

**УТВЕРЖДЕНО**  
**Академическим советом ОП**  
**Компьютерная лингвистика**  
**факультета гуманитарных наук**  
**Национального исследовательского университета**  
**Высшей школы экономики**  
**Протокол от 01.03.2017 № 8**

**Концепция образовательной программы магистратуры**  
**«Компьютерная лингвистика»**

**1. ОБЩАЯ ХАРАКТЕРИСТИКА ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ.**

Образовательная программа магистратуры «Компьютерная лингвистика» разработана в соответствии с образовательным стандартом НИУ ВШЭ направления подготовки магистра 45.04.03 «Фундаментальная и прикладная лингвистика».

Нормативный срок освоения программы составляет 2 года.

Трудоемкость программы составляет 120 кредитов.

Программа предполагает только очную форму обучения.

Программа предусматривает возможность обучения на бюджетных и коммерческих местах.

В рамках программы студенты обучаются основам прикладной и компьютерной лингвистики, преподавание которых представляет собой интеграцию академической и практико-ориентированной компоненты. Это означает, что студенты в равной степени приобретают компетенции, связанные с научно-исследовательской деятельностью (планирование и самостоятельное осуществление научных исследований в разных областях лингвистики и/или компьютерной обработки естественного языка) и компетенции, связанные с прикладной деятельностью лингвиста (моделирование лингвистических компонент, разработка лингвистических приложений, словарей, интерактивных тренажеров для обучения языку и пр.) Особое внимание программа уделяет выработке компетенций, связанных с проектной деятельностью, в частности: планирование и реализация в срок задач проекта, эффективная работа в команде, представление результатов проекта.

Программа «Компьютерная лингвистика» ориентирована на прием как российских, так и иностранных учащихся. Поэтому преподавание на программе будет вестись на русском и английском языках. При этом будет возможно освоение учебного цикла программы полностью на английском языке, или же частично на английском и частично на русском языках в зависимости от выбора и языковых компетенций студентов. Соответственно магистерская диссертация может быть написана по-русски или по-английски.

Программа предусматривает партнерство (студенческий обмен и стажировки) со следующими университетами: университеты Ланкастера и Шеффилда (Великобритания), университет Хельсинки (Финляндия), университет Тромсе (Норвегия), университет INALCO (Франция).

Руководителем программы является к.ф.н, доцент факультета филологии НИУ ВШЭ А.А.Бонч-Осмоловская ([abonch@gmail.com](mailto:abonch@gmail.com))

В разработке программы, кроме руководителей программы, участвовали д.ф.н., профессор факультета филологии НИУ ВШЭ Е.В.Рахилина ([rakhilina@gmail.com](mailto:rakhilina@gmail.com)), старший преподаватель факультета филологии НИУ ВШЭ Д.А.Алексеевский ([me.dendik@gmail.com](mailto:me.dendik@gmail.com)), к.ф.н, доцент факультета филологии НИУ ВШЭ С.Ю. Толдова ([toldova@yandex.ru](mailto:toldova@yandex.ru)), к.ф.н, профессор факультета филологии НИУ ВШЭ О.Н.Ляшевская ([olesar@yandex.ru](mailto:olesar@yandex.ru))

## **2. АКТУАЛЬНОСТЬ, ЦЕЛИ, ЗАДАЧИ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ**

В современном информационном обществе объем доступных языковых данных, касающихся самых разных областей, экспоненциально вырос. Одновременно сформировались новые типы дискурса, а старые приобрели иной статус и характеристики. Помимо устного и письменного вариантов языка возникла интернет-коммуникация, и разговорная норма проникла в письменное цифровое пространство, а глобализация вызвала качественно новый виток межязыковой интерференции. Ввиду этого в самых разных социальных и научных областях образовалась острая потребность в теоретических и прикладных лингвистах, способных структурировать, анализировать, и обрабатывать значительные массивы данных такого рода. Сотрудничая с профессионалами разных предметных областей, на стыке разных специальностей, в смешанных командах, они создают и используют для решения сложных информационных задач новые инструменты – размеченные корпуса текстов, текстовые базы данных, электронные словари, интерактивные и краудсорсинговые языковые ресурсы различных типов и др.

При этом рынок специалистов, прошедших базовую подготовку по математике и информатике в сочетании с лингвистической подготовкой, опирающейся на современные теории о принципах организации естественного языка, до сих пор не насыщен, несмотря на то, что в последние годы в разных университетах России открыто несколько бакалаврских программ соответствующего профиля.

Более того, в связи с очень быстрым развитием рынка информационных технологий, связанных с автоматической обработкой естественного языка, спрос на лингвистов, обладающих современной комплексной подготовкой, постоянно растет.

Актуальность магистерской программы по теории языка и компьютерной лингвистике состоит в том, что она позволяет адаптировать и расширить профессиональные компетенции выпускников - традиционных лингвистических бакалавриатов, ориентированных в первую очередь на свободное владение иностранными языками, к потребностям современного информационного сообщества. Одновременно структура программы дает возможность профессиональной лингвистической подготовки для выпускников нелингвистических специальностей, в первую очередь тех, кто закончил бакалавриат или специалитет по прикладной математике или информатике.

Таким образом, целью программы является подготовка специалистов современного уровня в области теоретической и прикладной лингвистики через продолжение базового лингвистического образования, прежде всего по бакалаврской программе «Фундаментальной и компьютерная лингвистика», реализуемой в НИУ ВШЭ, а также через расширение профессиональных компетенций выпускников других программ.

Реализация этой цели предполагает в том числе полноценный выход программы на международный рынок с учетом уже установившихся и развивающихся партнерских отношений с близкими по направленности программами иностранных университетов. Международная ориентированность программы отражается в ее структуре, включающей в себя полностью англоязычный трек из основных и элективных курсов.

В задачи программы входит полноценная реализация двух профессиональных специализаций.

Задачи специализации по теории языка включают в себя углубленную подготовку специалистов прежде всего в следующих областях:

- Русский язык в диахроническом и типологическом освещении
- Русский как иностранный и как неродной
- Социолингвистические срезы русского языка и анализ языковой ситуации для разных его вариантов
- Проблемы языковой интерференции
- Типология языков и полевая лингвистика
- Современные формальные и функциональные теории и модели языка

Задачи специализации по компьютерной лингвистике ориентированы в первую очередь на подготовку специалистов в таких областях, как:

- Автоматический анализ естественного языка
- Электронные языковые ресурсы (языковые корпуса, электронные словари, базы данных и пр.);

- Цифровые технологии в гуманитарных науках
- Квантитативные и статистические аспекты теории языка
- Прикладные лингвистические инструменты
- Оценка лингвистических систем и компонентов

### **3. ЦЕЛЕВАЯ АУДИТОРИЯ ОП: КРИТЕРИИ НАБОРА СТУДЕНТОВ, ВЕЛИЧИНА ПРЕДПОЛАГАЕМОГО ПОТОКА (В Т.Ч. БЮДЖЕТНОЙ ЕГО ЧАСТИ, ПРИ ЕЕ НАЛИЧИИ)**

Целевая аудитория образовательной программы может быть очерчена следующим образом:

Для специализации по теории языка:

#### Российская аудитория

- Выпускники программ бакалавриата ОС ВПО «Фундаментальная и прикладная лингвистика» и «Лингвистика», ставящие своей целью получение углубленных профессиональных компетенций в области теории языка, в частности типологии и социолингвистики, в области прикладных технологий в русистике, в частности, преподаванием русского как иностранного или неродного, а также в областях языковой документации, языкового планирования

#### Зарубежная аудитория

- Выпускники зарубежных бакалавриатов, имеющие мэйджор или майнор по лингвистике, интересующиеся теорией языка, типологией, в частности, полевой лингвистикой на материале языков России, а также имеющие образование в области славистики, желающие пройти углубленный курс теории русского языка, методологии преподавания русского как иностранного и неродного

Для специализации по компьютерной лингвистике:

#### Российская аудитория

- Выпускники программ бакалавриата ОС ВПО «Фундаментальная и прикладная лингвистика», желающие продолжить свое образование и получить специализации в области автоматического анализа естественного языка или расширить его в области цифровых технологий для гуманитарных наук (digital humanities)
- Выпускники программ бакалавриата ОС ВПО «Лингвистика», и «Филология», желающие получить дополнительные компетенции в области программирования, математики, автоматических технологий обработки естественного языка, электронного представления гуманитарных знаний
- Выпускники бакалавриата технического или математического направления (прикладная математика, информатика, математика и др.), интересующиеся компьютерной лингвистикой, применением математических и компьютерных методов для гуманитарных наук и желающие получить дополнительные знания по лингвистике
- Выпускники других специальностей, желающие получить дополнительное образование в области компьютерной лингвистики

#### Зарубежная аудитория

- Выпускники зарубежных бакалавриатов, желающие получить специализацию в области компьютерной лингвистики и электронного представления гуманитарного знания, овладеть дополнительными компетенциями в области программного решения задач, связанных с обработкой текста, современных методов и алгоритмов автоматической обработки естественного языка и цифровой репрезентации культурного наследия

Набор студентов будет осуществляться с помощью оценки и ранжирования портфолио, существенным компонентом которого будет являться мотивационное письмо. Поскольку в основном предлагаемая программа будет вестись на английском языке, абитуриенту будет необходимо подтвердить достаточное для обучения на программе знание английского языка либо с помощью предоставления сертификата тестов, либо с помощью прохождения квалификационного устного экзамена.

Величина предполагаемого набора на программу составляет 50- 60 человек, в том числе в бюджетной части 30-50 человек.

#### **4. МЕЖДУНАРОДНЫЙ И ОТЕЧЕСТВЕННЫЙ ОПЫТ В ИЗБРАННОЙ СФЕРЕ, ОСОБЕННОСТИ ПРОЕКТА В СВЕТЕ ЭТОГО ОПЫТА**

Основные тренды развития современной лингвистики соотносятся с ключевыми векторами изменения научных подходов в целом, и, в особенности, в гуманитарных науках, а именно:

- Развитие междисциплинарных исследований на стыке наук
- Интеграция разных направлений общей научной области
- Сближение теоретической и прикладной науки: прикладные исследования рассматриваются не только как способ применения научного знания, но и как его источник
- Расширение областей научного исследования, включение новых объектов
- Распространение количественных методов в гуманитарные научные области
- Внедрение экспериментального компонента в теоретические исследования
- Использование программных средств и ресурсов для качественного проведения эксперимента и интерпретации экспериментальных данных

В лингвистике обозначенные тенденции проявляются следующим образом:

- Интенсивное развитие получают дисциплины, находящиеся на стыке лингвистики и других наук, в частности, социолингвистики и нейролингвистики, интегрирующие научные методы лингвистики и других наук, в данном случае, социологии и нейропсихологии. Таким образом, развитие в этих областях требует специализированных знаний и дополнительной подготовки
- Разные области лингвистики, такие как, с одной стороны, теория языковой структуры и способов выражения смыслов, изучения языкового разнообразия, с другой стороны, лингвистические прикладные сферы, связанные с лексикографией, преподаванием языков, а с третьей стороны, компьютерная лингвистика, берущая свое начало из задач искусственного интеллекта, имеют все больше точек пересечений. Так, например, для теории языка новым объектом исследования является изучение языковых девиаций, в частности, допускаемых изучающими тот или иной язык. С помощью методов компьютерной лингвистики строятся языковые теоретические модели, основывающиеся на данных большого объема. Для решения задач преподавания языков, подготовки словарей и лексикографического описания используются языковые корпуса, изначально предназначенные для нужд теоретической лингвистики. В итоге формируется некоторый новый фундамент – baseline – которым должен владеть современный лингвист, чем бы он ни занимался, включающий в себя, , понимание современных методов теоретической лингвистики, основных возможностей количественных алгоритмов, применяемых в компьютерной лингвистике, и типов данных, с которыми работают специалисты по прикладной лингвистике.
- Современная лингвистика использует статистические и вероятностные модели, а также во многом опирается на методы работы с большими данными, прежде всего корпусами. Методы работы с текстовой неструктурированной информацией, заимствованные из компьютерной лингвистики, применяются не только для исследования языковых структур, но и для моделирования гуманитарного знания (истории и теории литературы), опирающегося на текст. Таким образом, утверждения, делаемые в теоретической плоскости, должны в современном понимании быть подтверждены статистическими выкладками на реальных данных. Такой подход требует не только

владения комплексом умений, связанных с обработкой больших данных, но и принципиально нового понимания того, что является доказательством в лингвистике. Вне всякого сомнения, распространение алгоритмов статистической обработки текстовых данных требует владения соответственными программными пакетами и комплексами.

Основной целью ОП «Теория языка и компьютерная лингвистика» является подготовка современного лингвиста, обладающего одной из актуальных специализаций. Представление о комплексе профессиональных компетенций современного лингвиста сформировано с учетом тенденций, обозначенных выше. Эти профессиональные компетенции (знакомство с междисциплинарными областями, интеграция знаний и умений разных лингвистических направлений, владение современными статистическими методами обработки данных) формируются в соответствующих курсах, на научно-исследовательском семинаре, во время практики и других видах проектных неаудиторных работ. Мы считаем, что именно понимание современных направлений развития лингвистики и отражение их в структуре и концепции программы делает нашу программу конкурентоспособной не только на российском, но и на внешнем рынке.

## **5. ХАРАКТЕРИСТИКА СЕГМЕНТА РЫНКА ОБРАЗОВАТЕЛЬНЫХ УСЛУГ, ОСНОВНЫЕ КОНКУРЕНТЫ, СРАВНИТЕЛЬНЫЕ ПРЕИМУЩЕСТВА ОП**

### Россия

Российское образование предлагает несколько видов лингвистической подготовки следующего за бакалавриатом уровня, осуществляемых в рамках ОС «Лингвистика», «Филология» и «Фундаментальная и прикладная лингвистика». В основном вузы открывают программы по общему лингвистическому стандарту, направленному на подготовку специалистов, владеющих иностранными языками на профессиональном уровне, достаточном для преподавания и перевода, или филологическому стандарту подготовки, ориентированному на концепцию интегрированной словесности (язык и литература). Программа ОС «Фундаментальная и прикладная лингвистика» для магистратуры, кроме НИУ ВШЭ, реализуется всего в трех вузах: в МГУ, в РГГУ и в УРФУ. На сайте последнего имеются сведения о том, что прием на эту программу осуществляется, однако, содержание программы (набор дисциплин, концепция) никак не отражена. Также подробная концепция и учебный план соответствующей программы разработан в Новосибирском Государственном Университете, однако, на сайте нет сведений о том, что ВУЗ в настоящий момент осуществляет по ней подготовку.

Тем не менее, поскольку подготовка бакалавров по стандарту ФиПЛ ведется в более чем 20 вузах, для многих выпускников программ встает вопрос о продолжении профессионального образования.

Программы, предлагаемые московскими вузами (НИУ ВШЭ, РГГУ, и МГУ), в каждом из которых, имеется и свой бакалавриат по аналогичной специальности, имеют ряд существенных отличий. Если говорить в целом, то программа НИУ ВШЭ рассчитана на существенно более широкую аудиторию, как российскую, так и международную. Программа читается в основном по-английски. В программе не установлены фильтры в виде вступительных экзаменов. В рамках программы предлагается четыре образовательных траектории, выбор которой зависит в первую очередь от образовательного бекграунда студента, а также от того, что именно он хочет получить от программы. Программа МГУ, напротив, является прямым продолжением бакалавриата программы ФИПЛ, предлагает узконаправленное углубленное изучение теории языка и ориентируется в первую очередь на выпускников МГУ. Программа РГГУ занимает срединное положение, поскольку предлагает две специализации.

Преподавание на английском языке дает возможность выделить дополнительные кредиты в ОП НИУ ВШЭ для изучения еще одного языка с нуля (славянского или восточного). Эта возможность не предусматривается альтернативными магистерскими программами. Наконец, хочется отметить больший вес практической и научно-исследовательской составляющей в программе НИУ ВШЭ относительно программ конкурентов.

Приведенные выше параметры обобщены в таблице 1

Таблица 1. Сравнение трех программ магистратуры по фундаментальной и прикладной лингвистике

| Параметр       | ОП НИУ ВШЭ                            | ОП МГУ                 | ОП РГГУ                        |
|----------------|---------------------------------------|------------------------|--------------------------------|
| целевая группа | расширение за счет приема выпускников | сужение, ориентация на | сужение за счет вступительного |

|   |   |   |   |
|---|---|---|---|
| абитуриентов  | непрофильного бакалавриата. Прием на программу по портфолио, без вступительного экзамена по специальности                     | интегрированную с бакалавриатом МГУ программу, экзамен по специальности               | экзамена по специальности   |
| наличие адаптационных курсов  | <b>интегрированный адаптационный курс по компьютерной лингвистике</b>   | <b>нет</b>  | <b>нет</b>  |
| количество предлагаемых образовательных траекторий                          | <b>4 траектории</b> в рамках двух специализаций   | <b>1 специализация</b> с широким вариативным компонентом                              | <b>2 специализации</b> с вариативным компонентом                                |
| математическая подготовка   | min кредитов – <b>6</b> , max кредитов – <b>12</b>  | min кредитов - <b>6</b> , max кредитов - <b>9</b>                                     | min кредитов – <b>0</b> , max кредитов - <b>4</b>                               |
| обучение программированию   | <b>6 кредитов</b>   | <b>нет</b>  | <b>2 кредита</b>  |
| научно-исследовательский семинар как концептуальная единица программы       | есть, <b>12 кредитов</b>  | <b>нет</b>  | <b>нет</b>  |
| наличие единого блока практик, проектных и научно-исследовательских работ   | единый блок, включает в себя: нис, мастер-класс, курсовую работу, научно-исследовательскую практику, <b>всего 40 кредитов</b> | блок практики и исследовательская работа оригинального стандарта – <b>30 кредитов</b> | блок практики – <b>21 кредит</b>  |
| основной язык преподавания  | <b>английский</b>   | <b>русский</b>  | <b>русский</b>  |
| дополнительный иностранный язык (кроме языка профессиональной коммуникации) | <b>есть, 16 кредитов</b>  | <b>нет</b> , язык профессиональной коммуникации (англ) – <b>15 кредитов</b>           | <b>есть, 3 кредита</b> , язык профессиональной коммуникации - <b>8 кредитов</b> |

Таким образом, предлагаемая программа обладает рядом существенных отличий от конкурентов, обеспечивающих ей возможность более четкого и продуманного позиционирования как в среде выпускников бакалаврских программ по фундаментальной и прикладной лингвистике, так и среди прочих абитуриентов.

Ориентированность программы на широкую аудиторию абитуриентов с разным бекграундом и ее гибкая структура дает возможность ей занять особую нишу относительно программ, не связанных с лингвистикой: программа расширяет компетенции студентов, дает возможность овладеть новой специальностью или же дополнить уже имеющиеся профессиональные навыки совершенно новыми умениями. Таким образом, программа готовит разносторонних междисциплинарных специалистов и не находится в прямой конкуренции с другими нелингвистическими программами.

Это свойство программы дает возможность рассчитывать на ее продвижение на международном рынке. В частности, программа будет интересна:

Выпускникам вузов стран СНГ, заинтересованных в получении дополнительных специализаций по социолингвистике и типологии, компьютерной лингвистике, которые очень редко преподаются в значительном объеме в бакалаврских программах и магистратурах стран содружества. Кроме того, безусловно привлекательной является перспектива обучения современным технологиям и практикам в области русского как неродного. Кроме того, привлекательным моментом программы является возможность стажировок в университетах Франции, Норвегии, Великобритании, Финляндии.

В странах Восточной Европы мало представлена специализация по digital humanities (электронному представлению гуманитарного знания), которое будет реализовано в программе в качестве одной из траекторий компьютерной лингвистики.

В целом для стран Европы и США привлекательной является компонент полевых практик, предлагаемый специализацией по теории языка, поскольку языковое разнообразие России дает возможность сбора уникальных данных. Мы также рассчитываем на интерес к программе выпускников славянских отделений (славистов), желающих продолжить углубленное изучение не только практических, но и теоретических аспектов русского языка. Несмотря на то, что на западе открыто множество магистратур по компьютерной лингвистике, конкурентным преимуществом нашей программы является отсутствие пререквизитов, активная проектная работа, интеграция с теорией языка и корпусной лингвистикой.

## **6. «ПОРТРЕТ ВЫПУСКНИКА» ОП, АНАЛИЗ ВОСТРЕБОВАННОСТИ И ВОЗМОЖНОСТЕЙ ТРУДОУСТРОЙСТВА ВЫПУСКНИКОВ**

Выпускник магистратуры по компьютерной лингвистике факультета гуманитарных наук НИУ ВШЭ должен:

- иметь хороший теоретический бэкграунд в теоретической лингвистике, математических методах обработки данных;
- знать, как устроена современная электронная языковая среда, уметь увидеть в ней новые проектные задачи; уметь их формулировать и формализовать;
- уметь работать с данными: понимать, какие именно данные нужны для решения поставленной задачи, как взаимодействуют лингвистические и иные данные, откуда их можно получить;
- понимать, какие языковые структуры лежат за той или иной локальной задачей, каковы лингвистические и экстралингвистические свойства текстов, используемых для решения задач;
- знать, какие лингвистические технологии следует применять в зависимости от условий задачи, понимать, на какие лингвистические данные и ресурсы может опираться технология;
- уметь программировать прототипы и модели решений; подготавливать необходимые лингвистические ресурсы;
- уметь проектировать цепочку процесса обработки языковых данных и интерпретировать результаты автоматической обработки лингвистических данных;
- оценивать трудоемкость разных решений и пороги допустимых ошибок и выбирать оптимальную методологию тестирования.
- понимать основные подходы к цифровому моделированию гуманитарных данных в области истории и литературы, основные тенденции развития этой области знаний

Возможности трудоустройства выпускника специализации по компьютерной лингвистике, кроме собственно академической карьеры в России или за рубежом включают в себя работу в качестве компьютерного лингвиста в коммерческих компаниях. Среди российских работодателей интерес к выпускникам программы проявляют такие «технологические гиганты», как компании Яндекс, АБВУУ, Mail.ru и др. В то же время, компьютерные лингвисты востребованы и в стартапах, небольших компаний, чей бизнес связан с наукоемкими и информационными технологиями. Выпускники, выбравшие трек по digital humanities, смогут, кроме вышеперечисленного, работать в

архивах, библиотеках, музеях, в современных городских проектах, связанных с обработкой текстовых данных и сохранения культурного наследия.

## **7. СТРУКТУРА УЧЕБНОГО ПЛАНА, КАК ОНА ОБЕСПЕЧИВАЕТ РЕАЛИЗАЦИЮ ЦЕЛЕЙ И ЗАДАЧ ОП**

Программа имеет следующую структуру учебного плана:

- I. Адаптационный интегрированный курс по компьютерной лингвистике, предназначенный для студентов, не имевших подготовки по математике, программированию и базовым понятиям компьютерной лингвистики в бакалавриате ( не входит в общее число кредитов по программе)

### **II. ОБЩАЯ БАЗОВАЯ ЧАСТЬ**

- В общей базовой части представлены три ключевых дисциплины, обеспечивающих высокий уровень лингвистической подготовки по обеим специализациям. Задачей этих дисциплин состоит в том, чтобы познакомить студентов с современными подходами к анализу языковых структур и научить его пользоваться современными методами и инструментами для анализа данных. Эти задачи обеспечиваются курсами
  - ✓ **Функциональные и когнитивные модели в лингвистике / Functional and cognitive models in linguistics**
  - ✓ **Формальные модели в лингвистике / Formal models in linguistics**
  - ✓ **Анализ языковых данных: квантитативные методы и визуализация / Linguistic data: quantitative analysis and visualisation**

### **ОБЩАЯ ВАРИАТИВНАЯ ЧАСТЬ**

- Общий электив (1 из 2) расширяет кругозор студентов в области возможных исследований в лингвистике и предлагает два междисциплинарных курса:
  - ✓ **Экспериментальная лингвистика / Experimental Linguistics**
  - ✓ **Антропология / Cultural Antropology**

### **III. ДИСЦИПЛИНЫ ПРОГРАММЫ**

- Пул обязательных дисциплин по компьютерной лингвистике включает в себя собственно курс по **компьютерной лингвистике**, а также два необходимых сопровождающих и тематически скоординированных с ним курса: курс **математических оснований компьютерной лингвистики** и курс программирования на языке **Python**. Кроме того, обязательным является курс по **машинному обучению**, исключительно важный с точки зрения профессионального роста и формирования компетенций компьютерного лингвиста.
- Дисциплины по выбору первого года обучения дают возможность получить дополнительную концентрацию в области электронных гуманитарных технологий. Это обеспечивается элективом в формате два из трех, в котором два курса непосредственно относятся в области digital humanities (**теоретический и практический курс**), а третий – курс **машинного перевода**, отдельной области компьютерной лингвистики
- Дисциплины по выбору второго года обучения завершают цикл специализации, концентрируясь на наиболее актуальных подходах и алгоритмах современной компьютерной лингвистики:
  - ✓ **Социальные сети / Social networks**
  - ✓ **Онтологии / Ontologies**
  - ✓ **Компьютерная семантика / Computational semantics**

### **IV. ПРАКТИКИ И НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА**

- Научно-исследовательская работа формируется из трех компонентов: проектного **научно-исследовательского семинара**, в рамках которого студенты выполняют проекты в группах, индивидуальной научно-исследовательской работы, в результате которой студент пишет и защищает **курсовую работу**, и **мастер-классов**, на которых студенты знакомятся с профессионалами и экспертами в изучаемых областях.
- Практики у студентов связываются с выбранными специализациями и треками. Для специализации по теории языка практика может проходить в центре РКИ НИУ ВШЭ

или же осуществляться в формате полевой работы или экспедиций. Студенты, специализирующиеся в области компьютерной лингвистики проходят практику на предприятиях.

#### V. ИТОГОВАЯ ГОСУДАРСТВЕННАЯ АТТЕСТАЦИЯ

- Итоговая государственная аттестация проходит в формате защиты дипломной работы с привлечением внешних оппонентов, в том числе (по возможности) из зарубежных вузов или институтов

#### 8. КОНЦЕПЦИЯ НАУЧНО-ИССЛЕДОВАТЕЛЬСКОГО (ПРОЕКТНОГО) СЕМИНАРА

Научно-исследовательский семинар рассматривается как ключевой образовательный компонент магистерской программы. Целью семинара является подготовка магистров, специализирующихся по теории языка и компьютерной лингвистике к самостоятельной научно-исследовательской и проектно-организационной работе. Научный семинар включает в себя два вида работ: групповую проектную работу и индивидуальную исследовательскую работу, таким образом, семинар может быть разделен на проектную и исследовательскую части. Проектная часть НИСа проходит в 1-3 модуле первого года обучения и в 1-2 модулях второго года обучения. Общая идея проектного семинара состоит в реализации группой студентов проекта (продукта), представляющего научный или общественный интерес, который соответствовал бы научным векторам развития школы и направления в целом. Студенты самостоятельно проходят все стадии работы над проектом, начиная от пилотного исследования, формирования технического задания, реализации поставленных задач и финального оформления результатов. Для студентов, специализирующихся в области теории языка, возможные проекты включают в себя документацию малых языков, создание лингвистических корпусов, словарные проекты, описание русской грамматики, тренажеры для обучения языку. Студенты, осваивающие специализацию по компьютерной лингвистике, могут выбрать в качестве задачи проектной деятельности разработку модуля автоматического анализа естественного языка в рамках определенных задач, разработку лингвистических ресурсов и программ, дигитальное представление культурного наследия. **Проектный** семинар является основной точкой плодотворного взаимодействия студентов разных специализаций и индивидуальных треков. Планирование работы над проектом выглядит следующим образом:

##### Первый год

- 1 модуль: обзор возможных задач, выбор задачи для проекта
- 2 модуль: пилотные исследования, формирование технического задания
- 3 модуль: реализация пилотной версии проекта
- 4 модуль: завершение работы над пилотной версией проекта в фоновом режиме

##### Второй год

- 1-2 модуль: финальная часть работы над проектом, подготовка презентации результатов в виде статьи или открытого кода

Рабочие проектные группы работают вместе с кураторами, которыми могут быть преподаватели программы или внешние приглашенные эксперты. Подробный дневник работы над проектом, включающий в себя релевантные материалы, результаты обсуждений, данные, ведется участниками проекта онлайн в системе семейства lms. Одной из задач проектного семинара является обучение грамотному планированию и документированию наукоемкой проектной деятельности, а именно той, которой свойственен экспериментальный, заранее неизвестный сценарий развития

На семинарах рабочие группы должны показать остальным участникам результаты своего труда. Формат выступления может носить характер сообщения-реферата по релевантной литературе, брейнсторма по полученным результатам и их анализу и краудсорсинга с привлечением всех участников семинара.

Общий **исследовательский** семинар проходит в четвертом модуле первого года обучения и в третьем-четвертом модулях второго года обучения. Предметом коллективного обсуждения на семинаре являются индивидуальные научные исследования студентов – курсовые на первом году и дипломы на втором. Оценка студента по НИС формируется из оценки текстов, отражающих разные этапы работы над магистерской диссертацией, и презентаций в ходе публичных выступлений на НИС

**9. ХАРАКТЕРИСТИКА КАДРОВОГО ПОТЕНЦИАЛА, СОБСТВЕННОГО И ПРИВЛЕКАЕМОГО СО СТОРОНЫ, А ТАКЖЕ ИМЕЮЩИХСЯ И ТРЕБУЕМЫХ ДЛЯ РЕАЛИЗАЦИИ ОП РЕСУРСОВ.**

В реализации программы принимают участие преподаватели направления фундаментальной и прикладной лингвистики Общего Гуманитарного Факультета НИУ ВШЭ, а также преподаватели, работающие в рамках программ международного найма и приглашенные преподаватели. Академические статусы и научные интересы участников программы представлены в списке ниже:

- Академический руководитель программы «Теория языка и компьютерная лингвистика», кандидат филологических наук, доцент А.А. Бонч-Осмоловская (компьютерная лингвистика, корпусная лингвистика, digital humanities)
- Доктор филологических наук, профессор Е.В. Рахилина (теоретическая семантика, лексическая типология, теоретическая грамматика русского языка, корпусная лингвистика)
- PhD Stanford University, профессор Т.В. Никитина (грамматическая типология, синтаксическая типология, теоретический синтаксис, языки Африки)
- Кандидат филологических наук, PhD University of South California, доцент В.Ю. Апресян (русистика, теоретическая семантика, теоретическая лексикография)
- Кандидат филологических наук, профессор О.Н. Ляшевская (теоретическая грамматика русского языка, корпусная лингвистика)
- Кандидат филологических наук, доцент А.Б. Летучий (грамматическая типология, теоретическая грамматика русского языка, корпусная лингвистика)
- Кандидат филологических наук, доцент Н.Р. Добрушина (грамматическая типология, теоретическая грамматика русского языка, корпусная лингвистика, языки Кавказа, социолингвистика)
- Кандидат филологических наук, доцент С.Ю. Толдова (грамматическая типология, синтаксическая типология, малые языки, языки Кавказа, уральские языки, автоматическая обработка текста)
- Кандидат филологических наук, доцент Ю.А. Ландер (грамматическая типология, синтаксическая типология, малые языки, языки Кавказа)
- Кандидат филологических наук, преподаватель Т.А. Архангельский (корпусная лингвистика, армянский язык, греческий язык, малые языки, автоматическая обработка текста, программирование)
- Кандидат физико-математических наук, доцент кафедры высшей математики И.В.Щуров (математика, математические основания компьютерной лингвистики)
- Кандидат филологических наук, доцент Б.В. Орехов (корпусная лингвистика, digital humanities, программирование, компьютерная лингвистика)
- Преподаватель Г.А. Мороз (фонетическая типология, малые языки, языки Кавказа)
- Преподаватель Д.А. Алексеевский (программирование, компьютерная лингвистика)
- Кандидат филологических наук, Phd Utrecht University, доцент Н.А. Слюсарь, tenure track (теория языка, морфология, синтаксис, психолингвистика)
- PhD: Massachusetts Institute of Technology, А.В.Подобряев, postdoc (теория языка)
- Phd University of Southern California, Associate professor of Dartmouth College Michael Gronas (Russian as foreign language, Digital Humanities)
- Phd Tartu University, professor of Tartu University Roman Leibov (Digital Humanities)