

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Нижегородский государственный лингвистический университет
им. Н.А. Добролюбова»
(НГЛУ)

УТВЕРЖДАЮ
Ректор
_____ / Н.В. Авралев
«__» _____ 2026 г.

ДОПОЛНИТЕЛЬНАЯ ОБРАЗОВАТЕЛЬНАЯ ПРОГРАММА
ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

Компьютерная лингвистика

Нижний Новгород
2026

Руководитель программы:
Старший научный сотрудник
научно-образовательного центра
дизайна и проектирования
инновационной среды
дополнительного образования,
канд. филол. наук, доцент

М.Б. Чиков

Разработчик программы:
Старший научный сотрудник
научно-образовательного центра
дизайна и проектирования
инновационной среды
дополнительного образования,
канд. филол. наук, доцент

М.Б. Чиков

Согласовано:
Проректор по учебной работе
канд. филол. наук, доцент

Н.А. Кохан

И.о. директора
Института непрерывного образования
канд. пед. наук

Л.С. Павлова

Оглавление

1. Общие положения	4
2. Цель и задачи программы.....	4
3. Планируемые результаты освоения программы	5
4. Учебный план	6
5. Календарный учебный график.....	7
6. Содержание программы	8
7. Организационно-педагогические условия реализации программы	8
8. Формы аттестации.....	8
9. Фонд оценочных средств.....	9
10. Учебно-методическое обеспечение	9

1. Общие положения

1.1. Актуальность, новизна, особенности

Программа «Введение в компьютерную лингвистику» разработана с учётом устойчивого спроса на специалистов, владеющих базовыми инструментами анализа языковых данных. Актуальность программы обусловлена ростом применения корпусных методов, автоматизированных систем перевода и технологий обработки устной речи в лингвистических исследованиях и преподавании. Программа сохраняет проверенную структуру предыдущих лет, но актуализирует ключевые инструменты: используются корпуса **Национальный корпус русского языка (НКРЯ)** и немецкий **DeReKo**, система машинного перевода, библиотека **NLTK** для Python и программы **Praat** и **ELAN** для анализа устной речи. Особенность программы — её практико-ориентированность и адаптация к асинхронному онлайн-формату.

1.2. Нормативная база

Программа разработана с учётом требований следующих нормативно-правовых актов:

- Федеральный закон от 29.12.2012 № 273-ФЗ «Об образовании в Российской Федерации»;
- Приказ Министерства науки и высшего образования Российской Федерации от 04.03.2025 № 266 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам»;
- Национальная стратегия развития искусственного интеллекта на период до 2030 года (Указ Президента РФ от 10.10.2019 № 490);
- Профессиональные стандарты в области информационных технологий и цифровой экономики.
- Устав Федерального государственного бюджетного образовательного учреждения высшего образования «Нижегородский государственный лингвистический университет им. Н.А. Добролюбова», утвержден приказом Министерства науки и высшего образования Российской Федерации от 14.12.2018 г. № 1166.
- Локальные нормативно-правовые акты НГЛУ, касающиеся организации образовательной деятельности.

1.3. Категория слушателей

К обучению приглашаются: преподаватели вузов и школ, лингвисты, переводчики, исследователи, а также все желающие освоить базовые инструменты компьютерной лингвистики.

Для обучения необходимо наличие законченного высшего или среднего профессионального образования.

1.4. Объем программы, срок обучения

Общий объем программы составляет 72 академических часа, включая все виды учебной работы. Срок обучения — 1 месяц.

1.5. Форма обучения

Форма обучения — заочная, с применением дистанционных образовательных технологий (асинхронный онлайн-курс). Обучение проходит без отрыва от основной деятельности.

2. Цель и задачи программы

2.1. Цель программы

Формирование у слушателей систематизированных знаний и практических навыков в области компьютерной лингвистики, включая работу с корпусами, основы

автоматического перевода, базовое программирование на Python с использованием NLTK, а также анализ устной речи с помощью специализированного ПО.

2.2. Задачи программы

1. Познакомить с основными понятиями и направлениями компьютерной лингвистики.
2. Обеспечить практическое освоение национальных корпусов (НКРЯ, DeReKo).
3. Раскрыть принципы работы систем машинного перевода на примере Freyja.
4. Научить базовой обработке текстов с помощью библиотеки NLTK в Python.
5. Обеспечить освоение методов анализа устной речи в программах Praat и ELAN.

3. Планируемые результаты освоения программы

В результате освоения программы слушатель должен:

Знать:

- Основные задачи и методы компьютерной лингвистики.
- Структуру и функционал корпусов НКРЯ и DeReKo.
- Архитектурные принципы систем машинного перевода.
- Базовые возможности библиотеки NLTK для обработки текста.
- Назначение и интерфейс программ Praat и ELAN.

Уметь:

- Выполнять запросы в корпусах НКРЯ и DeReKo.
- Интерпретировать результаты корпусного поиска.
- Проводить простую обработку текста (токенизация, частеречная разметка) с помощью NLTK.
- Анализировать акустические параметры речи в Praat.
- Аннотировать устные высказывания в ELAN.

Владеть:

- Терминологией компьютерной лингвистики.
- Навыками самостоятельной работы с корпусными ресурсами.
- Базовыми командами NLTK.
- Методикой описания речевых данных.

4. Учебный план

-	-	Формы пром. атт.	Итого акад. часов					Курс 1				Закрепленная кафедра	
								-					
Индекс	Наименование	Зачет	Экспертное	По плану	Ауд.	СР	Контроль	Итого	Ауд	СР	Контроль	Код	Наименование
Дисциплины (модули)			72	72		71	1	72		71	1		
01	Введение в компьютерную лингвистику. Основные направления и задачи		14	14		14		14		14			Институт непрерывного образования
02	Корпусные исследования: НКРЯ и DeReKo		14	14		14		14		14			Институт непрерывного образования
03	Автоматизация перевода: принципы работы системы переводческой памяти		14	14		14		14		14			Институт непрерывного образования
04	Основы программирования на Python: библиотека NLTK		14	14		14		14		14			Институт непрерывного образования
05	Обработка устной речи: Praat и ELAN		15	15		15		15		15			Институт непрерывного образования
06	Итоговое тестирование		1	1			1	1			1		Институт непрерывного образования

5. Календарный учебный график

с – самостоятельная работа, * – неучебные дни, = – неучебные недели

Мес	Сентябрь					Октябрь				Ноябрь				Декабрь					Январь				Февраль				Март					Апрель				
Пн	1	8	15	22	29	6	13	20	27	3	10	17	24	1	8	15	22	29	5	12	19	26	2	9	16	23	2	9	16	23	30	6	13	20	27	
Вт	2	9	16	23	30	7	14	21	28	4	11	18	25	2	9	16	23	30	6	13	20	27	3	10	17	24	3	10	17	24	31	7	14	21	28	
Ср	3	10	17	24	1	8	15	22	29	5	12	19	26	3	10	17	24	31	7	14	21	28	4	11	18	25	4	11	18	25	1	8	15	22	29	
Чт	4	11	18	25	2	9	16	23	30	6	13	20	27	4	11	18	25	1	8	15	22	5	12	19	26	5	12	19	26	2	9	16	23	30		
Пт	5	12	19	26	3	10	17	24	31	7	14	21	28	5	12	19	26	2	9	16	23	6	13	20	27	6	13	20	27	3	10	17	24	1		
Сб	6	13	20	27	4	11	18	25	1	8	15	22	29	6	13	20	27	3	10	17	24	7	14	21	28	7	14	21	28	4	11	18	25	2		
Вс	7	14	21	28	5	12	19	26	2	9	16	23	30	7	14	21	28	4	11	18	25	1	8	15	22	1	8	15	22	5	12	19	26	3		
Нед	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	
Пн																			*						*											
Вт										*																										
Ср																			*																	
Чт	=	=	=	=	=	=	=	=	=		=	=	=	=	=	=	=	*	*			=	=	=	=			18	18	18	18	=	=	=		
Пт																		*																*		
Сб																		*																		
Итого																												72 с								

*Программа реализуется по индивидуальному графику для каждого обучающегося

6. Содержание программы

Модуль 1. Введение в компьютерную лингвистику. Основные направления и задачи

Теория: Предмет и задачи компьютерной лингвистики. История развития. Связь с NLP, искусственным интеллектом, теоретической и прикладной лингвистикой.

Практика: Чтение обзорных материалов, тест по терминологии.

Модуль 2. Корпусные исследования: НКРЯ и DeReKo

Теория: Понятие языкового корпуса. Структура, метаданные, типы запросов. Сравнение НКРЯ и DeReKo.

Практика: Выполнение корпусных запросов (частотность, контексты, коллокации).

Модуль 3. Автоматизация перевода: принципы работы системы переводческой памяти

Теория: Архитектура rule-based и пример-based систем. Этапы перевода: анализ, трансфер, генерация.

Практика: Анализ примеров перевода, сравнение с современными нейросетевыми системами.

Модуль 4. Основы программирования на Python: библиотека NLTK

Теория: Установка Python и NLTK. Основные функции: токенизация, стемминг, лемматизация, POS-теггинг.

Практика: Написание простых скриптов для обработки текста.

Модуль 5. Обработка устной речи: Praat и ELAN

Теория: Акустические параметры речи (F0, интенсивность, форма). Принципы аннотации.

Практика: Загрузка аудиофайла в Praat, измерение параметров; создание аннотации в ELAN.

Модуль 6. Практические задания и мини-проекты

Практика: Выполнение 5 заданий по модулям + итоговый мини-проект (например: анализ корпуса + скрипт NLTK + описание речевого фрагмента).

7. Организационно-педагогические условия реализации программы

7.1. Условия привлечения профессорско-преподавательского состава

Реализация дополнительной профессиональной программы (программы повышения квалификации) обеспечивается научно-педагогическими кадрами НГЛУ, допускается привлечение внешних экспертов в области корпусной лингвистики и NLP.

7.2. Требования к квалификации профессорско-преподавательского состава

Педагогические работники должны иметь профильное филологическое или лингвистическое образование, опыт работы с корпусами, NLTK, Praat/ELAN, ученую степень и звание.

8. Формы аттестации

8.1. Текущий контроль

Проводится в форме выполнения онлайн-тестов по 5 модулям + проверка практических заданий в ЭИОС дополнительного образования НГЛУ.

8.2. Итоговая аттестация

Проводится в форме защиты мини-проекта или итогового тестирования.

8.3. В соответствии с частью 3 и частью 10 статьи 60 Федерального закона от 29 декабря 2012 г. N 273-ФЗ «Об образовании в Российской Федерации» лицам, успешно прошедшим итоговую аттестацию, выдается документ о квалификации (удостоверение о повышении квалификации), оформленный на основании Методических рекомендаций

Министерства образования и науки Российской Федерации по разработке, заполнению, учету и хранению документов о квалификации от 12 марта 2015 г. N АК-610/06.

В соответствии с частью 12 статьи 60 Федерального закона от 29 декабря 2012 г. N 273-ФЗ «Об образовании в Российской Федерации» лицам, не прошедшим итоговую аттестацию или получившим на итоговой аттестации неудовлетворительные результаты, а также лицам, освоившим часть дополнительной профессиональной программы (программы повышения квалификации) и (или) отчисленным из организации, осуществляющей образовательную деятельность, выдается справка об обучении или о периоде обучения.

9. Фонд оценочных средств

9.1. Примерные вопросы

1. Какие типы запросов поддерживаются в НКРЯ?
2. В чём разница между rule-based и neural MT?
3. Какие функции NLTK используются для токенизации?
4. Какие параметры можно измерить в Praat?
5. Для чего используется ELAN?

9.2. Критерии оценивания

Успешное освоение — выполнение всех заданий. Успешное прохождение тестирования в автоматическом режиме предполагает правильный ответ на 60 и более процентов вопросов. Освоение программы оценивается по двухбалльной системе: «зачтено», «не зачтено».

10. Учебно-методическое обеспечение

Основная литература

1. Bird, S., Klein, E., Loper, E. Natural Language Processing with Python. O'Reilly, 2009.
2. Национальный корпус русского языка: <https://ruscorpora.ru>
3. DeReKo: <https://www.ids-mannheim.de/derewo>

Дополнительная литература

1. Freyja: Popel, M. et al. The Freyja Machine Translation System. Prague Bulletin of Mathematical Linguistics, 2017.
2. Voersma, P., Weenink, D. Praat: doing phonetics by computer.
3. ELAN Manual. Max Planck Institute.

Интернет-ресурсы

1. NLTK Book: <https://www.nltk.org/book/>
2. Официальный сайт Praat
3. Руководство по ELAN