



УТВЕРЖДАЮ
Первый проректор-проректор по учебной работе
А.Б. Безбородов
«24» июня 2016 г.

**Образовательная программа высшего образования по направлению подготовки
45.04.03 Фундаментальная и прикладная лингвистика
Магистерская программа: Компьютерная лингвистика**

Квалификация (степень) – *магистр*

Форма обучения – *очная*

Образовательная программа разработана в соответствии с *Федеральным государственным образовательным стандартом высшего образования по направлению подготовки 45.04.03 Фундаментальная и прикладная лингвистика (уровень магистратуры), утвержденным приказом Минобрнауки России от 03.12.2015 г. № 1407).*

Срок освоения образовательной программы – 2 года;
Трудоемкость образовательной программы – 120 *зачетные единицы.*

Цель программы – подготовка профессиональных лингвистов, способных эффективно работать в области разработки инновационных языковых компьютерных технологий. Компьютерная лингвистика является относительно новой областью научной и инженерной деятельности. Актуальность создания программы определяется тем, что в последние годы наблюдается бурное развитие этой области, связанное со все растущим влиянием Интернета и появлением огромного количества новых технических устройств, важнейшей частью которых являются естественно-языковые интерфейсы. Кроме того, в современной лингвистике происходит стремительный переход от традиционных методов получения языковых данных к корпусным методам, требующим серьезного развития компьютерных технологий. Поэтому целью программы является подготовка специалистов, способных решать, в частности, такие задачи компьютерной лингвистики, как разработка новых исследовательских лингвистических методов и технологий, основанных на широком использовании методов машинного обучения, автоматического лингвистического аннотирования, синтаксического и семантического анализа текстов.

Характеристика профессиональной деятельности выпускников

Область профессиональной деятельности:

фундаментальные, функциональные и формальные исследования по общей теории языка, фонетике, морфологии, лексикологии, синтаксису, дискурсу, семантике различных языков, в области психолингвистики, социолингвистики и когнитивной лингвистики, в области типологического, ареального и генеалогически обусловленного разнообразия и диахронического развития языков, направленные на создание теоретической базы для компьютерно-лингвистических и иных приложений лингвистических знаний;

разработку электронных языковых ресурсов различного назначения (корпусов текстов и звучащей речи, словарей, онтологий, баз данных, лингвистического обеспечения для

электронных систем, предполагающих автоматическую обработку звучащей речи и письменных текстов на естественном языке

Объекты профессиональной деятельности:

феномены всех уровней и планов языковой структуры (фонетики, морфологии, лексики, синтаксиса, дискурса, семантики);

электронные языковые ресурсы (текстовые, речевые и мультимодальные корпуса, словари, тезаурусы, онтологии, фонетические, лексические, грамматические и иные базы данных и базы знаний);

лингвистические технологии, применяемые в электронных системах различного назначения (поисковых системах, системах машинного перевода, системах управления, системах обработки звучащей речи, экспертных системах, электронных языковых ресурсах)

Виды профессиональной деятельности:

научно-исследовательская, педагогическая, производственно-практическая, экспертно-аналитическая, организационно- управленческая

Планируемые результаты освоения образовательной программы

В результате освоения программы у выпускника должны быть сформированы общекультурные, общепрофессиональные и профессиональные компетенции. Выпускник, освоивший программу должен обладать:

способностью к абстрактному мышлению, анализу, синтезу (ОК-1);

готовностью действовать в нестандартных ситуациях, нести социальную и этическую ответственность за принятые решения (ОК-2);

готовностью к саморазвитию, самореализации, использованию творческого потенциала (ОК-3);

способностью свободно вести коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном языке для решения задач профессиональной деятельности (ОПК-1);

готовностью руководить коллективом в сфере своей профессиональной деятельности, толерантно воспринимая социальные, этнические, конфессиональные и культурные различия (ОПК-2);

владением знаниями об истории лингвистических учений и современного состояния лингвистической науки (ОПК-3);

способностью к осознанию современного состояния в области компьютерной лингвистики и информационных технологий (ОПК-4);

способностью анализировать, сопоставлять и критически оценивать различные лингвистические направления, теории и гипотезы (ОПК-5);

способностью адаптироваться к новым теориям и результатам мировой науки и расширять сферу научной деятельности, участвовать в междисциплинарных исследованиях на стыке наук (ОПК-6);

способностью выбирать оптимальные теоретические подходы и методы решения конкретных научных задач в области лингвистики и новых информационных технологий (ОПК-7);

способностью проводить самостоятельные исследования и получать новые научные результаты в области теории языка, лингвистики конкретных языков, прикладной и компьютерной лингвистики (ПК-1);

способностью изучать и осваивать современные технические средства и информационные технологии, служащие для обеспечения лингвистической деятельности (ПК-2);

способностью самостоятельно планировать и проводить научные эксперименты (в том числе, при наличии подобного оборудования, с использованием высокоточных методов регистрации мозговой активности (электроэнцефалография (вызванные потенциалы) и функциональная магнитно-резонансная томография) и движений глаз) (ПК-3);

способностью планировать, организовывать и реализовывать образовательный процесс по отдельным видам учебных занятий (лабораторные, практические и семинарские занятия) по лингвистическим дисциплинам (модулям) в образовательных организациях высшего образования (ПК-4);

способностью разрабатывать под руководством специалиста более высокой квалификации учебно-методическое обеспечение реализации учебных дисциплин (модулей) или отдельных видов учебных занятий программ бакалавриата и дополнительных профессиональных программ для лиц, имеющих или получающих соответствующую квалификацию, рецензировать и проводить экспертизу научно-методических и учебно-методических материалов по лингвистическим дисциплинам (модулям) (ПК-5);

владеть навыками участия в организации научно-исследовательской, проектной, учебной, профессиональной и иной деятельности обучающихся по программам бакалавриата и дополнительным образовательным программам, в профориентационных мероприятиях со школьниками, навыками проведения научно-популярных и просветительских мероприятий, навыками педагогической поддержки профессионального самоопределения обучающихся по программам бакалавриата и дополнительным образовательным программам (ПК-6);

способностью разрабатывать лингвистические компоненты электронных языковых ресурсов (лингвистические корпуса, словари, онтологии, базы данных) (ПК-7);

способностью разрабатывать системы автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистические компоненты интеллектуальных и информационных электронных систем (ПК-8);

способностью разрабатывать и внедрять в практику компьютерные системы обучения (ПК-9);

способностью разрабатывать и совершенствовать системы автоматизации и информационной поддержки лингвистических исследований (ПК-10);

владением навыками перевода научной литературы по лингвистике и смежным дисциплинам с иностранных языков на государственный язык Российской Федерации и с государственного языка Российской Федерации на иностранный язык со снабжением ее необходимым редакторским и издательским комментарием и научным аппаратом (ПК-11);

способностью проводить экспертный анализ различных типов устного и письменного дискурса с целью извлечения знаний, определения логической структуры и тональности текста, определения социолингвистических характеристик или эмоционального состояния говорящего, идентификации личности говорящего, его коммуникативных намерений и с другими прикладными целями, в том числе с целью ответа на входящие в компетенцию лингвиста вопросы при проведении экспертного исследования письменных текстов и фонограмм в процессе судопроизводства (ПК-12);

способностью вести мониторинг информационных массивов и готовить на этой основе аналитические материалы (ПК-13);

способностью производить систематизацию произвольной предметной области, разрабатывать для нее классификаторы, рубрикаторы, онтологии, проводить типологизацию данных и моделирование предметной области, разрабатывать универсальные онтологии (ПК-14);

способностью давать экспертные оценки и разрабатывать рекомендации в сфере

*языковой политики и языкового планирования (ПК-15);
способностью давать экспертные оценки и разрабатывать рекомендации в сфере
языковой политики и языкового планирования (ПК-16);
владением навыками участия в организации и проведении конференций, симпозиумов,
семинаров, переговоров и других мероприятий с использованием нескольких рабочих
языков (ПК-17).*

Преимущества и особенности образовательной программы в РГГУ

Концепция данной магистерской программы соответствует общей концепции кафедры компьютерной лингвистики РГГУ, в основе которой лежит положение, что область деятельности, связанная с решением задач автоматической обработки Естественного Языка (ЕЯ) и именуемая «Компьютерная лингвистика», требует подготовки специалистов двух принципиально отличающихся направлений: собственно, лингвистов и инженеров. Эти направления базируются на двух совершенно разных системах образования и компетенциях:

1. «Компьютерная лингвистика для инженеров» является частью так называемой Computer Science. В рамках направления готовятся инженеры, способные эффективно решать задачи автоматической обработки ЕЯ, опираясь на необходимые для конкретной задачи существующие лингвистические ресурсы и модели. Кафедра способствует появлению таких специалистов, взаимодействия с техническими вузами. В частности, с участием кафедры компьютерной лингвистики РГГУ создана «параллельная» магистерская программа по компьютерной лингвистике для инженеров в МФТИ.
2. «Компьютерная лингвистика для лингвистов» является разделом теоретической и прикладной лингвистики. В рамках этого направления готовятся лингвисты, способные решать задачи создания формальных языковых моделей и выполненных на их основе лингвистических ресурсов, обладающих необходимыми свойствами для их применения в задачах автоматической обработки ЕЯ. Именно это направление и реализуется данной программой.

Важнейшим обстоятельством, которое учитывается программой, является то, что специалисты, подготовленные в рамках этих двух направлений, являются необходимыми участниками любых серьезных проектов в области автоматической обработки ЕЯ. И хотя они выполняют существенно различающиеся функции, умение эффективно взаимодействовать между собой является ключевым фактором в успехе таких проектов. Основы такого взаимодействия должны быть заложены в программах за счет серьезной инженерной и математической подготовки лингвистов (и соответствующей лингвистической подготовки инженеров).

В отличие от авторов некоторых альтернативных программ подготовки компьютерных лингвистов мы не рассматриваем в качестве приоритетной для выпускников область так называемых стартапов: небольших проектов, в которых участникам по необходимости приходится решать сразу несколько задач: от лингвистики до программирования и маркетинга. Подготовка, полученная магистром данной специальности, может быть достаточной и для решения таких задач, но анализ ситуации и мнения экспертов показывают, что в таких проектах конкурентное преимущество имеют профессиональные инженеры, получившие направленную экспресс-подготовку по лингвистике.

Таким образом, подготовка магистров компьютерной лингвистики в данной программе основывается на глубоком изучении фундаментальных основ лингвистики с акцентом на методы создания операциональных формальных моделей языковой системы, адекватных сложности таких задач обработки естественного языка, как распознавание и синтез речи, машинный перевод, семантический анализ и понимание текста, интеллектуальный поиск.

Спецификой предлагаемой программы является наличие в ней нескольких блоков, отражающих междисциплинарный характер предлагаемого образования:

1. Формальные модели языка с акцентом на перспективах прикладного использования. Этот блок основывается на ряде курсов, общих с программой «Теория языка» того же базового направления подготовки **45.04.03** «Фундаментальная и прикладная лингвистика».
2. Математическая и инженерная подготовка. Статистика, формальные грамматики, представление о методах машинного обучения, эвристические методы Искусственного Интеллекта, экспертные системы, представление знаний.
3. Инструментальный блок: специализированные языки и среды разработки для лингвистов (типа Natural Language Toolkit R т.п.), специализированные лингвистические базы данных, открытые лингвистические ресурсы (от грамматик и парсеров до онтологий).
4. Прикладной задачный блок. Здесь рассматриваются отдельные наиболее актуальные задачи АОТ.

В процессе реализации программы решаются следующие основные **образовательные задачи**:

- Получение углубленных знаний в области теоретического языкознания, теоретической и практической лексикографии и смежных дисциплин.
- Владение современными методами научных исследований языка с применением новейших компьютерных технологий.

- Приобретение навыков создания операциональных формальных моделей языковой системы, адекватных сложности таких задач обработки естественного языка, как распознавание и синтез речи, машинный перевод, семантический анализ и понимание текста, интеллектуальный поиск.
- Последовательный типологический подход к задачам лингвистического моделирования, объектом которого может быть любой язык планеты и любые межъязыковые преобразования.
- Владение международными стандартами в области лингвистической разметки и аннотирования лингвистических ресурсов и современными методами оценки качества работы систем АОТ.

Таким образом, подготовка магистров по компьютерной лингвистике в данной программе основывается на глубоком изучении фундаментальных основ лингвистики с акцентом на методах операционального, типологически адекватного моделирования и создания универсальных стандартизованных лингвистических ресурсов.

Иновационность предлагаемой программы обеспечивается:

- использованием в учебных курсах результатов научной деятельности Института лингвистики в области исследования и формального моделирования языков и процессов языковой коммуникации;
- опорой на самые актуальные процессы в области компьютерного анализа языка, связанные с широким участием сотрудников кафедры в крупных исследовательских и коммерческих проектах в области АОТ.

Программа опирается одновременно и на уникальный кадровый потенциал Института лингвистики, и на широкую интеграцию процесса обучения магистров с мейнстримом современной компьютерной лингвистики через участие преподавателей и собственно обучаемых в российских и международных проектах и конференциях, таких, например, как международная конференция по компьютерной лингвистике «Диалог», соучредителем которой является Институт лингвистики РГГУ.

В числе разработчиков и ведущих преподавателей программы авторы базовых учебников по лингвистическим специальностям, специалисты мирового уровня, руководители проектов крупных систем автоматической обработки языка: Я.Г.Тестелец (синтаксис), И.М.Богуславский (лексическая семантика и лингвистические онтологии), В.И.Беликов (компьютерная социолингвистика), В.И.Подлеская (типология), А.Ч.Пипрески (корпусная лингвистика), В.П.Селегей (компьютерная лексикография, компьютерный анализ языков в системах ABYU Lingvo и ABYU Comprono), Л.Л.Иомдин (вычислительная семантика, компьютерный анализ языка в системе Этап-2 и

Этап-3), А.С.Старостин (синтаксический анализ в системе TRITON, лингвистические онтологии), С.А.Шаров (технологии компьютерного анализа и классификации текстов).

Планируется широкое привлечение к преподаванию сотрудников компаний, являющихся мировыми лидерами в области компьютерной лингвистики: IBM (система Watson), Яндекс и Google (информационный поиск, статистический машинный перевод), АBBYU (системы Lingvo, FineReader, Compreno).

Принципиальной особенностью подготовки магистров по данной программе является проектный подход и привлечение студентов к научно-исследовательской работе в области компьютерной лингвистики. Производственная и научно-исследовательская практики магистрантов организуются на базе действующих исследовательских и научно-образовательных проектов. Практика будет проводиться как на базе РГГУ: участие магистрантов в компьютерных проектах, специально организованных для этой цели, в том числе – совместных проектах со студентами кафедр компьютерной лингвистики других вузов (МФТИ, МГУ, ВШЭ), так и на базе компаний, занимающихся разработкой программ в области АОТ (Яндекс, АBBYU и др.). Основные направления: сбор и разметка лингвистических корпусов, корпусное исследование лексических, грамматических и дискурсивных явлений на материале русского, а также других языков и др.

Структура учебного плана программы, в соответствии с поставленными образовательными задачами, включает в себя следующие основные блоки дисциплин.

- Блок лингвистических дисциплин представлен курсами, «Введение в фундаментальную лингвистику», «Типология, компаративистика, ареальная лингвистика», «Современные синтаксические теории», «Корпусные и экспериментальные методы в семантике», «Корпусные и экспериментальные методы в семантике», «Введение в компьютерную лингвистику», «Компьютерная социолингвистика». Этот блок имеет ряд общих курсов с программой «Теория языка» и нацелен на глубокое изучение формальных моделей языка с акцентом на перспективах прикладного использования.
- Блок дисциплин математической и инженерной подготовки представлен курсами «Математические основы лингвистики», «Статистические модели в лингвистике», «Методы искусственного интеллекта в компьютерной лингвистике». Он нацелен на овладение магистрантами необходимыми знаниями в области практического применения статистики, формальных грамматик, методов машинного обучения, эвристических методов искусственного интеллекта, экспертных систем, представления знаний.

- Инструментальный блок представлен курсами «Программирование лингвистических задач», «Лингвистическое аннотирование/разметка текстов», «Специализированные лингвистические базы данных», «Методы оценки систем АОТ».
- Прикладной задачей блок представлен курсами «Модели и методы компьютерной лингвистики», «Методы классификации и машинное обучение», «Лингвистические основы машинного перевода», «Компьютерный синтаксический анализ», «Анализ устной речи».

Изучение иностраных языков в рамках данной программы ориентировано на их профессиональное использование как в качестве инструмента, так и в качестве объекта исследования. Одна из важнейших задач этого блока – обучить магистрантов навыкам написания научных статей, выступлений с докладами и другим формам академической коммуникации на английском языке (фактически выполняющем роль латыни в мировой науке) на высоком уровне требований, предъявляемых в современном профессиональном сообществе. Курс «Английский язык для профессиональной коммуникации» читается в течение 3-х семестров.

Карьерные перспективы

Выпускники могут продолжить научную карьеру как профессиональные лингвисты, получившие глубокие знания в области использования в профессиональной деятельности современных компьютерных методов, и в то же время – как специалисты в области создания и применения языковых моделей и ресурсов для задач автоматического анализа естественных языков, для задач компьютерной лексикографии и др. Выпускники программы будут профессионально задействованы в высших учебных заведениях, научно-исследовательских институтах, в компаниях, производящих программные продукты и в интернет-компаниях.