто описывается этот знак. Для Льва тоже отмечается лексика, связанная с работой (*шеф*, *офис*, *специалист*), однако еще и с указанием на такие, действительно, приписываемые ему черты, как успешность в карьере и лидерские качества (*победитель*, *руководить*).

Несмотря на отмечаемые в некоторых случаях различия, проведенный анализ указывает на необходимость обращения к гороскопу как к единому дискурсивному полю, отличающемуся слабо выраженной дистинктивностью (по крайней мере, если в качестве дифференцирующего параметра рассматривать знак зодиака). В связи с этим мы провели еще один эксперимент, а именно сравнили частотные распределения лемм в пользовательском корпусе, описанном в этом исследовании (и условно названном «гороскопы в соцсетях»), и в подкорпусе гороскопов Национального корпуса русского языка (НКРЯ)[[10]](#footnote-371). Подкорпус «гороскопы» НКРЯ насчитывает 130 текстов суммарным объемом 105877 словоупотреблений. Как позволяют судить метаданные, гороскопы, включенные в НКРЯ, были составлены в период с 1991 по 2010 гг. (включительно) и представляют собой принципиально другой вид медиа. Так, их источниками выступают журналы, газеты, а также, что немаловажно, встречаются авторские астрологические прогнозы.

В результате сравнения верхней зоны частотных словарей стала заметна ориентированность гороскопов, публикуемых в социальных сетях, на прогнозирование событий завтрашнего дня, в то время как гороскопы в журналах и газетах, что несколько ожидаемо, охватывали конкретный *день* или *неделю* целиком. Объяснение находится в особенностях сферы функционирования этих текстов, а также в исключении из нашего корпуса не ежедневных предсказаний. Кроме того, кажется, что предсказания из соцсетей стремятся быть более личными, отсутствие чего в подкорпусе гороскопов из НКРЯ, вероятно, является метой ограничений, накладываемых на периодические издания, их внешней модерируемостью.

В обоих корпусах заметна тема рабочих отношений. Однако в подкорпусе НКРЯ лемма *работа* имеет ранг 24, а в корпусе «гороскопов в соцсетях» — 50. Похожая тенденция на «опущение» по рангу заметна и для некоторых других лексем: *здоровье* — 31 vs. 234, *любовь* — 34 vs. 87, *деньги* — 37 vs. 289. По всей видимости, для ежедневных онлайн-предсказаний, хоть и важна, но в меньшей мере свойственна привязка к какому-либо конкретному аспекту жизни, в результате чего они представляются более общими и несколько абстрактными. В то время как для публикуемых в официальных СМИ гороскопах характерна выраженная категоризация «по сфере действия». Следующим шагом этой работы стало выявление особенностей эмоционального воздействия гороскопических текстов, определяемых на лексическом уровне. Для этого с использованием русскоязычного словаря оценочной лексики RuSentiLex [348] и программного обеспечения Orange[[11]](#footnote-372) был проведен анализ тональности. Расчет оценки тональности в настоящем исследовании производился в соответствии с [349; 350] по следующей формуле:

,

гдe pos — количество положительных слов (в типах), neg — количество отрицательных слов (в типах), а document\_length — длина документа (количество слов в тексте).

В результате были получены следующие распределения по количеству текстов: положительные — 12093 (42,96%), нейтральные — 10788 (38,32%), негативные — 5268 (18,71%), что согласуется с заключениями, представленными в [351]. Так, и в нашем случае, кажется, что лексико-эмоциональное наполнение текстов имеет целью вызвать у читателя положительные эмоции: спокойствие, надежду, стабильность. Это гарантируется чтением предсказания, что функционально выступает действием эквивалентным возможности контролировать свое будущее. При этом, как видно исходя из [рисунок 6.15](#fig-spb_8), знак зодиака не влияет на эмоциональную направленность текста, что вновь свидетельствует о генерализованном характере астрологических предсказаний.

Приведем несколько примеров положительных (216)–(217), негативных (218) и нейтральных предсказаний (219)–(220):

1. *Сегодня вы одержите пусть небольшую, но победу. Этот день принесет вам уважение окружающих.*
2. *Сегодня вы просто обречены на успех в амурных делах. Не играйте в карты!*
3. *Сегодня вы рискуете допустить ошибку или в чем-то просчитаться. Вам могут предоставить недостоверную информацию, так что будьте бдительны.*
4. *Сегодня лучше действовать сообща. Одному столько не выпить.*
5. *На работе не будет происходить ничего особенного. Будете выполнять рутинные обязанности и мелкие поручения.*

Интересным показалось посмотреть, какое выражение принимает эмоциональная компонента в зависимости от цели текста: гороскопы общего характера превалируют (58,01%), затем следуют любовные гороскопы (32,01%) и бизнес-гороскопы (9,98%). Положительная лексическая тональность главным образом характерна для прогнозов в сфере любви. Обратившись к долям постов, сгруппированных по гороскопическим поджанрам и классам тональности, удалось подтвердить это наблюдение: так, 46,77% любовных гороскопов и 42,22% общего характера вне зависимости от знака зодиака имеют положительную «заряженность». Наиболее негативными, составляя 21,58% постов своей подгруппы, в свою очередь являются бизнес-гороскопы (вероятно, так проявляется свойственный гороскопам этого поджанра мотив *предостережения о неприятностях на работе*), а наименее — вновь любовные (доля отрицательных постов внутри этой категории — 16,54%).

Частотный анализ позволяет проследить некоторую лексику, тематически соответствующую архетипическим образам знаков зодиака, однако из-за высокой генерализованности предсказаний (наблюдается значительное количество повторений) и слабых различий между тематическими группами этот анализ скорее говорит о том, что гороскопы имеет смысл рассматривать как некоторый речевой жанр, но с мало выраженной внутренней различимостью, по крайней мере по параметру адресата, выраженного представителем определенного знака зодиака. Представляется, что гороскоп в современном медиаполе функционирует как «аффирмирующее послание», на что указывает также и превалирующая положительная тональность этих текстов, но не как персонализированная рекомендация.

## 6.5 Внедрение результатов исследовательской работы по изучению цифровыми методами конвергентных процессов в современной русской речевой практике

### 6.5.1 Лингвистические цифровые ресурсы и прикладные разработки

#### 6.5.1.1 Модернизация и совершенствование системы векторного поиска картин по пользовательскому описанию

В рамках разработки поисковой функции для выбора картин Эрмитажа через чат-бот в 2025 г. был модернизирован существующий конвейер обработки запросов. Объем коллекции картин, доступных для поиска в боте, оставлен неизменным, 2000 картин; основное внимание было уделено улучшению качества и предсказуемости работы самой поисковой подсистемы при сохранении или снижении нагрузки на вычислительные ресурсы.

Первоначально функция поиска имела микросервисную архитектуру и опиралась на несколько сервисов Yandex Cloud (Yandex.Embeddings, Yandex.OpenSearch) . В ходе модернизации архитектура была упрощена: поисковая логика была целиком перенесена в технологию виртуального ассистента, Yandex AI Assistant, реализующего подход Retrieval-Augmented Generation (RAG) через sdk подключение, [рисунок 6.16](#fig-spb_9).

|  |
| --- |
| Рисунок 6.16 – Текущая архитектура поисковой функции |

В качестве базы знаний ассистента использованы текстовые описания картин и их метаданные, собранные с сайта «Цифровая коллекция Эрмитажа». Каждый документ был проиндексирован и получил уникальный идентификатор, используемый на этапе поиска и сопоставления с пользовательским запросом.

Каждая пользовательская сессия (тред, threads) в виртуальном ассистенте хранит историю запросов, промежуточные результаты поиска и рассуждения модели. Это позволяет оперативно отслеживать ошибки и аномалии в работе бота. Ранее для этих целей использовалась отдельная база данных для логирования действий пользователя, однако такой подход сопровождался задержками обновления и усложнял диагностику. Переход к механизму тредов обеспечил практически мгновенное обновление истории взаимодействий.

Дополнительно, благодаря сохранению контекста диалога в треде, поисковая функция получила возможность учитывать предыдущие реплики пользователя. Это позволяет уточнять исходный запрос, опираться на уже найденные результаты и использовать диалоговый контекст для более точного подбора картин.

Внесенные изменения позволили существенно снизить нагрузку на вычислительные ресурсы без ухудшения пользовательского опыта. Для команды разработки стала доступна удобная инфраструктура для анализа промежуточных результатов работы модели и оперативного реагирования на возможные ошибки.

Кроме того, расширение поиска за счёт учёта контекста диалога создает предпосылки для дальнейшего повышения релевантности выдачи и улучшения восприятия системы пользователями.

#### 6.5.1.2 Динамика роста корпуса русского рассказа

Корпус русского рассказа XX века — филологический цифровой ресурс, разрабатываемый на базе Корпуса русского рассказа 1900-1930 гг., целью создания которого является популяризация и изучение русской малой прозы. Корпус включает репрезентативную коллекцию аннотированных текстов и охватывает творчество широкого круга писателей XX века — от классиков до малоизвестных авторов. В 2025 году основная работа по корпусу состояла в завершении вычитки и корректуры 1000 текстов представительной выборки 1006 писателей, творивших на протяжении столетия. Также проводилась доразметка по всем принятым в корпусе принципам аннотации Корпуса-240 для последующей публикации результатов. Наконец, была начата инвентаризация всего корпуса с целью систематизации материалов.

В 2025 году велась работа над завершением оцифровки, вычитки и корректуры электронных версий рассказов 1000 авторов, творивших на протяжении всего XX века. Составленный в прошлом году представительный список 1000 русских прозаиков снова претерпел изменения, поскольку выяснилось, что 12 текстов не соответствовали жанру рассказа, описывали дореволюционные события, были написаны в эмиграции, адресованы детской аудитории, что противоречит принципам составления корпуса [352; 353]. После изменений список насчитывает 1006 персоналий (887 мужчин и 119 женщин), так как 6 рассказов написано в соавторстве [354]. На данный момент выборка полностью откорректирована и готова к проведению исследований на ее материале.

Данная коллекция текстов представляет собой аннотированную часть корпуса, которая тщательно отбирается и вручную размечается для проведения исследований языка и стиля русской малой прозы. Создание экспертно аннотированного подкорпуса является ключевым моментом разработки ресурса, так как именно эти данные становятся потом обучающей выборкой для исследования больших объемов литературных произведений. Всего в корпусе принято 5 уровней аннотации:

* экспертная (ручная) тематическая разметка рассказов с опорой на расширенный набор тегов, впервые предложенный в [355];
* оценка эмоционального влияния текстов на читателей, для чего используется метод, описанный ранее в [356]: после прочтения текста респонденты должны оценить его с двух точек зрения: указать свое общее впечатление по десятибалльной шкале и определить, какие базовые по [357] эмоции (радость, печаль, отвращение, удивление, гнев, страх) и в какой степени вызвало у них чтение этого рассказа (по шкале от 0 до 2);
* разведение текстов на речь автора и речь персонажей с использованием инструмента для извлечения прямой речи, апробированного в [358]. Наряду с этим выделяется и собственно информация о «говорящих» героях русского рассказа: составляются их «социобиографические профили», включающие в себя полное имя, информацию является ли герой главным в рассказе или нет, его роль в произведении, профессию, семейное положение, характеристику от автора, пол, возраст, социальное происхождение и пр. [359];
* разметка эмоциональной лексики в соответствии с 11 эмоциями (страх, грусть, злоба, отвращение, стыд, счастье, наслаждение, веселье, удивление, гордость, облегчение) по методике, описанной в [360];
* выделение диегетического звука в тексте [361].

В 2025 году было протеггировано тематически 109 текстов, благодаря чему завершилась задача тематической разметки рассказов, входящих в Корпус-1000. Для тех же 109 рассказов были заполнены таблицы с оценкой силы вызываемых у читателя эмоций, однако точное количество этих таблиц уточняется. Помимо разметки литературного материала, продолжался поиск биографической информации о писателях, включенных в список 1006 персоналий, по методике, описанной в [362]. В 2025 году были заполнены таблицы с биографическими данными для 124 писателей, рассказы которых включены в Корпус-1000.

Количественные показатели прироста корпуса вынесены в [таблица 6.17](#tbl-9).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.17 – Динамика роста Корпуса русского рассказа XX века в 2024-2025 гг.   |  | Кол-во рассказов/таблиц |  | | --- | --- | --- | |  | 2024 | 2025 | | — | — | — | | Вычитка, корректура | 90 | 109 | | Тематическое теггирование | 90 | 109 | | Оценка эмоционального влияния | 90 | 109 | | Биографические данные | 17 | 124 | |

Для демонстрации принципов разметки литературного материала, использующихся в корпусе, было принято решение провести полное аннотирование Корпуса-240 с последующей публикацией результатов. Данный подкорпус уже размечен по 4 из 5 уровней, описанных выше, поэтому для обозначенной цели выбран именно он, а не Корпус-1000.

Корпус-240, включающий 240 рассказов 1901–2000 гг. (по 2-3 рассказа на каждый год) общим объемом 531725 словоупотреблений, был предназначен для апробации экспериментальных уровней аннотирования — на уровне эмоций и диегетического звука [360]. В 2025 году выполнялось разведение текстов подкорпуса на речь автора и речь персонажа. На данный момент сегментирован 161 рассказ из 240. Также были выделены говорящие персонажи, заполнены таблицы с их социобиографической информацией.

Таким образом, в 2025 году работа над Корпусом русского рассказа XX века достигла значимых результатов: была завершена вычитка и корректура аннотированной части корпуса, начата систематизация его материалов, а также доразметка текстов для демонстрации принципов аннотации.

#### 6.5.1.3 Динамика развития Корпуса устной речи студентов

Корпус устной речи молодёжи и студентов (КУРС) — проект, «направленный на создание мультимедийного языкового ресурса, предназначенного для изучения русского языка повседневного общения на материале живой неподготовленной устной речи, которую мы используем в бытовой и профессиональной коммуникации» [363; 364]. Он основан на методике непрерывной записи речи, разработанной для корпуса «Один речевой день» (ОРД), запись которого велась с 2007 по 2016 год [365–368].

Для проекта по созданию КУРС основополагающей задачей остается увеличение объёма корпуса. Добровольцы, принимающие участие в записи корпуса, проживают свой обычный день с диктофоном, фиксируя речь в естественной среде. Информантам предоставляется возможность записывать как на профессиональный диктофон (Zoom H1n или Tascam DR-05X), так и на диктофон, встроенный в собственный мобильный телефон. Чаще всего информанты выбирают запись на профессиональную технику.

За период с начала 2025 года объём аудиозаписей в корпусе увеличился на 29%, получено 562 часа исходного материала. В настоящий момент общий объём корпуса достигает 1823 часов. Количество информантов увеличилось на 38%: 57 человека, из которых 34 женщины и 23 мужчины. За 2025 год прирост информантов-мужчин максимальный, 68%, однако по-прежнему гендерная выборка остаётся недостаточно сбалансированной.

Предполагается, что один речевой день — это 12-14 часов исходного материала, то есть период с утра до вечера. Собранные за прошедший год данные насчитывают 12 речевых дней (записи длительностью от 8 часов, сделанные без пауз в рамках одного дня), 11 неполных речевых дней (к ним были отнесены как записи, сделанные в рамках одного дня, но с паузами между записями, так и записи длительностью от 5 до 8 часов).

В среднем каждый из 57 новых информантов предоставляет 10 часов аудио. Суммарная продолжительность записей, полученных от одного человека, варьируется. Так, максимальное количество полученных за 2025 год от новых информантов часов звукозаписей составляет 30 часов, а минимальная длительность равна 1 часу. За всё время сбора КУРС максимальное число часов, записанное одним информантом, составило 134 часа, а минимум равнялся 20 минутам.

Помимо аудиофайлов речи информантов и их коммуникантов, в распоряжении КУРС имеется социологическая информация об участниках, которую они добровольно предоставляют, заполняя анкету для участия в исследовании. Так, средний возраст новых информантов составляет 21 год. В большинстве случаев их родной язык — русский. Также среди новых информантов были носители молдавского и персидского, до этого не представленных среди информантов. 41% новых информантов являются студентами НИУ ВШЭ преимущественно гуманитарных специальностей. Тем не менее за 2025 год корпус пополнился записями речи представителей других профессиональных сфер: информационные технологии, история, экономика, медицина и другие.

На втором этапе работы с Корпусом устной речи молодёжи и студентов из полученных от информантов звуковых файлов извлекается важная информация — речь. Для этого файлы сегментируются на макроэпизоды, «крупные эпизоды, объединённые местом коммуникации, её условиями и участниками» [369], продолжительностью 20-40 минут, из которых удаляются неречевые фрагменты.

Каждый из эпизодов описывается в соответствии с методикой формального описания макроэпизодов, разработанной для ОРД. Она включает в себя аннотирование по типу, условию, месту коммуникации, а также указание социальных ролей информантов и коммуникантов. Работа по сегментации исходных звуковых файлов и описанию макроэпизодов выполняется вручную, после чего они передаются для автоматической расшифровки. За отчетный период были обработаны записи от 29 информантов: из 196 часов исходного материала получено 147 часов речи, 284 новых эпизода.

Следующим этапом является расшифровка аудиоматериала в текст с использованием стека технологий Whisper v3 (ASR) для автоматического распознавания речи и pyannote для диаризации (деления спикеров). После проводится ручная проверка и вычитка текстов. Использование акустической модели значительно ускоряет работу по обработке материала, тем не менее остаётся значительное число существенных ошибок, что требует ручной экспертной проверки. В [таблица 6.18](#tbl-10) приведены количественные результаты работы по увеличению объёма корпуса за 2025 год.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.18 – Прирост Корпуса устной речи молодёжи и студентов за 2025 год   |  | 2022-2024 | 2025 | прирост 2025 | итог | | --- | --- | --- | --- | --- | | количество информантов | 94 | 57 | 38% | 151 | | женщины | 83 | 34 | 29% | 117 | | мужчины | 11 | 23 | 68% | 34 | | количество исходно записанных часов | 1261 | 562 | 31% | 1823 | | количество сегментированных часов | 548 | 196 | 26% | 744 | | объём чистой речи | 380 | 147 | 28% | 527 | | количество макроэпизодов | 952 | 284 | 23% | 1236 | | количество иформантов | 44 | 29 | 40% | 73 | | количество расшифрованных часов | 84 | 46 | 35% | 130 | | количество эпизодов | 258 | 93 | 26% | 351 | |

Одним из практических результатов работы над проектом по созданию Корпуса устной речи молодёжи и студентов является публикация материалов корпуса в открытом доступе. Для этой цели был создан сайт [esc-corpus.ru](http://esc-corpus.ru/), работающий в тестовом режиме. В 2025 году был увеличен объём данных, доступных в демо-версии корпуса, в [таблица 6.19](#tbl-11) отражены ключевые аспекты этого процесса.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Таблица 6.19 – Динамика наполнения сайта КУРС   |  | 2024 | 2025 | процент от общего количества | | --- | --- | --- | --- | | количество информантов | 12 | 16 | 19% | | женщины | 11 | 9 | 17% | | мужчины | 1 | 7 | 24% | | возрастная группа: 18-24 | 11 | 15 | 19% | | возрастная группа: 25-35 | 0 | 1 | 13% | | возрастная группа: 36-54 | 1 | 0 | 33% | | возрастная группа: 55+ | 0 | 0 | 0% | | количество токенов | 55000 | 200000 | н/д | | количество часов речи | 13 | 40 | 41% | |

Кратный прирост корпуса затруднителен, так как значительное число записей представляет собой частные разговоры информантов, требующие экспертного внимания. Для предотвращения разглашения персональных данных материалы КУРС проходят обязательную процедуру анонимизации. Также в соответствии с Федеральным Законом от 05.04.2013 № 34-ФЗ (ред. от 01.07.2021) «О внесении изменений в статью 4 Закона Российской Федерации «О средствах массовой информации» и статью 13.21 Кодекса Российской Федерации об административных правонарушениях», в котором говорится о запрете публикации ненормативной лексики в печати и онлайн, в текстах расшифровок устной речи кодируется непечатная лексика (часть буквенных символов, входящих в состав четырёх корней, которые признаны нецензурными, заменены на символ «\*»).

Анонимизация осуществляется по утверждённой методике и включает замену всей личной информации. Имена и фамилии, являясь важной частью бытовой коммуникации, заменяются на вымышленные аналоги. Ключевой принцип — сохранение ритмической структуры и количества слогов оригинала (например, «Саша» → «Маша%», «Сонечка» → «Тонечка%»). Адреса, номера телефонов, места работы/учёбы, названия организаций, которые прямо или косвенно могут идентифицировать говорящего, заменяются на обобщающие категории в угловых скобках (<адрес>, <компания>, <номер>).

В целом результаты 2025 года показывают устойчивую положительную динамику развития корпуса: рост объёма аудиоматериала, расширение круга информантов, появление новых языковых и социокультурных групп, а также внедрение автоматизированных технологий обработки данных. Полученные данные представляют уникальный материал, отражающий живую русскую речь XXI века, и формирующие базу для ее междисциплинарных исследований.

# 7. Заключение

Подводя итог данному разделу, отметим, что в рамках проведенных в 2025 году типологических исследований детально описаны императивные междометия в языках Кавказа; именные части речи, сложные глаголы, грамматический род и дейксис в иранских языках; морфосинтаксис пространственной сферы в северных языках. Целостное описание получили лексические группы «мешать» на материале большого числа языков и «менять» в русском языке в типологической перспективе. Получены существенные результаты для малых иранских (памирских) языков в отношении таджикских заимствований в шугнанском, исконной лексики в объеме списка Сводеша для шугнанского и бартангского. Зафиксированы типологические особенности в сфере образования звуков, ударения и интонации в языках северного ареала, а также закономерности фонетической адаптации таджикских заимствований в шугнанском. Созданы и совершенствуются электронные ресурсы для машинного перевода на базе шугнанского и других памирских языков.

В целом сопоставительные исследования языков разных ареалов в типологической перспективе позволяют выявить общие и специфические черты родственных и неродственных идиомов, таким образом расширяя представление современной науки о языке в сфере теории универсалий и идиоэтнического лингвистического своеобразия. Особый интерес представляют оригинальные методики лексико-типологического анализа, реализованные на материале двух семантических зон. Изучение разных языковых уровней и аспектов (звукового строя и интонации, грамматики, лексики) позволяет выявить и интерпретировать такие черты языковых систем и функционирования языков, которые в перспективе могут скорректировать знание об устройстве и вариативности человеческой коммуникации в целом, об особенностях соотношения языковых черт, проявляющихся в конкретном языке совместно или порознь. Создание цифровых ресурсов на материале отдельных языков (корпусов, баз, словарей) служит общей задаче накопления лингвистических данных для их последующего сопоставления и уточнения. Существенно, что такие ресурсы вносят вклад в сохранение малых языков и дают возможность сопоставлять их друг с другом и с крупными языками, а также формируют основу для диахронического анализа. В процессе реализации проекта 2025 года существенно расширены представления о языках трех ареалов: Кавказа, Памира и севера, включая Арктику. Особенно ценны описания, построенные на материале уникальных эмпирических данных, собранных в ходе экспедиций.

Представьте, что здесь заключение для частноязыкового.

Таким образом, в результате исследований были получены уточнения формальных моделей (как в области формальной семантики, так и синтаксиса) для следующих областей: семантика и синтаксис видо-временных категорий, анафора и связывание, свойства логических операторов, дистантные передвижения в полипредикации. Для этого привлекались данные русского, хантыйского, абазинского, кильдинского саамского, мишарского татарского и др. Для исследований русского языка активно применялись экспериментальные методы, которые позволяют верифицировать модели, часто основанные на интроспекции. Кроме того, экспериментальные методы позволили выявить новые нетривиальные особенности согласования в русском языке, ранее не учитываемые в теории. Материалы других языков были получены в ходе полевых исследований методом элицитации. Эти методы получили дальнейшее развитие, в частности, применялись современные методы семантического опроса.

Для развития эмпирической базы для дальнейших исследований был разработан ряд ресурсов, некоторые из которых открыты для общего доступа (в частности, корпус текстов на хантыйском языке, типологическая база нефинитных форм в языках северной Евразии).

В результате разработан эксплицитный формальный анализ семантики наречного сдвига в казымском хантыйском, позволяющий предсказать несовместимость относительного времени с индексикальными наречиями: относительное время связывается локально доступным временем матричной клаузы, а наречия интерпретируются относительно момента речи. Также разработана эксплицитная семантическая модель причастного показателя -(m)a в кильдинском саамском, объясняющая все употребления этого показателя (показателя упорядочивающего аспекта, вводящего отношение предшествования). Исследование синтаксических свойств дистантного скрэмблинга в мишарском татарском показало, что такой скрэмблинг возможен вне зависимости от падежного маркирования субъекта, что ставит под сомнение описательную и объяснительную адекватность ряда существующих моделей. Основными результатами исследования грамматических свойств логических операторов стало уточнение эмпирической картины — необходимый этап при верификации гипотез в основе эксплицитной формальной модели соответствующих явлений. Выявлены факты, ставящие под сомнение анализ русских местоимений на -либо как выражений отрицательной полярности, и выдвинута альтернативная гипотеза, подлежащая дальнейшей разработке. В результате экспериментальных исследований уточнены ограничения на лицензирование отрицательных и реципрокальных местоимений в русском языке, а также ограничения на согласование по роду и числу, в том числе в зависимости от референциального статуса контролёра согласования. Исследование синтаксического поведения анафорических элементов в казымском хантыйском позволило описать ограничения на употребления двух классов местоимений — демонстративов и прономиналов, заложив основу для разработки модели их поведения. Исследования в рамках уточнения формальных моделей отдельных языковых явлений, а также более глубокое их рассмотрение с использованием экспериментальных методов велись в следующих направлениях: семантика и синтаксис грамматических глагольных категорий, анафора и связывание, исследование согласования экспериментальными методами, грамматические свойства логических операторов.

Исследование взаимовлияния логических и лингвистических факторов в процессах обработки информации рациональными агентами с ограниченными логико-когнитивными ресурсами позволило построить формальные модели для разрешения ряда затруднений логико-семиотического характера, возникающих в условиях языкового и мультиагентного многообразия.

В результате исследования была осуществлена реактуализация семиотического наследия Ч-С. Пирса. Методология синехизма, намеченная в его поздних трудах, применена к решению актуальных задач обоснования междисциплинарного трансфера в научных коммуникациях, включающих искусственный интеллект. Разработана семиотическая модель трансформации структурных элементов интернет-мемов, использующая концептуальный аппарат семиотики Ч.-С. Пирса (понятия иконичности, типа, знака и привычки вывода). В актуальной ситуации плюрализма логических систем предложена реконструкция классического аргумента Г. Фреге против “логических чужаков”, обуславливающего признание когнитивного многообразия рационально рассуждающих агентов психологизмом в логике.

Предложены различные модификации формальной семантики перформативов, позволившие сформулировать аргументы за и против скептицизма относительно значения. Разработаны новые подходы к разрешению парадоксов неопределенности, основанные на критике ее эпистемической интерпретации. С использованием аппарата эпистемической логики построены соответствующие формализации.

Выявлены и экспериментально подтверждены различия в дистрибуции пропозициональных глаголов и степени коммитмента в контекстах со слифтами и стандартными вложенными конструкциями. На основе дихотомии типов обновления дискурса (ассерций и импозиций) построена формальная модель этих различий. Предложена формальная модель эвиденциальных конструкций, верифицированная анализом морфем первого и второго прошедших времен коми-зырянского языка.

В данном разделе был рассмотрен язык учебных текстов и язык профессионалов, работающих на данном языке. В качестве профессионального языка было предложено использовать тексты различных научных работ и новостные тексты. Были разобраны не только логически корректные рассуждения, но и рассуждения, обладающие теми или иными недостатками в рассуждениях, например, фейковые новости или новости с неполными рассуждениями. В разделе показывается, что синтаксические связи между парами и тройками слов может служить для анализа сходства структуры предложений в разных языках. Также предлагается новый инструмент, позволяющий обучающимся автоматизировать проверку своих работ и за счёт этого сократить время отклика от преподавателя.

Таким образом, получены данные о том, что существующие алгоритмы автоматической суммаризации не всегда справляются с задачей краткого пересказа художественного текста. Расхождения во влиянии разных параметров на результат суммаризации указывают на то, что они влияют на результат не так сильно, как обучающая выборка каждой конкретной модели. В перспективе исследования разработка нового инструмента суммаризации путем тонкой настройки моделей с использованием улучшенных версий эталонных кратких содержаний в качестве обучающей выборки.

В данном разделе были описаны результаты экспериментов с большими мультимодальными моделями (VLMs), а также большими языковыми моделями (LLMs), в которых проверялись возможности моделей в решении задач, связанных с интерпретативной эвристической деятельностью: описание картин из коллекции Государственного Эрмитажа, суммаризация содержания художественных текстов. Так, было установлено, что модели дают более согласованные описания картин, чем информанты, а расхождения с человеческими текстами увеличиваются при усложнении сцены или переходе к абстрактным картинам. Модели пока показывают недостаточное качество суммаризации художественных рассказов, хотя при этом фактов влияния композиционно-эстетической сложности текстов на метрики качества не установлено.

Кроме того, проведены эксперименты по оценке качества существующих систем автоматического распознавания речи. В качестве материала для такой оценки использовались данные атипичной речи (корпус речи пациентов с афазией) и устной речи в зашумленных условиях, собранные сотрудниками подразделения и представляющие ценность для дальнейших исследований в качестве источника данных. Сформированы бенчмарки для оценки отечественных движков для автоматического распознавания нетипичной речи и для оценки качества диаризации устной спонтанной речи. На основе анализа пяти ведущих отечественных ASR-движков охарактеризован существующий уровень инклюзивности российского сегмента рынка продуктов для ASR.

Представлен эффективный пайплайн для сентимент-анализа отзывов на культурные институции. С помощью многоступенчатого промптинга одна из отечественных больших языковых моделей была дообучена для задачи аспектного сентимент-анализа неструктурированных данных на русском языке. Значения метрик оценки качества сентимент-анализа опередили соответствующие метрики словарного метода и метода анализа с помощью нейронный сетей.

Наконец, в данном разделе описан один из вариантов исследовательской логики при изучении корпуса текстов русскоязычных социальных сетей. На примере текстов зоообъявлений и гороскопов продемонстрировано, что дискурс социальных сетей, с одной стороны, формирует паттерны нормативной картины мира, а с другой, сам по себе имеет манипулятивный характер.

В данном разделе были описаны структура и содержание трех разработок: системы векторного поиска картин по Цифровой коллекции Государственного Эрмитажа, разработанной в сотрудничестве с компанией “Яндекс.Облако”, а также двух лингвистических корпусов: Корпуса русского рассказа ХХ в., а также Корпуса устной речи студентов.

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Цветкова Л. С. Афазия и восстановительное обучение. — Москва: Просвещение, 1988. — С. 207.

2. Лурия А. Р. Основы нейропсихологии. — Москва: Изд-во Моск. ун-та, 1973. — С. 374.

3. Лурия А. Р. Высшие корковые функции человека. — Санкт-Петербург: Питер, 2008. — С. 431.

4. Dingemanse M. [Interjections](https://til.im/ru/slovar) // The oxford handbook of word classes / ed. by Lier E. van. — Oxford: Oxford University Press, 2023. — P. 477–491.

5. Aikhenvald A. Y. Imperatives and commands. — Oxford: Oxford University Press, 2010.

6. Gusev V. Ju. Tipologija imperativa [typology of the imperative]. — Moscow: Jazyki slavjanskoj kulʹtury, 2013.

7. Daniel M., Filatov K., Maisak T., Moroz G., Mukhin T., Naccarato C., Verhees S. [Typological atlas of the languages of daghestan (TALD), v. 2.0.0.](https://lingconlab.ru/tald) — Moscow: Linguistic Convergence Laboratory, NRU HSE, 2025.

8. Ganieva F. A. Xinalugsko-russkij slovarʹ [Khinalug-Russian dictionary]. — Makhachkala: DNC RAN, 2002.

9. Xalilov M. Š., Isakov I. A. Ginuxsko-russkij slovarʹ [hinuq-russian dictionary]. — Makhachkala: DNC RAN Institut jazyka, literatury i iskusstva im. G. Cadasy i Institut èvoljucionnoj antropologii im. M. Planka, 2005.

10. Jusupov X. A. Darginsko-russkij slovarʹ (rukopisʹ) [dargwa-russian dictionary (manuscript)]. — Makhachkala: Institut istorii, jazyka i literatury im. G. Cadasy., 2014.

11. Bammatov B. G., Gadžiaxmedov N. È. Kumyksko-russkij slovarʹ [Kumyk-Russian dictionary]. — Makhachkala: IJaLI, 2013.

12. Talibov B. B., Gadžiev M. M. Lezginsko-russkij slovarʹ [lezgian-russian dictionary]. — Moscow: Sovetskaja Ènciklopedija, 1966.

13. Zerzele V. Plural marking on imperatives and prohibitives // ed. by Daniel M., Filatov K., Maisak T., Moroz G., Mukhin T., Naccarato C., Verhees S. — Moscow: Linguistic Convergence Laboratory, 2024.

14. Dobrushina N. Moods of Mehweb // The mehweb language. Essays on phonology, morphology and syntax / ed. by Daniel M., Dobrushina N., Ganenkov D. — Berlin: Language Science Press, 2019. — P. 117–165.

15. Dixon R. M. W. Adjective classes: A cross linguistic typology // Adjective classes: A cross linguistic typology / ed. by Dixon R. M. W., Aikenvald A. Y. — Oxford: Oxford University Press, 2004. — Vol. 1. — P. 1–49.

16. Helmbrecht J. Are there adjectives in Hocąk (Winnebago)? // Lexical categories and root classes in Amerindian languages / ed. by Lois X., Vapnarsky V. — Bern: Peter Lang International Academic Publishers, 2006. — P. 289–316.

17. Sokolov S. N. The Avestan language. — М.: Наука, 1967.

18. Карамшоев Д. К. Категория рода в памирских языках (шугнано-рушанская группа). — Душанбе: Дониш, 1986. — Т. 2.

19. Dixon R. M. W. Where Have All the Adjectives Gone? // Studies in Language. — Amsterdam/Philadephia: John Benjamins Publishing Company, 1977. — Vol. 1, no. 1. — P. 19–80.

20. Карамшоев Д. К. Шугнанско-русский словарь в трёх томах. А—З. — М.: Наука, 1988. — Т. 1.

21. Wilkins D. P. The Demonstrative Questionnaire: “THIS” and “THAT” in Comparative Perspective // Demonstratives in cross-linguistic perspective / ed. by Levinson S. C., Cutfield S., Dunn M. J., Enfield N. J., Meira S. — Cambridge: Cambridge University Press, 2018. — P. 43–71.

22. Ростовцев-Попель А. А. Типология демонстративов: Средние дейктики // Вопросы языкознания. — 2009. — Is. 2. — P. 22–34.

23. Карамшоев Д. К. Баджувский диалект шугнанского языка. — Душанбе: Академия наук Таджикской ССР, 1963.

24. Parker C. A Grammar of the Shughni language. — Montreal: McGill University, 2023.

25. Samvelian P. Specific features of Persian syntax. The Ezāfe Construction, Differential Object Marking and Complex Predicates // The Oxford handbook of Persian linguistics / ed. by Sedighi A., Shabani-Jadid P. — Oxford: Oxford University Press, 2018. — P. 226–269.

26. Modarresi F., Krifka M. Pseudo Incorporation and Anaphoricity: Evidence from Persian // Glossa: a journal of general linguistics. — Open Library of Humanities, 2021. — Vol. 36, no. 1. — P. 1–32.

27. Dabir-Moghaddam M. Compound verb in Persian // Studies in the Linguistic Science. — 1997. — Vol. 27, no. 2. — P. 25–59.

28. Driemel I. Pseudo-incorporation and its movement patterns // Glossa: a journal of general linguistics. — Open Library of Humanities, 2020. — Vol. 5, no. 1. — P. 1–40.

29. Boersma P., Weenink D. Praat: Doing phonetics by computer. Available online at: Http://www.praat.org/. — 2025.

30. Kopecka A., Vuillermet M. Source-Goal (a)symmetries across languages // Studies in Language. — 2021. — Vol. 45, no. 1. — P. 2–35.

31. Ylikoski J. The origins of the western Uralic s-cases revisited: Historiographical, functional-typological and Samoyedic perspectives // Finnisch-Ugrische Forschungen. — 2016. — Vol. 63. — P. 6–78.

32. Nikolaeva I., M. T. A grammar of Udihe. — Berlin, New York: De Gruyter Mouton, 2001.

33. Рахилина Е. В., Резникова Т. И. Фреймовый подход к лексической типологии // Вопросы языкознания. — 2013. — Вып. 2. — С. 3–31.

34. Рахилина Е. В. Когнитивный анализ предметных имен. — М.: Азбуковник, 2010. — С. 448.

35. François A. Semantic maps and the typology of colexification: Intertwining polysemous networks across languages // From Polysemy to Semantic Change: Towards a Typology of Lexical Semantic Associations / ed. by Vanhove, Martine. — Amsterdam/Philadephia: John Benjamins Publishing Company, 2008. — P. 163–215.

36. Рыжова Д. А., Добрушина Н. Р., Бонч-Осмоловская А. А., Выренкова А. С., Кюсева М. В., Орехов Б. В., Резникова Т. И. ЕВРика! Сборник статей о поисках и находках к юбилею Е.В. Рахилиной. — М.: Лабиринт, 2018. — С. 392.

37. Резникова Т. И. Глаголы прятания: Типология систем // Вопросы языкознания. — 2022. — Вып. 4. — С. 66–94.

38. Багирокова И. Г., Рыжова Д. А. Глаголы прятания и особенности их оформления локативными аффиксами в адыгейском языке // Вопросы языкознания. — 2022. — Вып. 4. — С. 95–114.

39. Рахилина Е. В., Рыжова Д. А. О принципах глагольной колексификации // Вопросы языкознания. — 2025. — Вып. 2. — С. 7–25.

40. Lehrer A. Semantic fields and lexical structure. — Amsterdam: North-Holland, 1974. — P. 225.

41. Лапшина К. М., Копылова Е. В., Матюхина Е. П. Семантическое поле ’менять’ в финском языке // Урало-алтайские исследования. — 2024. — Т. 52, № 1. — С. 22–41.

42. Казкенова А. К. Глаголы семантического поля ’менять’ в казахском языке // Acta Linguistica Petropolitana. Труды Института лингвистических исследований. — 2026. — Вып. 1. — С. in print.

43. Goldberg A., Suttle L. Construction grammar // WIREs Cognitive Science. — 2010. — Vol. 1, no. 4. — P. 468–477.

44. Janda L. A., Endresen A., Zhukova V., Mordashova D., Rakhilina E. From data to theory: An emergent semantic classification based on the large-scale Russian constructicon // Constructions and Frames. — John Benjamins Publishing Company Amsterdam/Philadelphia, 2023. — Vol. 15, no. 1. — P. 1–58.

45. Петрова З. Н., Коноплина Е. Д. Словарь русского языка XVIII века. Вып. 19 (пенат — плангерд). — СПб: Наука, 2011. — С. 239.

46. Алексеев М. Е., Атаев Б. М., Магомедов М. А., Магомедов М. И., Мадиева Г. И., Саидова П. А., Самедов Дж. С. Современный аварский язык. — Махачкала: ИЯЛИ ДНЦ РАН, Изд-во АЛЕФ, 2012.

47. Zakirova A. Number marking across different word classes in the east caucasian languages: PhD thesis. — Moscow: HSE University, 2023.

48. Кибрик А. Е. Багвалинский язык: Грамматика, тексты, словари. — Москва: ИМЛИ РАН Наследие, 2001.

49. Sumbatova N. R., Mutalov R. O. A grammar of Icari Dargwa. — Lincom Europa, 2003.

50. Babaliyeva A. Études sur la morphosyntaxe du tabasaran littéraire: PhD thesis. — École Pratique des Hautes Études, 2013.

51. Johanson L. The structure of turkic // The turkic languages / ed. by Johanson L., Csató É. Á. — Routledge, 1998. — P. 30–66.

52. Абдуллаева А. З., Э. Г. Н., С. К. К., А. К. И., Х. О. Н., М. Х. Д. Современный кумыкский язык / под ред. Н. Э. Гаджиахмедов. — Махачкала: ИЯЛИ ДНЦ РАН, 2014.

53. T. R., J. B., T. Y. [Mari. An essential grammar for international learners](https://grammar.mari-language.com). — University of Vienna, 2022.

54. М. З. П., И. Р. Л. Карельско-русский словарь. — Петрозаводск: Периодика, 1999.

55. Ослон М. В. Язык котляров-молдован. Грамматика кэлдэрарского диалекта цыганского языка в русскоязычном окружении. — Москва: Издательский Дом ЯСК, 2018.

56. Bates D. [Fitting linear mixed-effects models using lme4](https://arxiv.org/abs/1406.5823). — 2014.

57. Nakagawa S., H. S. A general and simple method for obtaining R2 from generalized linear mixed-effects models // Methods Ecol Evol. — 2013. — Is. 4. — P. 133–142.

58. Daniel M., Dobrushina N., Knyazev S. Highlanders’ russian: Case study in bilingualism and language interference in central daghestan // Slavica Helsingiensia. — 2010. — Vol. 40. — P. 65–93.

59. Shagal K. Contact-induced grammatical phenomena in the russian of erzya speakers // Mordvin languages in the field. — Helsinki: University of Helsinki, 2016. — P. 363–377.

60. Стойнова Н. М., Шлуинский А. Б. Русская речь лесных энцев: Зарисовки исследователей вымирающего языка // Slavica Helsingiensia. — 2010. — С. 153–165.

61. Бубрих Д. В. Русский язык в его отличиях от мордовского. // Проблемы мордовско-русского билингвизма. — Мордов. кн. изд-во, 1985. — С. 146–165.

62. Khomchenkova I., Pleshak P., Stoynova N. Gender disagreement in the contactinfluenced russian of northern siberia and russian far east. — 2018. — P. 14–15.

63. Panova A., Philippova T. [When a cross-linguistic tendency marries incomplete acquisition: Preposition drop in russian spoken in daghestan](https://doi.org/10.1177/1367006921990442) // International Journal of Bilingualism. — 2021. — Is. 3. — P. 640–667.

64. Ашмарин Н. И. Опыт исследования чувашского синтаксиса: Часть первая. — Типолитография Клучникова, 1903.

65. Андреев Н. А. Краткий грамматический очерк чувашского языка. // Чувашско-русский словарь. — Государственное издательство иностранных и национальных словарей, 1961.

66. Павлов И. П. Современный чувашский язык (vol. 2). — Чувашский гос. ин-т гуманитарных наук, 2017.

67. Savelyev A. [Chuvash and the bulgharic languages](https://doi.org/10.1093/oso/9780198804628.003.0028) // The oxford guide to the transeurasian languages. — 2020. — P. 446–464.

68. Резюков Н. А. Сопоставительная грамматика чувашского и русского языков. — Гос. изд. Чувашской АССР, 1959.

69. Мызников С. А. Русские говоры среднего поволжья. — Наука, 2005.

70. К. А. Б. (Иванова), A. Х. М., А. Д. Е. (Кожемякина), А. Р. Е., Е. Р. Т., А. С. А., К. Т. Н., А. З. В., А. М. В., Б. П. А., Р. Д. Н. Корпус русской речи чувашии. — Москва: Международная лаборатория языковой конвергенции, НИУ ВШЭ, 2018.

71. Gehrke B., Lekakou M. [How to miss your preposition](https://www.academia.edu/2092553/How_to_miss_your_preposition_with_Marika_Lekakou_) // Studies in Greek Linguistics. — 2013. — Vol. 33. — P. 92–106.

72. Caponigro I., Pearl L. [Silent prepositions: Evidence from free relatives.](https://chooser.crossref.org/?doi=10.1075%2Fla.120.18cap) // The syntax and semantics of spatial p. — 2008. — P. 365–385.

73. Naccarato C., Moroz G. [Non-standard numeral constructions in L2 russian: A corpusbased study](https://doi.org/10.1177/13670069251323203) // International Journal of Bilingualism. — 2025.

74. Levshina N. [How to do linguistics with r](https://doi.org/10.1075/z.195.website). — John Benjamins, 2015.

75. Рахилина Е. В., Казкенова А. Нестандартные метафоры движения в текстах казахско-русских билингвов // Russian Language Journal. — 2025. — Vol. 75, no. 1.

76. Кувшинская Ю. М., Аксенова А. А. Современные тенденции в употреблении коннектора то есть // Русская грамматика: Cистема – узус – варьирование. — Берлин, 2021. — P. 341–357.

77. Лю Ч., Титкова С. И. Можно ли начинать предложение с потому что? (К вопросу об особенностях употребления и написания союза потому что) // Вестник Чувашского государственного педагогического университета им. ИЯ Яковлева. — 2016. — Vol. 90, no. 2. — P. 36–45.

78. Кувшинская Ю. М. Предложения с союзом «потому что» как независимые высказывания в современном устном дискурсе. Доклад, представленный на VII Международном научном симпозиуме «Русская грамматика: Константы, контексты, перспективы». Тюмень, 10-14 октября 2023 г. — 2023.

79. Прияткина А. Ф. Русский синтаксис в грамматическом аспекте (синтаксические связи и конструкции). — Владивосток: Изд-во Дальневосточного ун-та, 2007.

80. Баранов А. Н., Крейдлин Г. Е. Иллокутивное вынуждение в структуре диалога // Вопросы языкознания. — 1992. — Is. 2. — P. 84–99.

81. Граудина Л. К., Ширяев Е. Н. Культура русской речи: Учебник для студентов высших учебных заведений. — М.: Норма, 2008.

82. Кувшинская Ю. М. Употребление неопределенных местоимений серии на -либо в контексте прямого отрицания // Труды ИРЯ РАН. — 2024. — Vol. 40, no. 2. — P. 150–174.

83. Кибрик А. А., Подлесская В. И. Рассказы о сновидениях: Корпусное исследование устного русского дискурса. — М.: ЯСК, 2009.

84. Апресян В. Ю., Пекелис О. Е. Материалы для проекта корпусного описания русской грамматики (http://rusgram.ru). — М., 2012.

85. Van Dijk T. A. Episodes as units of dicourse analysis // Analysing discourse: Text and talk / ed. by Tannen D. — Washington, D.C.: Georgetown University Press, 1982.

86. Ванников Ю. В. Парцелляция // Лингвистический энциклопедический словарь / ed. by Ярцева В. Н. — М., 1990. — P. 369.

87. Иванчикова Е. А. Парцелляция, ее коммуникативно-экспрессивные и синтаксические функции // Русский язык и советское общество: Морфология и синтаксис современного русского литературного языка. — М.: Наука, 1968. — P. 277–301.

88. Сковородников А. П. О функциях парцелляции в современном русском литературном языке // Русский язык в школе. — М.: Наш язык, 1980. — Vol. 5. — P. 86–91.

89. Пинегина Я. Н. Парцеллированные конструкции и их коммуникативно-прагматические функции в современных медиатекстах. Дис. ... Канд. Филол. наук.: PhD thesis. — Ростов н/Д, 2005.

90. Каркошко О. П. Парцелляция: Структура, семантика, функция: На материале русского и немецкого языков: Автореферат дис. ... Кандидата филологических наук: PhD thesis. — Майкоп, 2011.

91. Цумарев А. Э. Парцелляция в современной газетной речи. Автореф. Дис. Канд. Филолог. наук.: PhD thesis. — М., 2003.

92. Апресян В. Ю., Летучий А. Б. Конкуренция дательного падежа и предлога для при глаголах и предикативах в русском языке: Семантика и синтаксис // Russian Linguistics. — 2023. — Vol. 47, no. 2. — P. 167–207.

93. Летучий А. Б., Джонова М. Способы оформления одушевленного участника при предикативах в русском и болгарском языках // Русская речь. — 2022. — P. 26–46.

94. Сердобольская Н. В., Толдова С. Ю. Конструкции с оценочными предикативами в русском языке: Участники ситуации оценки и семантика оценочного предиката // Acta Linguistica Petropolitana. Труды института лингвистических исследований. — 2014. — Vol. 10, no. 2. — P. 443–477.

95. Кузьмина С. М. Семантика и стилистика неопределенных местоимений // Грамматические исследования: функционально-стилистический аспект. — М.: Наука, 1989. — P. 158–231.

96. Николаева Т. М. Функциональная нагрузка неопределенных местоимений в русском языке // Изв. АН СССР, сер. лит. и яз. — 1983. — Vol. 42, no. 4. — P. 342–358.

97. Падучева Е. В. Местоимения неизвестности. Материалы для проекта корпусного описания русской грамматики, http://rusgram.ru. — 2016.

98. Pekelis O. E. Russian indefinite markers to and nibud’: A microdiachtonic approach // Russian Linguistics. — 2023. — Vol. 47. — P. 253–277.

99. Rozovskaya A., Roth D. Grammar error correction in morphologically rich languages: The case of Russian // ed. by Celikyilmaz A., Reichart R., Hakkani Tur D. — Cambridge, Massachusetts: MIT Press, 2019. — P. 1–17.

100. Kosakin D., Obiedkov S., Smirnov I., Rakhilina E., Vyrenkova A., Zalivina E. Towards Error-Cause Annotation for L2 Russian // ed. by Calzolari N., Kan M.-Y., Hoste V., Lenci A., Sakti S., Xue N. — Torino, Italy: ELRA; ICCL, 2024. — P. 14240–14258.

101. Warriner A. B., Shore D. I., Schmidt L. A., Imbault C. L., Kuperman V. Sliding into happiness: A new tool for measuring affective responses to words // Canadian Journal of Experimental Psychology. — Ottawa, Canada: CPA, 2017. — Vol. 71, no. 1. — P. 71–88.

102. Bradley M. M., Lang P. J. Affective norms for English words (ANEW): Instruction manual and affective ratings. — US: University of Florida, 1999.

103. Cheek J. M., Buss A. H. Shyness and sociability // Journal of Personality and Social Psychology. — Washington, D.C.: APA, 1981. — Is. 41. — P. 330–339.

104. Jolliffe D., Farrington D. P. Development and validation of the Basic Empathy Scale // Journal of Adolescence. — USA: Wiley, 2006. — Vol. 29, no. 4. — P. 589–611.

105. Carver C. S., White T. L. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales // Journal of Personality and Social Psychology. — Washington, D.C.: APA, 1994. — Vol. 67, no. 2. — P. 319–333.

106. Кибрик А. А. Моделирование многофакторного процесса: Выбор референциального средства в русском дискурсе // Вестник Московского университета. — 1997. — Is. 4. — P. 94–105.

107. Kibrik A. A. Cognitive inferences from discourse observations: Reference and working memory / ed. by Hoek K. van, Kibrik A. A., Noordman L. — Amsterdam: Benjamins, 1999. — P. 29–52.

108. Givón T. [Topic continuity in discourse: The functional domain of switch-reference](https://doi:10.1075/tsl.2.06giv) // Switch reference and universal grammar / ed. by Haiman J., Munro P. — Amsterdam: John Benjamins Publishing Company, 1983. — Vol. 2. — P. 51–82.

109. Kibrik A. [Reference in discourse](https://doi:10.1093/acprof:oso/9780199215805.001.0001). — Oxford: Oxford University Press, 2011.

110. Bickel B. [Referential density in discourse and syntactic typology](https://doi:10.1353/lan.2003.0205) // Language. — 2003. — Vol. 79, no. 4. — P. 708–736.

111. Arkadiev P., Panova A. Independent pronouns in languages with pronominal affixes: Evidence from abaza // Состав науки: Сборник статей к юбилею веры исааковны подлесской / ed. by А. К. Н., Р. С. Н. — Москва: Буки Веди, 2024. — P. 30–58.

112. Недялков В. П., Яхонтов С. Е. Типология результативных конструкций // Типология результативных конструкций (результатив, статив, пассив, перфект) / под ред. В. П. Недялков. — Ленинград: Наука, 1983. — С. 5–41.

113. Володин А. П. Каузативные глаголы и каузативные глагольные структуры в ительменском языке // Типология каузативных конструкций. Морфологический каузатив / под ред. А. А. Холодович. — Ленинград: Наука, 1969. — С. 170–178.

114. Pakendorf B., Krivoshapkina I. V. Ėven nominal evaluatives and the marking of definiteness // Linguistic Typology. — 2014. — Vol. 18, no. 2. — P. 289–331.

115. Сопоставительно-ономасиологический словарь диалектов карельского, вепсского, саамского языков / под ред. Ю. С. Елисеев, Н. Г. Зайцева. — Петрозаводск, 2007.

116. Окнина Л. Б., Подлепич В. В., Вологдина Я. О., Зибер И. А., Машеров Е. Л., Слезкин А. А., Стрельникова Е. В., Канцерова А. О. Особенности восприятия собственного имени человеком по сравнению с восприятием имен сходного и отличного звучания: Анализ потенциалов, связанных с событиями // Физиология человека. — 2025. — Т. 51, № 2. — С. 3–13.

117. Гаспаров М. Л. Очерк истории русского стиха: Метрика, ритмика, рифма, строфика. — Москва: Наука, 1984.

118. Zwicky A. M. Well, this rock and roll has got to stop. Junior’s head is hard as a rock // Papers from the twelfth regional meeting of the chicago linguistic society / ed. by Mufwene S. S., Walker C. A., Steever S. B. — Chicago: Chicago Linguistic Society, 1976. — P. 676–697.

119. Bikina D. Syntax of non-finite relative clauses in kazym khanty. — Moscow: HSE University, 2019.

120. Bikina D., Rakhman D., Potseluev V., Starchenko A., Toldova S. [Non-finite constructions in khanty: Their unity and diversity](https://doi.org/10.1515/flin-2022-2043) // Folia Linguistica. — 2022. — Is. 0.

121. Arregui A., Kusumoto K. [Tense in temporal adjunct clauses](https://doi.org/10.3765/salt.v0i0.2814) / Semantics and linguistic theory. — 1998.

122. Kusumoto K. [On the quantification over times in natural language](https://doi.org/10.1007/s11050-005-4537-6) // Natural Language Semantics. — 2005. — Vol. 13, no. 4. — P. 317–357.

123. Stowell T. [The syntactic expression of tense](https://doi.org/10.1016/j.lingua.2005.08.003) // Lingua. — 2007. — Vol. 117, no. 2. — P. 437–463.

124. O’Leary M. C. [Locality constraints in nominal evaluation times](https://doi.org/10.3765/plsa.v6i1.5041). — 2021. — Vol. 6, no. 1. — P. 938.

125. O’Leary M. C. About time: Lexical, structural, and discourse constraints on the temporal interpretation of nominal predicates: PhD thesis. — Los Angeles: University of California, Los Angeles; University of California, Los Angeles, 2022.

126. Keshet E. R. Good intensions: Paving two roads to a theory of the de re/de dicto distinction: PhD thesis. — Cambridge, MA: Massachusetts Institute of Technology, 2008.

127. Keshet E. [Situation economy](https://doi.org/10.1007/s11050-010-9059-1) // Natural Language Semantics. — 2010. — Vol. 18, no. 4. — P. 385–434.

128. Armenante G. IPG effects on adverbials in attitude reports / Proceedings of NELS 55. — Yale University: University of Massachusetts Amherst, 2024.

129. Klein W. Time in language. — London: Routledge, 1994.

130. Bohnemeyer J. Aspect vs. Relative tense: The case reopened // Natural Language & Linguistic Theory. — 2014. — Vol. 32, no. 3. — P. 917–954.

131. Nishiyama A., Koenig J. P. What is a perfect state? // Language. — 2010. — Vol. 86, no. 3. — P. 611–646.

132. Wegner D. The properties of perfect(ive) and (eventive) passive participles: An identity approach // Glossa: A Journal of General Linguistics. — 2019. — Vol. 4, no. 1.

133. Bertrand A., Aonuki Y., Chen S., Davis H., Gambarage J., Griffin L., Huijsmans M., Matthewson L., Reisinger D., Rullmann H. Nobody’s perfect // Languages. — 2022. — Vol. 7, no. 2.

134. Chen S., Vander Klok J., Matthewson L., Rullmann H. The “experiential” as an existential past: Evidence from javanese and atayal // Natural Language & Linguistic Theory. — 2021. — Vol. 39, no. 3. — P. 709–758.

135. Cinque G. Adverbs and functional heads: A cross-linguistic perspective. — Oxford: Oxford University Press, 1999.

136. Ramchand G. C. Situations and syntactic structures: Rethinking auxiliaries and order in english. — MIT Press, 2018. — Vol. 77.

137. Мечкина Е. И. Па̄ррнэ па̄ль мушштлэз. [Воспоминания из детства]. — Мурманск, Россия, 2014.

138. Татевосов С. Г. Акциональность в лексике и грамматике. Глагол и структура события. — Москва: Языки славянской культуры, 2015.

139. Grønn A., Stechow A. von. The perfect // The wiley blackwell companion to semantics / ed. by Gutzmann D., Matthewson L., Meier C., Rullmann H., Zimmermann T. — John Wiley & Sons, 2020.

140. Coppock E., Champollion L. Invitation to formal semantics. — Boston University; New York University, 2023.

141. Grønn A., Stechow A. von. Tense // The cambridge handbook of formal semantics / ed. by Aloni M., Dekker P. — Cambridge University Press, 2016. — P. 313–341.

142. Partee B. H. Some structural analogies between tenses and pronouns in english // Journal of Philosophy. — 1973. — Vol. 70, no. 18. — P. 601–609.

143. Julien M. On negation, tense, and participles in finnic and sámi // Syntax over time / ed. by Biberauer T., Walkden G. — Oxford University Press, 2015. — P. 163–178.

144. Toosarvandani M. Languages without tense // Language and Linguistics Compass. — 2025. — Vol. 19, no. 4.

145. Matthewson L. Temporal semantics in a superficially tenseless language // Linguistics and Philosophy. — 2006. — Vol. 29, no. 6. — P. 673–713.

146. Bondarenko T., Davis C. Cross-clausal scrambling and subject case in balkar: On multiple specifiers and the locality of overt and covert movement // Syntax. — 2024.

147. Chomsky N. Minimalist inquiries: The framework // Step by step: Essays on minimalist syntax in honor of howard lasnik. — Cambridge, MA: MIT Press, 2000. — P. 89–155.

148. Boˇskovi´c Z. On the locality of left branch extraction and the structure of NP // Studia linguistica. — 2005. — Vol. 59, no. 1. — P. 1–45.

149. Erlewine M. Y. Anti-locality and optimality in kaqchikel agent focus // Natural Language & Linguistic Theory. — 2016. — Vol. 34, no. 2. — P. 429–479.

150. Richards N. What moves where when in which language? PhD thesis. — Cambridge, MA: Massachusetts Institute of Technology, 1997.

151. Подобряев А. В. Синтаксис простого предложения // Элементы татарского языка в типологическом освещении. Мишарский диалект / под ред. С. Г. Татевосов, А. Г. Пазельская, Д. Ш. Сулейманов. — Буки Веди, 2017. — С. 374–384.

152. Подобряев А. В. Финитные сентенциальные дополнения // Элементы татарского языка в типологическом освещении. Мишарский диалект / под ред. С. Г. Татевосов, А. Г. Пазельская, Д. Ш. Сулейманов. — Буки Веди, 2017. — С. 560–588.

153. Падучева Е. В. Снятая утвердительность и неверидиктальность (нa примере русских местоимений отрицательной полярности) / suspended assertion and nonveridicality: The case of russian negative polarity items // Russian Linguistics. — 2015. — Т. 39, № 2. — С. 129–162.

154. Pereltsvaig A. Monotonicity-based vs. Veridicality-based approaches to negative polarity: Evidence from russian // Formal approaches to slavic linguistics (FASL) / ed. by King T. H., Sekerina I. A. — Ann Arbor: Michigan Slavic Publishers, 2000. — P. 328–346.

155. Pereltsvaig A. Negative polarity items in russian and the “bagel problem” // Negation in slavic / ed. by Przepiorkowski A., Brown S. — Bloomington: Slavica Publishers, 2004. — P. 153–178.

156. Bondarenko T. Anatomy of an attitude: PhD thesis. — Massachusetts Institute of Technology, 2022.

157. Zwarts F. Three types of polarity // Plurality and quantification / ed. by Hamm F., Hinrichs E. — Dordrecht: Springer Netherlands, 1998. — P. 177–238.

158. Chierchia G. Logic in grammar: Polarity, free choice, and intervention. — Oxford: OUP, 2013.

159. Dayal V. Variation in english free choice items // Universals and variation. Proceedings of GLOW in asia VII / ed. by Mohanty R., Menon M. — Hyderabad: EFL University Press, 2009. — P. 237–256.

160. Yanovich I. Choice-functional series of indefinite pronouns and hamblin semantics // Proceedings of semantics and linguistic theory (SALT) 15 / ed. by Georgala E., Howell J. — Ithaca, NY: Cornell University, 2005. — P. 309–326.

161. Kholodilova M. (I)nter-clausal negative concord in russian. — Poster presented at The pragmatics of grammar: Negation and polarity. Caen, 2015.

162. Voznesenskaia A. What backward negative concord tells us about control. — Unpublished ms., Stony Brook University, 2024.

163. Zehr J., Schwarz F. [PennController for internet based experiments (IBEX)](https://doi: 10.17605/OSF.IO/MD832). — 2018.

164. Burukina I. [Mandative verbs and deontic modals in russian: Between obligatory control and overt embedded subjects](https://doi: 10.5334/gjgl.905) // Glossa: a journal of general linguistics. — 2020. — Is. 5. — P. art. 1.

165. Тестелец Я. Т. Введение в общий синтаксис. — Москва: Российский государственный гуманитарный университет, 2001.

166. Rappaport G. On anaphor binding in russian // Natural Language and Linguistic Theory. — 1986. — Vol. 4. — P. 97–120.

167. Голубева М. И. Синтаксис лексических реципроков в русском языке. — Москва: Выпускная квалификационная работа, Российский государственный гуманитарный университет; Российский государственный гуманитарный университет, 2020.

168. Голубева М. И. Синтаксические и семантические свойства русского взаимного местоимения друг друга. — Москва: Национальный исследовательский университет «Высшая школа экономики», 2022.

169. Летучий А. Б. Русский синтаксис: Структурные и неструктурные объяснения: Учеб. пособие. — Москва: Издательский дом Высшей школы экономики, 2024. — С. 416.

170. Miyauchi T. How to introduce instrumental agents: Evidence from binding in russian event nominal phrases // Advances in formal slavic linguistics 2017 / ed. by Marušič F., Mišmaš P., Žaucer R. — Berlin: Language Science Press, 2020. — P. 179–204.

171. Yadroff M., Franks S. The origin of prepositions // Current issues in formal slavic linguistics / ed. by Zybatow G., Junghanns U., Mehlhorn G., Szucsich L. — Frankfurt am Main: Peter Lang, 2001. — P. 69–79.

172. Герасимова А. А. Количественные методы исследования грамматических ограничений (на материале вариативного согласования в русском языке): PhD thesis. — Москва: Московский государственный университет, 2023.

173. Team R. C. R: A language and environment for statistical computing. — Vienna, Austria: R Foundation for Statistical Computing, 2018.

174. Плешак П. С. Некоторые сюжеты о структуре хантыйской ИГ. — 2018.

175. Masliukov V. [Possessive agreement and external possessors in kazym khanty](http://tipl.philol.msu.ru/application/files/9017/4523/9866/MYaBL2025_Marinina.pdf) / Syntax of uralic languages 5. — Гамбург, 2024.

176. Marinina V. Semantics of northern khanty demonstratives: Spatial deixis, discourse, and attention. — Москва: НИУ ВШЭ, 2024.

177. Лютикова Е. А. Согласование, признаки и структура именной группы в русском языке // Русский язык в научном освещении. — 2015. — Т. 30. — С. 44–74.

178. Панов М. В. Русский язык и советское общество: Морфология и синтаксис современного русского литературного языка. — Москва: Наука, 1968.

179. Corbett G. G. [Gender in russian: An account of gender specification and its relationship to declension](https://doi: 10.1007/BF03545848) // Russian linguistics. — 1982. — Vol. 6. — P. 197–232.

180. Corbett G. G. [The agreement hierarchy and (generalized) semantic agreement](https://doi: 10.16995/glossa.9164) // Glossa. — 2023. — Vol. 8. — P. 1–39.

181. Pesetsky D. Russian case morphology and the syntactic categories. — Cambridge, MA: MIT Press, 2013.

182. Privizentseva M. [Semantic agreement in russian: Gender, declension, and morphological ineffability](https://doi: 10.1007/s11049-023-09587-0) // Natural Language and Linguistic Theory. — 2024. — Vol. 42. — P. 767–814.

183. Steriopolo O. Mixed gender agreement in the case of russian hybrid nouns // Questions and Answers in Linguistics. — 2019. — Is. 5. — P. 91–105.

184. Garnham A., Yakovlev Y. The interaction of morphological and stereotypical gender information in russian // Frontiers on Psychology. — 2015. — Is. 6.

185. Pereltsvaig A. Small nominals // Natural Language & Linguistic Theory. — 2006. — Vol. 24, no. 2. — P. 433–500.

186. Matushansky O., Ruys E. Measure for measure // Slavic grammar from a formal perspective: The 10th anniversary FDSL conference. — 2015. — P. 317–330.

187. Rudnev P. Categorial selection and functional structure in the noun phrase: Revisiting russian small nominals // Russian Linguistics. — 2024. — Vol. 48, no. 1. — P. art. 1.

188. Толдова С. Ю., Холодилова М. А., Татевосов С. Г., Кашкин Е. В., Козлов А. А., Козлов Л. С., Кухто А. В., Привизенцева М. Ю., Стенин И. А. Элементы мокшанского языка в типологическом освещении / под ред. С. Ю. (отв. ред.) Толдова, М. А. (отв. ред.) Холодилова, С. Г. Татевосов, Е. В. Кашкин, А. А. Козлов, Л. С. Козлов, А. В. Кухто, М. Ю. Привизенцева, И. А. Стенин. — Москва: «Буки Веди», 2018. — С. xxiv + 1014.

189. Архангельский Т. А. Корпусная платформа tsakorpus и языки россии / Электронная письменность народов российской федерации – 2021 & IWCLUL 2021. Материалы междунар. Научно-практич. конф. — Сыктывкар: «Коми республиканская академия государственной службы и управления», 2021. — С. 23–24.

190. Toldova S. Yu., Sokur E. O. Building an open corpus and a morphological parser for corpus annotation for standard dargwa // Journal of Siberian Federal University. — 2024. — Vol. 17, no. 5. — P. 905–915.

191. Peirce C. S. Collected papers of charles sanders peirce. Volume VII. Science and philosophy and volume VIII. Reviews, correspondence and bibliography / ed. by Burks A. W. — Cambridge, MA: Belknap Press of Harvard University Press, 1966.

192. Haack S. Not cynicism, but synechism: Lessons from classical pragmatism // Trans. Charles S Peirce Soc. : A Quarterly Journal in American Philosophy. — 2005. — Vol. 41, no. 2. — P. 239–253.

193. Peirce C. S. Collected papers of charles sanders peirce. Volume i. Principles of philosophy and volume II. Elements of logic / ed. by Hartshorne C., Weiss P. — Cambridge, MA: Belknap Press of Harvard University Press, 1960.

194. Peirce C. S. Collected papers of charles sanders peirce. Volume v. Pragmatism and pragmaticism and volume VI. Scientific metaphysics / ed. by Hartshorne C., Weiss P. — Cambridge, MA: Belknap Press of Harvard University Press, 1960.

195. Lagasse E., Levin M. Future medicine: From molecular pathways to the collective intelligence of the body // Trends in molecular medicine. — Elsevier, 2023. — Vol. 29, no. 9. — P. 687–710.

196. Friston K. The free-energy principle: A rough guide to the brain? // Trends in cognitive sciences. — Elsevier, 2009. — Vol. 13, no. 7. — P. 293–301.

197. Friston K. The free-energy principle: A unified brain theory? // Nature reviews neuroscience. — Nature publishing group, 2010. — Vol. 11, no. 2. — P. 127–138.

198. Friston K., FitzGerald T., Rigoli F., Schwartenbeck P., Pezzulo G. Active inference and learning // Neuroscience & Biobehavioral Reviews. — Elsevier, 2016. — Vol. 68. — P. 862–879.

199. Pietarinen A.-V., Beni M. D. Active inference and abduction // Biosemiotics. — Springer, 2021. — Vol. 14, no. 2. — P. 499–517.

200. Dawkins R. The selfish gene. — New York: Oxford University Press, 1976.

201. Dawkins R. The god delusion. — Kent: Transworld Publishers, 2006.

202. Cannizzaro S. Internet memes as internet signs: A semiotic view of digital culture // Sign Systems Studies. — Tartu: Ulikooli Kirjastus, 2016. — Vol. 44, no. 4. — P. 562–586.

203. Bellucci F., Burton J. Observational advantages and occurrence referentiality / Diagrammatic representation and inference. 11th international conference, diagrams / ed. by Pietarinen A.-V., Chapman P., Bosveld-de Smet L., Giardino V., Corter J., Linker S. — Basel: Springer, 2020. — Vol. 12169. — P. 202–215.

204. Bellucci F., Pietarinen A.-V. Existential graphs as an instrument of logical analysis: Part i. alpha // The Review of Symbolic Logic. — Cambridge University Press, 2016. — Vol. 9, no. 2. — P. 209–237.

205. Bobrova A., Pietarinen A.-V. Thoughts, things and logical guidance // Peirce and husserl: Mutual insights on logic, mathematics and cognition / ed. by Shafiei M., Pietarinen A.-V. — Cham: Springer, 2019. — Vol. 46. — P. 43–58.

206. Peirce C. S. Collected papers of charles sanders peirce. Volume III. Exact logic (published papers) and volume IV. The simplest mathematics / ed. by Hartshorne C., Weiss P. — Cambridge, MA: Belknap Press of Harvard University Press, 1960.

207. Frege G. The basic laws of arithmetic. — Berkeley: University of California Press, 1964.

208. Conant J. In search of logical alien thought – descartes, kant, frege and the tractatus // Philosophical Topics. — 1991. — Vol. 20, no. 1. — P. 115–180.

209. Кант И. Логика, 1800 // Кант и. Собрание сочинений в восьми томах. Том 8. / под ред. А. В. Гулыга. — Москва: Издательство «Чоро», 1994. — С. 266–398.

210. Quine W. V. O. Carnap and logical truth // Synthese. — 1960. — Vol. 12, no. 4. — P. 350–374.

211. Achourioti T., Fugard A. J. B., Stenning K. The empirical study of norms is just what we are missing // Frontiers in Psychology. — 2014. — Vol. 5. — P. 1–15.

212. Ладов В. А. Иллюзия значения: Проблема следования правилу в аналитической философии. — Томск: Изд-во Томcкого университета, 2008.

213. Ладов В. А. Значение: Реализм vs скептицизм // Эпистемология и философия науки. — 2024. — Т. 61, № 2. — С. 43–50.

214. Борисов Е. В. Прямое решение проблемы крипке // Эпистемология и философия науки. — 2024. — Т. 61, № 2. — С. 23–32.

215. Борисов Е. В. Ответ оппонентам // Эпистемология и философия науки. — 2024. — Т. 61, № 2. — С. 89–94.

216. Austin J. L. How to do things with words. — Oxford: Clarendon Press, 1962.

217. Ginsborg H. Normativity of meaning // Proceedings of the Aristotelian Society. Supplementary Volume. — Oxford University Press, 2012. — Vol. 86. — P. 127–146.

218. Williamson T. The philosophy of philosophy. Second edition. — Hoboken; Chichester: Wiley-Blackwell, 2022.

219. Bacon A. Vagueness and thought. — Oxford: Oxford University Press, 2018.

220. Zardini E. A model of tolerance // Studia Logica. — Springer, 2008. — Vol. 90, no. 3. — P. 337–368.

221. Edgington D. Vagueness by degrees // Vagueness: A reader / ed. by Keefe R., Smith P. — Cambridge, MA: MIT Press, 1997. — P. 294–316.

222. Wright C. On being in a quandary. Relativism, vagueness, logical revisionism // Mind. — 2001. — Vol. 110, no. 1. — P. 45–98.

223. Moltmann F. Propositions, attitudinal objects, and the distinction between actions and products // Canadian Journal of Philosophy. — University of Calgary Press, 2013. — Vol. 43, no. 5/6. — P. 679–701.

224. Moltmann F. Truthmaker semantics for natural language: Attitude verbs, modals, and intensional transitive verbs // Theoretical Linguistics. — Mouton de Gruyter, 2020. — Vol. 46, no. 3-4. — P. 159–200.

225. Moltmann F. Attitudinal objects and propositions // Routledge handbook of propositions / ed. by Tillman C., Murray A. R. — New York; London: Routledge, 2023. — P. 408–423.

226. Moltmann F. Objects and attitudes. — Oxford: Oxford University Press, 2024.

227. Ross J. R. Slifting // The formal analysis of natural languages / ed. by Gross M., Halle M., Schützenberger M.-P. — The Hague; Paris: Mouton de Gruyter, 1973. — P. 133–169.

228. Asher N. Truth conditional discourse semantics for parentheticals // Journal of Semantics. — 2000. — Vol. 17, no. 1. — P. 31–50.

229. Koev T. Gradient at-issueness vs. Uncertainty about binary at-issueness // Theoretical Linguistics. — 2021. — Vol. 44. — P. 130–140.

230. Jackendoff R. S. Semantic interpretation in generative grammar. — Cambridge, Mass.: MIT Press, 1972.

231. Roberts C. Information structure: Towards an integrated formal theory of pragmatics // Papers in semantics / ed. by Yoon J. H., Kathol A. — Columbus, OH: The Ohio State University Department of Linguistics, 1996. — Vol. 49. — P. 91–136.

232. Büring D. On d-trees, beans, and b-accents // Linguistics and Philosophy. — 2023. — Vol. 26, no. 5. — P. 511–545.

233. Simons M., Tonhauser J., Beaver D., Roberts C. What projects and why / Proceedings of semantics and linguistic theory 20 / ed. by Li N., Lutz D. — Ithaca, NY: Linguistic Society of America, 2010. — Vol. 20. — P. 309–327.

234. AnderBois S., Brasoveanu A., Henderson R. Crossing the appositive/at-issue meaning boundary / Proceedings of semantics and linguistic theory 20 / ed. by Li N., Lutz D. — Ithaca, NY: Linguistic Society of America, 2010. — Vol. 20. — P. 328–346.

235. Stalnaker R. Assertion // Pragmatics / ed. by Cole P. — New York: Academic Press, 1978. — Vol. 9. — P. 315–332.

236. Hooper J. B. On assertive predicates // Syntax and semantics / ed. by Kimball J. P. — Leiden; Boston: BRILL, 1975. — Vol. 4. — P. 91–124.

237. Farkas D. F., Bruce K. B. On reacting to assertions and polar questions // Journal of Semantics. — 2010. — Vol. 27. — P. 81–118.

238. Бубрих Д. В. Грамматика литературного коми языка. — Ленинград: ЛГУ им. А. А. Жданова, 1949.

239. Серебренников Б. А. Категория времени и вида в финно-угорских языках пермской и волжской группы. — Москва: Наука, 1960.

240. Цыпанов Е. А. К вопросу о факторах, определяющих употребление i и II прошедших времен в коми языке // Linguistica uralica. — 2002. — Т. XXXVIII. — С. 165–175.

241. Цыпанов Е. А. Грамматические категории глагола в коми языке. — Сыктывкар, 2005.

242. Лейнонен М., Цыпанов Е. А. Грамматическая семантика эвиденциальности в коми языке (на материале модального перфекта) // Linguistica uralica. — 2009. — Т. XLV. — С. 23–35.

243. Суббота К. А. Глагол в ижемском диалекте коми языка: Грамматические категории и словообразование (на материале казымского говора): PhD thesis. — ГОУ ВПО «Сыктывкарский Государственный Университет», 2008.

244. Leinonen M. Evidentiality in komi zyryan // Evidentials: Turkic, iranian and neighbouring languages / ed. by Johanson L., Utas B. — Berlin, Boston: De Gruyter Mouton, 2000. — P. 419–440.

245. Siegl F. The second past in the permic languages: Form, function, and a comparative analysis from a typological perspective: PhD thesis. — University of Tartu, 2004.

246. Leinonen M., Vilkuna M. Past tenses in permic languages // Tense and Aspect in the Languages of Europe. — 2008. — P. 495–514.

247. Anderson L. B. Evidentials, paths of change, and mental maps: Typologically regular asymmetries // Advances in Discourse Processes. — 1986. — Vol. 20. — P. 273–312.

248. Aikhenvald A. Evidentiality. — Oxford: Oxford University Press, 2004.

249. Faller M. Semantics and pragmatics of evidentials in cuzco quechua: PhD thesis. — Stanford University, 2002.

250. Izvorski R. The present perfect as an epistemic modal / Proceedings of semantics and linguistic theory 7 / ed. by Lawson A. — Ithaca, NY: Linguistic Society of America, 1997. — Vol. 7. — P. 222–239.

251. Matthewson L., Davis H., Rullmann H. Evidentials as epistemic modals: Evidence from st’ ́at’imcets’ // Linguistic Variation Yearbook. — 2007. — Vol. 7. — P. 201–254.

252. Murray S. Evidentiality and the structure of speech acts: PhD thesis. — Rutgers University, 2010.

253. Garrett E. J. Evidentiality and assertion in tibetan: PhD thesis. — UCLA, 2001.

254. Sauerland U., Schenner M. Embedded evidentials in bulgarian // Proceedings of Sinn und Bedeutung. — 2007. — Vol. 11. — P. 495–509.

255. Koev T. Evidentiality, learning events, and spatiotemporal distance: The view from bulgarian // Journal of Semantics. — 2017. — Vol. 34, no. 1. — P. 1–41.

256. Grice P. Logic and conversation // Speech acts / ed. by Cole P., Morgan J. L. — New York: Academic Press, 1975. — Vol. 3. — P. 43–58.

257. Murray S. Varieties of update // Semantics & Pragmatics. — 2014. — Vol. 7, no. 2. — P. 1–53.

258. Fintel K. von, Gillies A. S. Must... Stay... strong! // Natural Language Semantics. — 2010. — Vol. 18. — P. 350–383.

259. AnderBois S., Brasoveanu A., Henderson R. At-issue proposals and appositive impositions in discourse // Journal of Semantics. — 2015. — Vol. 32, no. 1. — P. 93–138.

260. Luebbering C. R. Displaying the geography of language: The cartography of language maps // The Linguistics Journal. — 2013. — Vol. 7, no. 1. — P. 39–67.

261. [WALS online](https://doi.org/10.5281/zenodo.13950591) / ed. by Dryer M. S., Haspelmath M. — 2013.

262. Moroz G. Lingtypology: Easy mapping for linguistic typology [r package]. — 2017.

263. Drude S. Why we need better language maps, and what they could look like. // Endangered languages and the land: Mapping landscapes of multilingualism. — London: FEL & EL Publishing, 2018. — P. 33–40.

264. Macaulay R. K. Linguistic maps: Visual aid or abstract art? // Studies in linguistic geography. — Routledge, 2014. — P. 172–186.

265. Naccarato C., Verhees S. [Morning greetings](https://10.5281/zenodo.6807070) // ed. by Daniel M., Filatov K., Maisak T., Moroz G., Mukhin T., Naccarato C., Verhees S. — Moscow: Linguistic Convergence Laboratory, 2021.

266. Meulemans W., Sondag M., Speckmann B. A simple pipeline for coherent grid maps // IEEE Transactions on Visualization and Computer Graphics. — 2020. — Vol. 27, no. 2. — P. 1236–1246.

267. Loecher M., Ropkins K. RgoogleMaps and loa: Unleashing r graphics power on map tiles // Journal of Statistical Software. — 2015. — Vol. 63. — P. 1–18.

268. McNeill G., Hale S. A. Generating tile maps // Computer graphics forum. — 2017. — Vol. 36, no. 3. — P. 435–445.

269. Peterson M. P. The tile-based mapping transition in cartography // Maps for the future: Children, education and internet. — Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. — P. 151–163.

270. Panse C. Rectangular statistical cartograms in r: The recmap package // Journal of Statistical Software. — 2018. — Vol. 86. — P. 1–27.

271. Moroz G. On phonology of east caucasian languages // ed. by Daniel M., Filatov K., Maisak T., Moroz G., Mukhin T., Naccarato C., Verhees S. — Moscow: Linguistic Convergence Laboratory, 2021.

272. Granger S., Paquot M. Customising a general EAP dictionary to meet learner needs // ed. by Granger S., Paquot M. — Louvain: Presses universitaires de Louvain, 2010. — P. 87–96.

273. Granger S., Paquot M. Electronic lexicography goes local: Design and structures of a needs-driven online academic writing aid // Lexicographica. — 2015. — Is. 31(1). — P. 118–141.

274. Paquot M. The LEAD dictionary-cum-writing aid: An integrated dictionary and corpus tool // ed. by Granger S., Paquot M. — Oxford: Oxford University Press, 2012. — P. 163–186.

275. Biber D., Johansson S., Leech G., Conrad S., Finegan E., Quirk R. Longman grammar of spoken and written english. — Longman, 1999.

276. Celce-Murcia M., Larsen-Freeman D. The grammar book. — Thomson Learning Ltd., 2008.

277. Tsang A. Judgement of countability and plural marking in english by native and non-native english speakers // Language Awareness. — 2017. — Is. 26(4). — P. 343–359.

278. Kodera M. An overview of the count/noncount distinction of english nouns // 阪南論集. 人文・自然科学編. — 2020. — Is. 55(2). — P. 23–51.

279. Fernández-Pena Y. Exploring the patterns of verbal agreement of collective noun-headed subjects in a parsed corpus of present-day english // ed. by Vargas-Sierra C. — 2017. — Is. 2. — P. 325–334.

280. Fernández-Pena Y. Modelling verb number agreement variation with complex collective subjects in inner-circle varieties of english // ed. by Krug M., Werner V., Schützler O., Vetter F. — Peter Lang, 2022. — P. 125–153.

281. Gao X. A cross-disciplinary corpus-based study on english and chinese native speakers’ use of linking adverbials in academic writing // Journal of English for Academic Purposes. — 2016. — Is. 24. — P. 14–28.

282. Gong Z., Liu Y., Liu Y. A comparative study of research questions written by L1 english authors and chinese EFL scholars // Journal of English for Academic Purposes. — 2024. — Is. 69. — P. 101383.

283. Cohen J. Statistical power analysis for the behavioral sciences (2nd ed.). — Routledge, 1988.

284. Morallo A. B. Corpus-based investigation of s-v concord patterns of nouns with latin plural endings // Malaysian Journal of ELT Research. — 2022. — Is. 19(2). — P. 97–116.

285. Al–Shaer I. The collapse of grammatical rules under the pressure of semantic content: Subject-verb concord // An-Najah University Journal for Research - B (Humanities). — 2005. — Is. 19(1). — P. 295–314.

286. Levin M. Agreement with collective nouns in english. — Lund: Lund University, 2001.

287. Martín P., León Pérez I. K. Convincing peers of the value of one’s research: A genre analysis of rhetorical promotion in academic texts // English for Specific Purposes. — 2014. — Is. 34. — P. 1–13.

288. Afros E., Schryer C. F. Promotional (meta)discourse in research articles in language and literary studies // English for Specific Purposes. — 2009. — Is. 28(1). — P. 58–68.

289. Wang W., Yang C. Claiming centrality as promotion in applied linguistics research article introductions // Journal of English for Academic Purposes. — 2015. — Is. 20. — P. 162–175.

290. Cao X., Lei L., Wen J. Promoting science with linguistic devices: A large-scale study of positive and negative words in academic writing // Learned Publishing. — 2021. — Is. 34(2). — P. 82–88.

291. Hyland K., Jiang F. “Our striking results demonstrate …”: Persuasion and the growth of academic hype // Journal of Pragmatics. — 2021. — Is. 182. — P. 189–202.

292. Xie S., Mi C. Promotion and caution in research article abstracts: The use of positive, negative and hedge words across disciplines and rankings // Learned Publishing. — 2023. — Is. 36(2). — P. 249–265.

293. Pho P. D. Research article abstracts in applied linguistics and educational technology: A study of linguistic realizations of rhetorical structure and authorial stance // Discourse Studies. — 2008. — Is. 10(2). — P. 231–250.

294. Hyland K. Metadiscourse: Exploring interaction in writing // Continuum. — 2008.

295. Bhatia V. K. Analysing genre: Language use in professional settings. — Longman, 1993.

296. Hyland K. Stance and engagement: A model of interaction in academic discourse // Discourse Studies. — 2005. — Is. 7(2). — P. 173–192.

297. Grieve J., Woodfield H. The language of fake news. Elements in forensic linguistics. — Cambridge: Cambridge University Press, 2023.

298. Piskorski J., Stefanovitch N., Nikolaidis N., Da San Martino G., Nakov P. Multilingual multifaceted understanding of online news in terms of genre, framing, and persuasion techniques. — 2023. — P. 3001–3022.

299. Da San Martino G., Yu A., Barrón-Cedeño A., Petrov R., Nakov P. Fine-grained analysis of propaganda in news articles // ed. by Inui K., Jiang J., Ng V., Wan X. — Hong Kong, China: Association for Computational Linguistics., 2019. — P. 5636–5646.

300. Jowett G., O’Donnell V. Propaganda and persuasion. — Thousand Oaks: Sage, 2006.

301. Dayter D., Rüdiger S. Corpus linguistics and persuasion // ed. by Nesi H., Milin P. — Amsterdam: Elsevier, 2025a.

302. Dayter D., Rüdiger S. Persuasion and influence in linguistic research // ed. by Rüdiger S., Dayter D. — Cambridge: Cambridge University Press, 2025b. — P. 3–18.

303. Lakoff R. Persuasive discourse and ordinary conversation, with examples from advertising / ed. by Tannen D. — Washington: Georgetown University Press, 1982. — P. 25–42.

304. Fairclough N. Critical discourse analysis: The critical study of language. 2nd ed. — London: Routledge, 2010.

305. Austin J. L. How to do things with words: The william james lectures delivered at harvard university. — Cambridge, M.A.: Harvard University Press, 1962.

306. Koteyko I. The language of press advertising in the UK: A multi-dimensional study // Journal of English Linguistics. — 2015. — Is. 43(4). — P. 259–283.

307. Kuiken J., Schuth A., Spitters M., Marx M. Effective headlines of newspaper articles in a digital environment // Digital Journalism. — 2017. — Is. 5(10). — P. 1300–1314.

308. Garassino D., Masia V., Brocca N., A. D. B. Politicians vs ChatGPT: A study of presuppositions in french and italian political communication // AI-Linguistica. — 2024. — Is. 1(1).

309. Giora R., Givoni S., O. F. Defaultness reigns: The case of sarcasm // Metaphor & Symbol. — 2015. — Is. 30(4). — P. 290–313.

310. Charteris-Black J. Politicians and rhetoric: The persuasive power of metaphor. — London: Palgrave Macmillan, 2011.

311. Claridge C. Lying, metaphor, and hyperbole // ed. by Meibauer J. — Oxford: Oxford University Press, 2018. — P. 370–381.

312. Cruse A. Meaning in language: An introduction to semantics and pragmatics. — Oxford: Oxford University Press, 2004.

313. Benali A. The impact of using automated writing feedback in ESL/EFL classroom contexts // English Language Teaching. — 2021. — Is. 14(12). — P. 189.

314. Wang E. L., Matsumura L. C., Correnti R., Litman D., Zhang H., Howe E., Magooda A., Quintana R. eRevis(ing): Students’ revision of text evidence use in an automated writing evaluation system // Assessing Writing. — 2020. — Is. 44. — P. 100449.

315. Woodworth J., Barkaoui K. Perspectives on using automated writing evaluation systems to provide written corrective feedback in the ESL classroom // TESL Canada Journal. — 2020. — Is. 37(2). — P. 234–247.

316. Yuan Z., Taslimipoor S., Davis C., Bryant C. Multi-class grammatical error detection for correction: A tale of two systems. — 2021. — P. 8722–8736.

317. Clark K., Luong M.-T., Le Q. V., Manning C. D. ELECTRA: Pre-training text encoders as discriminators rather than generators. Version number: 1. — 2020.

318. Raffel C., Shazeer N., Roberts A., Lee K., Narang S., Matena M., Zhou Y., Li W., Liu P. J. Exploring the limits of transfer learning with a unified text-to-text transformer. — 2023.

319. Rothe S., Mallinson J., Malmi E., Krause S., Severyn A. A simple recipe for multilingual grammatical error correction // ed. by Zong C., Xia F., Li W., Navigli R. — 2021. — P. 702–707.

320. Vinogradova O., Lyashevskaya O. Review of practices of collecting and annotating texts in the learner corpus REALEC // Lecture Notes in Computer Science: Text, Speech, and Dialogue / ed. by Sojka P., Horák A., Kopeček I., Pala K. — Springer International Publishing, 2022. — Is. 13502. — P. 77–88.

321. Touvron H., Lavril T., Izacard G., Martinet X., Lachaux M.-A., Lacroix T., Rozière B., Goyal N., Hambro E., Azhar F., Rodriguez A., Joulin A., Grave E., Lample G. Llama: Open and efficient foundation language models. — 2023.

322. Vinogradova O., Login N. The design of tests with multiple choice questions automatically generated from essays in a learner corpus // SSRN Electronic Journal. — 2017.

323. K. G., K. S., C. X. Typometrics: From implicational to quantitative universals in word order typology // Glossa: a journal of general linguistics. — 2021. — Is. 6(1). — P. 1–31.

324. Kolmogorova A. V., Nalobina P. A. Conceptualizing the space: How natural and artificial cognitive agents use topological semantics schemes (based on descriptions of paintings from the hermitage collection) // Epistemology and Philosophy of Science. — 2025. — Vol. 62, no. 1. — P. 170–197.

325. Lin Z., Chen X., Pathak D., Zhang P., Ramanan D. [Revisiting the role of language priors in vision-language models](https://arXiv:2306.01879). — 2024.

326. Subbiah M., Zhang S., Chilton L. B., McKeown K. Reading subtext: Evaluating large language models on short story summarization with writers // Transactions of the Association for Computational Linguistics. — Cambridge, MA: MIT Press, 2024. — Vol. 12. — P. 1290–1310.

327. Шмид В. Нарратология. — Москва: Яз. славян. культуры : Кошелев, 2003. — С. 311.

328. Lin C.-Y. [ROUGE: A package for automatic evaluation of summaries](https://aclanthology.org/W04-1013/) / Text summarization branches out. — Barcelona, Spain: Association for Computational Linguistics, 2004. — P. 74–81.

329. Papineni K., Roukos S., Ward T., Zhu W. J. [Bleu: A method for automatic evaluation of machine translation](https://aclanthology.org/P02-1040/) / Proceedings of the 40th annual meeting of the association for computational linguistics / ed. by P. I., E. C., D. L. — Philadelphia, Pennsylvania, USA: Association for Computational Linguistics, 2002. — P. 311–318.

330. Zhang T., Kishore V., Wu F., Weinberger K. Q., Artzi Y. [BERTScore: Evaluating text generation with BERT](https://api.semanticscholar.org/CorpusID:127986044) // ArXiv. — 2019. — Vol. abs/1904.09675.

331. Богданова-Бегларян Н. В., Шерстинова Т. Ю., Баева Е. М., Блинова О. В., Мартыненко Г. Я., Ермолова О. Б., Рыко А. И. Русский язык повседневного общения: Особенности функционирования в разных социальных группах. — Санкт-Петербург: ЛАЙКА, 2016. — С. 244.

332. Карпов А. А., Кипяткова И. С. Методология оценивания работы систем автоматического распознавания речи // Известия высших учебных заведений. Приборостроение. — 2012. — Т. 55, № 11. — С. 38–43.

333. Sherstinova T., Mikhaylovskiy N., Kolpashchikova E., Kruglikova V. Bridging gaps in russian language processing: AI and everyday conversations / 35th conference of open innovations association (FRUCT). — 2024. — P. 253–258.

334. Bredin H., Yin R., Coria J. M., Gelly G., Korshunov P., Lavechin M., Fustes D., Titeux H., Bouaziz W., Gill M. P. [Pyannote.audio: Neural building blocks for speaker diarization](https://arxiv.org/abs/1911.01255).

335. [Yandex SpeechKit](https://cloud.yandex.ru/services/speechkit).

336. [SaluteSpeech](https://developers.sber.ru/portal/products/smartspeech).

337. [Schöpot.ai](https://shopot.ai/).

338. [VoiceKit](https://www.tbank.ru/software/voicekit/).

339. MacWhinney B. [Analyses of AphasiaBank data](https://doi:10.3389/conf.fpsyg.2015.65.00055). — 2015.

340. Khudyakova M., Bergelson M., Akinina Y., Iskra E., Toldova S., Dragoy O. Russian CliPS: A corpus of narratives by brain-damaged individuals / RaPID-2016. — Saarbruecken, 2016. — P. 22–26.

341. Берджер Д. Зачем смотреть на животных? — Москва: Ад Маргинем Пресс, 2017. — С. 160.

342. Крылова К. Рынок удобных животных. — Москва: НЛО, 2023. — С. 408.

343. Markowitz D. M. Putting your best pet forward: Language patterns of persuasion in online pet advertisements // Journal of Applied Social Psychology. — 2020. — Vol. 50, no. 3. — P. 160–173.

344. Азначеева Е. Н. Астрологический дискурс: Семиотический и когнитивный аспекты // Вестник Челябинского государственного университета. — 2011. — Вып. 33. — С. 19–21.

345. Вепрева И. Т. Базовый лексикон астрологического прогноза: О словах с """"размытым"""" содержанием // Экология языка и коммуникативная практика. — 2017. — Вып. 1. — С. 57–63.

346. Лалетина А. О. Гороскоп как гендерно-маркированный медиажанр // Жанры и типы текста в научном и медийном дискурсе. — 2007. — С. 288–294.

347. Bird S., Klein E., Loper E. [Natural language processing with python](https://www.nltk.org/book). — O’Reilly Media Inc., 2009.

348. Лукашевич Н. В., Левчик А. В. Создание лексикона оценочных слов русского языка РуСентилекс // Труды конференции OSTIS-2016. — 2016. — С. 377–382.

349. Hu M., Liu B. Mining opinion features in customer reviews / Proceedings of AAAI conference on artificial intelligence. — 2004. — Vol. 4. — P. 755–760.

350. Колмогорова А. В., Колмогорова П. А., Куликова Е. Р. [О прошлом, но в разное время: Компьютерный анализ текстов учебников по истории СССР / россии для шести поколений студентов](https://DOI:10.17223/19986645/89/4) // Вестник Томского государственного университета. Филология. — 2024. — Вып. 89. — С. 73–103.

351. Ling Y. S. C. The use of modal auxiliary verbs in horoscope: A corpus-based study // Issues in Language Studies. — 2016. — Vol. 5, no. 2.

352. Мартыненко Г. Я., Шерстинова Т. Ю., Мельник А. Г., Попова Т. И. Методологические проблемы создания компьютерной антологии русского рассказа как языкового ресурса для исследования языка и стиля русской художественной прозы в эпоху революционных перемен (первой трети XX века). — 2018. — Вып. 2. — С. 97–102.

353. Мартыненко Г. Я., Шерстинова Т. Ю., Попова Т. И., Мельник А. Г., Замирайлова Е. В. О принципах создания корпуса русского рассказа первой трети XX века // Труды XV международной конференции по компьютерной и когнитивной лингвистике «TEL 2018». — 2018. — С. 180–197.

354. Шерстинова Т. Ю., Кирина М. А., Хлусова Я. К. Корпус русского рассказа XX века: Текущее состояние и перспективы развития. — 2025.

355. Skrebtsova T. Thematic tagging of literary fiction: The case of early 20th century russian short stories / CEUR workshop proceedings: Proceedings of the international conference """"internet and modern society"""" (IMS-2020). — 2020. — Vol. 2813. — P. 265–276.

356. Шерстинова Т. Ю., Колпащикова Е. О., Сейнова А. Р., Максименко П. И., Родионов Р. А. Русский рассказ 1900-1930-х и его восприятие читателем: Опыт квантитативного анализа оценки художественного текста. — 2023. — Вып. 2 (54). — С. 164–184.

357. Ekman P. Facial expressions. — Chichester: Wiley, 1999. — P. 301–320.

358. Кирина М. А., Лукьянчикова А. С. Голос эпохи: Лингвостатистические показатели прямой речи в русском рассказе XX века // Восьмая калининградская школа по гуманитарной информатике. Сборник докладов. Калининград, 12–14 декабря 2024 года. — 2024. — С. 69–73.

359. Шерстинова Т. Ю., Кирина М. А., Хлусова Я. К. Корпус русского рассказа как база для проведения социолингвистических исследований русской литературы // Информационные технологии в гуманитарных исследованиях: Материалы международной научно-практической конференции, красноярск, 25–28 сентября 2023 г. — 2023. — С. 200–211.

360. Moskvina A., Kirina M. Fear and loathing in russian literature: A case of emotion annotation of short stories of the 20th century / International conference on internet and modern society. — 2025. — P. 113–129.

361. Делазари И. Комиксы на слух: Диегетический звук как трансмедиальная категория. — 2023. — Т. 179, № 1. — С. 118–132.

362. Шерстинова Т. Ю. Биографическая база данных русских писателей (к созданию корпуса русского рассказа XX века) // Корпусная лингвистика-2019. — 2019. — С. 439–447.

363. Шерстинова Т. Ю., Петрова И. А. Моделирование повседневного речевого поведения: Корпус устной речи молодежи, или ОРД v. 2.0. — 2023. — Вып. 11.

364. Sherstinova T., Petrova I. ESC corpus of spoken russian: Everyday student conversations captured through continuous speech recording in natural communicative environments / 26th international conference on speech and computer SPECOM-2024, 25-28 november 2024, belgrade, serbia. — 2024.

365. Asinovsky A., Bogdanova N., Rusakova M., Ryko A., Stepanova S., Sherstinova T. The ORD speech corpus of russian everyday communication """"one speaker’s day"""": Creation principles and annotation / Text, speech and dialogue: 12th international conference (TSD 2009). — 2009. — P. 250–257.

366. Sherstinova T. The structure of the ORD speech corpus of russian everyday communication / International conference on text, speech and dialogue. — 2009. — P. 258–265.

367. Bogdanova-Beglarian N., Sherstinova T., Blinova O., Ermolova O., Baeva E., Martynenko G., Ryko A. Sociolinguistic extension of the ORD corpus of russian everyday speech / Speech and computer: 18th international conference (SPECOM 2016), 23 - 27 august 2016, budapest, hungary. — 2016. — P. 659–666.

368. Богданова-Бегларян Н. В., Блинова О. В., Мартыненко Г. Я., Шерстинова Т. Ю. Корпус русского языка повседневного общения «один речевой день» (ОРД): Текущее состояние и перспективы. — 2019. — Т. 21. — С. 100–110.

369. Шерстинова Т. Ю. Коммуникативные макроэпизоды в корпусе повседневной русской речи """"один речевой день"""": Принципы аннотирования и результаты статистической обработки // Корпусная лингвистика-2013: Тр. Междунар. конф. — 2013. — С. 449–456.

1. https://ruscorpora.ru/. [↑](#footnote-ref-90)
2. Предыдущая версия анализатора хранится на гитхабе по ссылке: <https://github.com/tbkazakova/even_corpus/tree/main/apertium-eve>. [↑](#footnote-ref-183)
3. Являясь партнером Лаборатории языковой конвергенции, Яндекс предоставил сотрудникам подразделения бесплатный доступ к модели. [↑](#footnote-ref-315)
4. <https://huggingface.co/BoloniniD/YandexGPT-5-Lite-8B-instruct-Q8_0-GGUF> [↑](#footnote-ref-336)
5. Здесь и далее частотность в корпусе указывается в IPM (item per million) с округлением до целых. [↑](#footnote-ref-352)
6. <https://huggingface.co/seara/rubert-tiny2-russian-emotion-detection-ru-go-emotions> [↑](#footnote-ref-361)
7. Указывается процент постов, которые были классифицированы как соответствующая эмоция. [↑](#footnote-ref-363)
8. Указывается процент постов, которые были классифицированы как соответствующая эмоция. [↑](#footnote-ref-364)
9. MyStem 3.1 = морфологический парсер [Электронный ресурс] // Yandex. URL: https://yandex.ru/dev/mystem (дата обращения: 15.06.2025). [↑](#footnote-ref-366)
10. Национальный корпус русского языка [Электронный ресурс]. URL: https://ruscorpora.ru/stats (дата обращения 01.07.2025). [↑](#footnote-ref-371)
11. Orange3 Text Mining = software [Electronic resource] // Laboratory of Bioinformatics, Faculty of Computer Science, University of Ljubljana. URL: https://orangedatamining.com (date of access: 01.07.2025). [↑](#footnote-ref-372)