Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Государственный институт русского языка им. А.С. Пушкина»

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Наименование дисциплины (модуля):

Квантитативная лингвистика

Б1.В.02.04

Уровень высшего образования:
магистратура

Направление подготовки:

45.04.01 Филология

Направленность (профиль) образования:

Лингвистическая аналитика и экспертиза текстов

Форма обучения:

очная

Рабочая программа дисциплины (модуля) разработана в соответствии с федеральным государственным образовательным стандартом высшего образования по направлению подготовки 45.04.01 Филология, утвержденным приказом Минобрнауки России № 980 от 12.08.2020.

1. Место дисциплины (модуля) в структуре ОПОП 2

В соответствии с учебным планом магистратуры 45.04.01 "Филология" по специальности «Лингвистическая аналитика и экспертиза текстов» дисциплина Б1.О.02 «Квантитативная лингвистика» относится к части, формируемой участниками образовательных отношений, и изучается во 2 семестре.

2. Цель и задачи освоения дисциплины (модуля)

Курс «Квантитативная лингвистика» является вариативной частью профессионального цикла и направлен на комплексную интеграцию классической филологической науки с современными цифровыми технологиями. Главной целью курса является ознакомление студентов с фундаментальными положениями современной квантитативной лингвистики, ее законами и закономерностями, а также возможностями использования количественных методов и новых информационных технологий в профессиональной научно-исследовательской деятельности

○

3. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников

Компетенции выпускников (коды)	Индикаторы достижения компетенций	Планируемые результаты обучения по дисциплине (модулю), соотнесенные с индикаторами достижения компетенций
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.3 Критически оценивает надежность источников информации, работает с противоречивой информацией из разных источников	Знать: системную структуру современного информационного пространства в гуманитарных науках: классификацию баз данных, корпусов текстов, цифровых архивов и электронных библиотек, а также их взаимосвязи. Уметь: самостоятельно проектировать базы данных для лингвистических исследований, проводить корпусный анализ, критически оценивать достоверность сетевых источников, а также применять методы автоматической обработки текста для решения конкретных научных и

УК-6 Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1 Находит, обобщает и творчески использует имеющийся опыт в соответствии с задачами саморазвития	прикладных задач. Владеть: техникой критического аудита информационных ресурсов Знать: основные этапы и методологию проведения научного исследования в лингвистике: от постановки проблемы до формулирования выводов; критерии оценки качества и репрезентативности лингвистических корпусов, статистических моделей и программных инструментов Уметь: находить, критически оценивать и отбирать релевантные количественные данные (корпусы, датасеты) и программные инструменты для решения конкретной задачи саморазвития Владеть: культурой академической саморефлексии: способностью письменно или в ходе дискуссии аргументировать выбор методов, оценивать ограничения своего исследования и определять дальнейшие шаги для

	УК-6.2 Планирует профессиональную траекторию с учетом профессиональных особенностей, а также других видов деятельности и требований рынка труда УК-6.3 Оценивает свои ресурсы и их пределы (личностные, ситуативные, временные), целесообразно их использует	профессионального роста. Знать: базовые метатеоретические концепции саморазвития исследователя (рефлексия, метапознание, перенос знаний), применимые к работе с данными Уметь: применять межпредметный перенос: адаптировать опыт из смежных дисциплин (статистика, программирование, Data Science) для решения новых лингвистических задач (например, использовать методы анализа временных рядов для диахронических корпусных исследований). Владеть: навыком построения индивидуальной исследовательской траектории в области квантитативной лингвистики на основе рефлексии собственного практического опыта.

		Знать: Специфику предметной области «Квантитативная лингвистика»: основные направления (корпусная лингвистика, стилометрия, социолингвистическое моделирование), типичные задачи и ограничения количественных методов Уметь: обобщать результаты собственных практических работ (лабораторных, мини-исследований), выявляя в них паттерны, типичные ошибки и точки роста; формулировать персональные исследовательские вопросы и гипотезы, выходящие за рамки типовых заданий курса. Владеть: техникой переноса освоенных методов (например, логистической регрессии для анализа социолингвистических переменных или векторных моделей для семантики) в новые предметные контексты и на новые типы данных.
ПК-1 Способен проводить	ПК-1.1. Выделяет и систематизирует значимые аспекты описания лексических единиц в соответствии с целями исследования и типом материала	Знать: принципы стратегического планирования информационного поиска и исследования: от

разноаспектное описание лексических единиц для решения теоретических и прикладных задач	ПК-1.2 Применяет цифровые и корпусные технологии для работы с лексическими данными	постановки научной проблемы до выбора оптимального комплекса инструментов и прогнозирования результатов. Уметь: осуществлять декомпозицию филологической проблемной ситуации на информационные составляющие (выделить корпус источников, историографию вопроса, фактологическую базу, недостающие данные). Владеть: навыком системного информационного мониторинга для комплексного решения филологической задачи. Знать: методологические ограничения цифровых инструментов (статистическую погрешность при работе с малыми корпусами, алгоритмические искажения поисковых систем) и способы их критической оценки. Уметь: выстраивать иерархию информационных ресурсов по степени их значимости для

	ПК-1.3 Представляет результаты разноаспектного описания лексических единиц в формате, пригодном для разных пользователей	конкретной исследовательской стратегии Владеть: навыком интеграции результатов системного анализа в итоговый научный продукт с наглядной демонстрацией логики принятых решений. Знать: типологию проблемных ситуаций в филологической деятельности, требующих системного анализа. Уметь: формулировать многоуровневую стратегию действий: определять приоритетные направления поиска, распределять ресурсы, выбирать методы обработки данных (количественные, качественные, текстологические) и предусматривать альтернативные сценарии при отсутствии данных. Владеть: практикой применения специализированного программного обеспечения для анализа данных в рамках единой стратегии

		исследования.
ПК-2 Способен осуществлять деятельность по созданию и редактированию текстового контента в соответствии с задачами профессиональной деятельности	ПК-2.1 Редактирует, трансформирует, реферирует различные типы текстов в зависимости от задач профессиональной деятельности	Знать: основные понятия и метрики квантитативной лингвистики: частотность (абсолютная и относительная), ранг, закон Ципфа, коэффициент Хердана, индекс Юла-К, TTR (Type-Token Ratio), MATTR (Moving-Average Type-Token Ratio), MSTTR (Mean Segmental Type-Token Ratio); принципы статистического анализа текстовых корпусов, включая оценку статистической значимости (p-value, хи-критерий Пирсона, точный тест Фишера). Уметь: проводить предредактирование текстов для их автоматизированной обработки: нормализация, очистка от шума, унификация метаданных; выполнять автоматическую и ручную разметку текстов, сегментировать их на функциональные блоки (введение, методы, результаты) для последующего количественного

	ПК-2.2 Способен осуществлять деятельность по созданию и редактированию текстового контента в соответствии с задачами профессиональной деятельности	анализа Владеть: методикой комплексного квантитативного анализа текста, включающего предобработку, сбор статистики, проверку статистических гипотез и лингвистическую интерпретацию результатов Навыками настройки параметров корпусных менеджеров и скриптов на Python/R для решения конкретных задач редактирования и трансформации текста. Знать: методы лемматизации, стемминга, POS-теггинга и их влияние на количественные показатели текста; типологию лингвистических корпусов (эталонные, динамические, параллельные, диахронические) и их репрезентативность; основные алгоритмы автоматического извлечения ключевых слов, коллокаций и построения конкордансов; специфику

		квантитативных параметров различных функциональных стилей и жанров (научный, официально-деловой, медийный) Уметь: применять специализированное ПО (AntConc, Sketch Engine, R с пакетами tidytext/quanteda, Python с библиотеками NLTK/spaCy/NLTK) для извлечения количественных данных из текстов; рассчитывать и интерпретировать лексическое богатство, синтаксическую сложность и плотность информации текста с помощью количественных индексов Владеть: навыками сопоставительного квантитативного анализа нескольких текстов или корпусов для выявления их жанровых, авторских или хронологических различий

4. Объем дисциплины (модуля) _3_з.е., в том числе __36__академических часа на контактную работу обучающихся с преподавателем; ___74__академических часа на самостоятельную работу обучающихся.

(В конце данного пункта следует отметить, если дисциплина или часть ее реализуется с применением электронного обучения и дистанционных образовательных технологий)

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов

Наименование и краткое содержание разделов и тем дисциплины (модуля). Форма промежуточной аттестации по дисциплине (модулю)	Всего (часы)	В том числе							
		Контактная работа (работа во взаимодействии с преподавателем) *Виды контактной работы, часы*					Самостоятельная работа обучающегося *Виды самостоятельной работы, часы*		
		Занятия лекционного типа	Занятия семинарского типа (практические/ лабораторные)	Групповые консультации	Индивидуальная работа с обучающимися	Всего	Задания для самостоятельной работы	Доклад	Всего
Тема 1. Математическое моделирование языка: от комбинаторики к структурно-вероятностным моделям. Эволюция	12	2	2			4		8	8

подходов, энтропия и избыточность в естественных языках. Сравнение классических вероятностных моделей с современными нейросетевыми архитектурами									
Тема 2. Корпусная лингвистика и проблема репрезентативности данных. Принципы формирования сбалансированных корпусов, оценка объема выборки, статистическая значимость, проблема Zipfian sparsity (разреженности данных) и методы ее компенсации	12	2	2			4		8	8
Тема 3. Количественные методы в лексикологии и лексикографии. Частотный анализ	14	2	2			4		10	10

лексем, статистические критерии выделения ключевых слов (keywords), методы дистрибутивной семантики и векторных представлений слов (word embeddings)									
Тема 4. Квантитативный синтаксис и теория сложности текста. Метрики читабельности (формулы Флеша — Кинкейда, индекс SMOG), синтаксическая насыщенность, автоматическое парсирование и статистические профили синтаксических конструкций	14	2	2			4		10	10
Тема 5. Авторская идиостилистика и атрибуция текста (стилометрия). Количественные маркеры	14	2	2			4		10	10

индивидуального стиля, методы главных компонент (PCA) и кластерного анализа для определения авторства анонимных текстов, проблема фальсификации данных									
Тема 6. Диахроническая квантитативная лингвистика. Отслеживание языковых изменений через большие временные корпуса, скорость лексической замены, закон регулярных соответствий и количественная верификация глоттохронологии	14	2	2			4		10	10
Тема 7. Квантитативные методы в социолингвистике и диалектологии. Статистический анализ вариативности,	12	2	2			4		8	8

логистическая регрессия для предсказания социолингвистических переменных, измерение расстояний между языковыми вариантами									
Тема 8. Прикладная квантитативная лингвистика: NLP и языковая статистика. Языковые модели, перплексия как мера качества модели, автоматическая обработка естественного языка, оценка качества машинного перевода и извлечение информации из текстов	14	2	2			4		10	10
Промежуточная аттестация *зачет*	*2*								
Итого	108	16	16					74	

6. Фонд оценочных материалов по дисциплине (модулю)

Типовые задания и иные материалы, необходимые для оценки результатов обучения по курсу «Квантитативная лингвистика»:

1. Ключевые слова (Keywords). С помощью статистических мер (логарифмическое правдоподобие — *Log-likelihood*, или взаимная информация — *MI*) выявить ключевые слова в корпусе текстов о нейросетях по сравнению с эталонным частотным словарем (например, Частотным словарем С. А. Шарова). Обосновать выбор метрики.

2. Коллокации. Найти статистически значимые коллокации (устойчивые сочетания) для слова «данные» в научном подкорпусе, используя меры *Dice*, *t-score* и *pMI*. Сравнить результаты и объяснить, почему разные меры выделяют разные типы сочетаний.

3. Динамика лексики. Проанализировать частотность слов «интернет», «нейросеть» и «цифровизация» в текстах СМИ за 2010–2025 годы. Построить график частотности, рассчитать коэффициент ранговой корреляции Спирмена для динамики этих слов и проверить гипотезу о синхронности их вхождения в язык.

4. Авторская атрибуция и стилометрия. Сравнить два подкорпуса прозы (например, ранние рассказы А. П. Чехова и Л. Н. Андреева) по синтаксическим признакам (средняя длина предложения, доля сложноподчиненных предложений, коэффициент гипотаксиса). Проверить статистическую значимость различий с помощью *t-критерия Стьюдента*.

5. Кластеризация. Выполнить иерархическую кластеризацию 10–15 стихотворений разных поэтов Серебряного века (А. Блок, А. Белый, О. Мандельштам) на основе векторов относительных частот служебных слов. Оценить качество кластеризации с помощью силуэтного коэффициента.

6. LDA. Построить тематическую модель (*LDA*) для корпуса «Ведомостей» XIX века. Определить оптимальное число тем (K) с помощью перплексии (*perplexity*) и когерентности (*topic coherence*), интерпретировать 5 наиболее вероятных топ-слов для каждой темы.

7. Семантические сдвиги. Построить дистрибутивные семантические модели (*Word2Vec* или *GloVe*) на корпусе текстов 1950-х и 2020-х годов. Вычислить косинусное расстояние между векторами слов «спутник», «партия», «сеть» в двух моделях и визуализировать семантические сдвиги на *t-SNE*-графике.

8. Анализ социальных сетей и прагматика. Провести анализ тональности (VADER или RuBERT) отзывов на маркетплейсе. Рассчитать метрики точности (*Precision*), полноты (*Recall*) и *F1-меры* путем ручной разметки тестовой выборки из 200 примеров.

9. Гендерная предикативность. Проверить гипотезу о наличии гендерных различий в использовании местоимений и частиц в постах Telegram-каналов с помощью критерия хи-квадрат и логит-регрессии.

10. Синтаксис и морфология Рассчитать объективные метрики читабельности (например, *Flesch–Kincaid* для английских текстов или индекс удобочитаемости *Оборина* для русских) для корпуса учебников разных классов. Построить регрессионную модель зависимости балла от средней длины слова и предложения.

17. Зависимостный парсинг. Оценить качество автоматического парсера (*UDPipe*, *SyntaxNet*) на 100 случайных предложениях из НКРЯ. Рассчитать метрики *UAS* (без учета меток) и *LAS* (с учетом меток), классифицировать типичные ошибки модели (например, неверное определение подлежащего).

18. Морфологическая продуктивность. Оценить продуктивность словообразовательной модели *пре-* в современном русском языке на корпусных данных. Рассчитать индекс Баайена (*Baayen's P*) для измерения неологизмов в реальном времени.

Рекомендации по структуре заданий:

- Каждое задание должно сопровождаться кратким теоретическим введением (0,5–1 страница), где обосновывается выбор статистического метода.
- Предполагается наличие не только численного ответа, но и его лингвистической интерпретации: что полученный коэффициент или график говорит о языке.

- Обязательный элемент — визуализация данных (графики распределений, тепловые карты, диаграммы рассеяния).

Задания по курсу для самостоятельной работы студентов:

1) Корпусная статистика и очистка данных. Студенту выдается «сырой» текстовый файл (например, логи чатов или оцифрованная художественная проза). Задача: написать скрипт (на Python или в R) для подсчета общего числа токенов, уникальных типов, среднего размера предложения и соотношения *type-token ratio* (TTR). Дополнительно требуется очистить данные от пунктуации, стоп-слов и привести слова к нижнему регистру, а затем пересчитать метрики и объяснить, как предобработка изменила картину.

2) Закон Ципфа и ранжирование частот. На основе данных из первого задания построить график распределения частот слов в логарифмических осях. Задача: проверить соответствие закону Ципфа, вычислить коэффициент наклона регрессионной прямой, найти отклонения (выбросы) в топ-10 самых частотных слов и объяснить их природу (например, влияние жанра или специфики корпуса).

3) Коллокации и статистические меры ассоциации. Извлечь из корпуса коллокации (устойчивые сочетания) для заданного опорного слова (например, «дело», «система» или «время»). Задача: рассчитать для топ-20 кандидатов меры Dice, MI (Mutual Information) и Log-Likelihood. Сравнить результаты: какие сочетания выявляет каждая мера, почему MI склонен отдавать предпочтение редким словам и какой метод лучше подходит для лексикографического описания.

4) Ключевые слова (Keywords) и контрастивный анализ. Сравнить два корпуса: например, новостные тексты 1990-х и 2020-х годов, или научные статьи по физике и лингвистике. Задача: используя меру Log-Likelihood или *t-критерий Стьюдента*, выявить ключевые слова, которые статистически значимо отличают один корпус от другого. Построить облака слов для визуализации и интерпретировать топ-10 ключевых слов с точки зрения диахронии или дисциплинарных различий.

5) Кластеризация слов по дистрибутивному сходству. На основе данных о совместной встречаемости слов в окне (например, 5 слов слева и справа) построить матрицу контекстов. Задача: применить метод *t-SNE* или *UMAP* для снижения размерности и алгоритм *K-means* или иерархическую кластеризацию для группировки слов. Проанализировать полученные семантические поля (например, кластер глаголов движения или кластер эмоциональной лексики) и оценить качество кластеризации.

6) Тематическое моделирование (LDA). Студенту предоставляется размеченный корпус текстов нескольких жанров (политика, спорт, культура). Задача: настроить модель латентного размещения Дирихле (*LDA*), подобрать оптимальное число тем (используя перплексию или когерентность), визуализировать распределение тем по документам и интерпретировать топ-10 слов в каждой теме. В качестве проверки нужно сопоставить доминирующие темы с исходными жанрами текстов.

7) Авторская атрибуция (стилометрия). Дается корпус текстов от 3–5 известных авторов. Задача: извлечь числовые векторы стиля (частоты служебных слов, средняя длина слова, разнообразие лексики, *Mendenhall's characteristic curves*) и обучить простой классификатор (например, логистическую регрессию или метод ближайших соседей). Оценить точность модели на отложенной выборке и определить, какие именно признаки оказались наиболее дискриминативными для каждого автора.

8) Анализ тональности (Sentiment Analysis). Используя размеченный датасет отзывов (например, о книгах или фильмах), обучить модель бинарной классификации (позитив/негатив). Задача: реализовать два подхода — «мешок слов» (*Bag-of-Words*) с регуляризацией и предобученные векторные представления (*Word2Vec* или *FastText*). Сравнить метрики качества (*Precision*, *Recall*, *F1-score*) и проанализировать типичные ошибки классификатора на сарказме и многозначной лексике.

9) Автоматическое извлечение терминов (Term Extraction). На корпусе узкоспециализированных текстов (например, по машинному обучению или молекулярной биологии) применить статистические (*C-value/NC-value*) и лингвистические (шаблоны *N-грамм*, POS-теги) методы для извлечения кандидатов в термины. Задача: составить глоссарий из 50 терминов, оценить точность и полноту ручной проверки, а также визуализировать семантическую сеть связей между извлеченными терминами.

10) Планирование лингвистического эксперимента и статистическая проверка гипотез. Формулируется гипотеза (например, «в русской разговорной речи предлоги "в" и "на" с предложным падежом конкурируют не случайно, а зависят от семантики места»). Задача: спроектировать сбор данных (выборка из устного подкорпуса НКРЯ), определить переменные, рассчитать необходимый объем выборки для статистической значимости и описать, какой тест будет применен (хи-критерий Пирсона или точный тест Фишера) для проверки нулевой гипотезы.

Каждое задание рассчитано на самостоятельное выполнение вне аудитории, требует написания кода, работы с данными и заканчивается кратким аналитическим отчетом (2–3 страницы) с визуализациями и лингвистической интерпретацией полученных цифр.

Шкала и критерии оценивания *(шкала и критерии оценивания могут быть едиными (типовыми) для всех дисциплин (модулей), входящих в основную образовательную программу)*

Оценка качества учебных достижений обучающихся по дисциплине осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Темы докладов

1. Цифровые корпуса текстов как инструмент лексикографического анализа (на примере Национального корпуса русского языка). Специфика использования подкорпусов для исследования исторической семантики.
2. Автоматизированные методы стилистического анализа текста: применение стилометров для атрибуции анонимных произведений и выявления авторского идиостиля.
3. Цифровая эдиция (Digital Scholarly Editing): принципы подготовки, визуализации и комментирования рукописного наследия в академических онлайн-проектах.
4. Технологии Text Mining и дистрибутивная семантика в исследовании эволюции поэтических образов (на материале русской лирики XIX–XXI веков).
5. Виртуальные реконструкции историко-культурных ландшафтов в филологической деятельности: от создания ГИС-карт литературных путешествий до геопоэтики.

6. Специфика работы с фольклорными информационными ресурсами: базы данных сказочных сюжетов (типа СУС) и методы их количественного анализа.
7. Машинное обучение в решении задач межтекстовых связей: алгоритмы автоматического поиска цитат, аллюзий и реминисценций в больших массивах прозы.
8. Репрезентация диалектной лексики в современных лексикографических базах данных: проблемы полевого сбора, метаданных и пользовательских интерфейсов.
9. Информационное обеспечение переводоведения: CAT-инструменты, переводческая память и терминологические базы данных в работе с художественным текстом.
10. Социальные сети и блоги как новый тип филологического источника: методы сбора, верификации и этические проблемы анализа пользовательского контента.
11. Создание и анализ мультимодальных литературных ресурсов: интеграция текста, аудио (начитки), видео и гипертекста в едином исследовательском пространстве.
12. Филологическая критика данных (Data Criticism): оценка репрезентативности, полноты и предвзятости (bias) литературных баз данных и цифровых библиотек.
13. Использование технологий дополненной и виртуальной реальности (AR/VR) для иммерсивного анализа пространства в литературе (на примере романов Достоевского или Булгакова).
14. Информационные ресурсы в преподавании родного и иностранного языков: проектирование адаптивных обучающих сред на основе анализа данных об успеваемости.
15. Библиометрические и альтметрические методы в оценке научного вклада филологов и динамики развития научных направлений (анализ цитируемости, упоминаний в медиа).
16. Архитектура и наполнение тезаурусов и семантических сетей для автоматической обработки текстов (на материале русского языка).
17. Онтологическое моделирование художественного мира писателя: формализация персонажей, хронотопа и сюжетных линий для семантического поиска.

18. Информационная поддержка сравнительно-исторического языкознания: базы данных родства языков, этимологические онлайн-словари и филогенетические алгоритмы.
19. Анализ тональности и сентимент-анализ в литературоведении: автоматическое выявление эмоциональных доминант в драматургии и эпистолярных жанрах.
20. Цифровое сохранение исчезающих малых языков России: создание электронных архивов, словарей и корпусов (на примере финно-угорских или тюркских языков).

Рекомендации по выбору и формулировке темы:

- Актуализация: любую из предложенных тем можно сузить, ограничив материал конкретным автором (например, «Цифровые методы анализа синтаксиса в романах Андрея Белого»), историческим периодом или типом ресурса.
- Методологический баланс: для магистерского реферата важно соблюсти баланс между филологической проблемой (что исследуется) и информационным ресурсом или технологией (как исследуется).

Практическая часть: большинство тем предполагает не только обзор литературы, но и описание работы с конкретным инструментом (например, выгрузка данных из НКРЯ, построение облака слов, анализ метаданных в электронной библиотеке).

Требования к структуре и содержанию доклада:

1) регламент выступления - 10 минут; наличие текста доклада и компьютерной презентации (презентация и текст доклада не должны дублировать друг друга);

2) использование самостоятельно подобранного текстового материала и его анализа в аспекте, определяемом темой доклада;

3) наличие учебного задания для студентов группы по теме доклада.

Критерии оценки доклада

Критерии	Показатели
Наличие необходимых элементов	Есть письменный текст доклада, компьютерная презентация, не дублирующие друг друга, студент выступил с докладом
Соблюдение требований к структуре и содержанию доклада	В докладе используется самостоятельно подобранный текстовый материал, есть его анализ и учебное задание

Шкала оценивания доклада (в баллах):

- все требования соблюдены, необходимые элементы есть, выполнено качественно: 15 б.

- все требования соблюдены, необходимые элементы есть, выполнено качественно, но презентация и текст доклада дублируют друга друга – 10 б

- не все требования соблюдены, не все необходимые элементы в наличии (нет текста доклада/презентации/нет самостоятельно подобранного текстового материала/задания – 5-7 баллов).

- доклад не представлен – 0 баллов.

7. Методические материалы

7.1. Перечень основной и дополнительной учебной литературы и иные материалы

Основная литература:

1. Шунейко А. А., Авдеенко И. А. Квантитативная лингвистика и новые информационные технологии. — М.: Юрайт, 2023. — 348 с. (ISBN 978-5-534-15446-7).

2) Гребенщикова А. В. Основы квантитативной лингвистики и новых информационных технологий: учебное пособие. — 3-е изд., стер. — М.: ФЛИНТА, 2018. — 152 с. (ISBN 978-5-9765-2137-7).

Дополнительная литература:

1. Копотев М. Введение в корпусную лингвистику: Учебное пособие для студентов филологических и лингвистических специальностей университетов. — М.: Языки славянской культуры, 2014. — 231 с.
2. Старостина С. А. Информационные технологии в филологии: учебное пособие / С. А. Старостина; Тамбовский государственный университет имени Г. Р. Державина. — Тамбов: Издательский дом "Державинский", 2022
3. Гуслякова А. В. Информационные технологии и лингвистика XXI века [Текст] : учебное пособие / А. В. Гуслякова ; М-во образования и науки Российской Федерации, Федеральное гос. бюджетное образовательное учреждение высш. образования Московский пед. гос. ун-т. - Москва : МГПУ, 2016. - 95 с. : ил.; 20 см.; ISBN 978-5-4263-0398-0

Словари и справочники:

1. Энциклопедия «Русский язык»/Институт русского языка им. В.В. Виноградова РАН, Гл. ред. А.М. Молдован. – 3-е изд., переработанное и дополненное. – М.: АСТ-ПРЕСС ШКОЛА, 2020. – 904 с. – ISBN: 978-5-907126-47-3

Периодические издания:

1. Вопросы языкознания: Научный журнал. - https://vja.ruslang.ru/ru/archive (Архив журнала с открытым доступом в Интернет).

7.2. Перечень лицензионного программного обеспечения

8. Ресурсное обеспечение

8.1. Перечень профессиональных баз данных и информационных справочных систем:

Наименование	Ссылка на ресурс	Доступность (свободный доступ / ограниченный доступ)
1	*2*	*3*
Информационно-справочные системы		
Библиографические базы данных по общественным наукам ИНИОН	http://inion.ru/	Свободный доступ
Методология: словарь системы основных понятий. Сайт академика РАО Новикова А.М.	http://www.anovikov.ru/dict/met_sl.htm	Свободный доступ

8.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

Наименование	Ссылка на ресурс	Доступность (свободный доступ / ограниченный

		доступ)
1	*2*	*3*
Электронно-библиотечные системы		
ЭБС Юрайт	https://urait.ru	После регистрации с компьютера Института – доступ с любого компьютера, подключенного к сети Интернет.
Профессиональные базы данных		
Справочно-информационный портал ГРАМОТА.РУ – русский язык для всех	http://gramota.ru/	Свободный доступ
«ИНФРА–М»	http://znanium.com	Свободный доступ
BOOK.ru	https://www.book.ru	После регистрации с компьютера Института – доступ с любого компьютера, подключенного к сети Интернет

Научная электронная библиотека РИНЦ	https://elibrary.ru/	Свободный доступ
Научная электронная библиотека «КиберЛенинка»	https://cyberleninka.ru/	Свободный доступ

9. Язык преподавания: русский

10. Разработчики программы

Лунькова Е. С., д. филол. наук, профессор кафедры общего и русского языкознания ФГБОУ ВО «Государственный институт русского языка им. А. С. Пушкина»