

Лаптева Наталья Евгеньевна
кандидат педагогических наук,
доцент, ректор
УО «Минский государственный
лингвистический университет»
г. Минск, Беларусь

Natalia Lapteva
PhD in Pedagogy, Rector
Minsk State Linguistic University
Minsk, Belarus
rector@mslu.by

Бетеня Елена Петровна
кандидат филологических наук,
доцент, проректор по учебной работе
УО «Минский государственный
лингвистический университет»
г. Минск, Беларусь

Elena Betenya
PhD in Philology
Vice-Rector for Academic Affairs
Minsk State Linguistic University
Minsk, Belarus
elbet@mail.ru

Шевцова Валентина Анатольевна
кандидат филологических наук
доцент, декан факультета немецкого языка
УО «Минский государственный
лингвистический университет»
г. Минск, Беларусь

Valentina Shevtsova
PhD in Philology
Dean of the German Faculty
Minsk State Linguistic University
Minsk, Belarus
schewzowa_w@mail.ru

Долматова Екатерина Дмитриевна
кандидат филологических наук,
доцент, заместитель декана
факультета английского языка
УО «Минский государственный
лингвистический университет»
г. Минск, Беларусь

Ekaterina Dolmatova
PhD in Philology
Vice-Dean of the English Faculty
Minsk State Linguistic University
Minsk, Belarus
engldek@gmail.com

ОПЫТ ПРОЕКТИРОВАНИЯ НОВОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ
БАКАЛАВРИАТА 6-05-0231-04
«ЦИФРОВАЯ ЛИНГВИСТИКА (С УКАЗАНИЕМ ЯЗЫКА)»
В МИНСКОМ ГОСУДАРСТВЕННОМ ЛИНГВИСТИЧЕСКОМ
УНИВЕРСИТЕТЕ

В статье представлен опыт проектирования образовательной программы бакалавриата 6-05-0231-04 «Цифровая лингвистика (с указанием языка)» в учреждении образования «Минский государственный лингвистический университет». Программа направлена

на подготовку лингвистов нового поколения, способных сочетать комплексные *лингвистические* компетенции с *техническими* компетенциями для эффективного лингвистического сопровождения разработки инновационных продуктов в сфере применения искусственного интеллекта, анализа и обработки больших данных, в том числе на иностранном языке. Такой междисциплинарный подход обусловил необходимость присвоения двух квалификаций по окончании образовательной программы – «Лингвист. Специалист по информационным технологиям». В статье рассматриваются причины открытия новой специальности бакалавриата «Цифровая лингвистика (с указанием языка)» в Республике Беларусь и этапы ее проектирования, в частности, процесс формирования учебного плана и матрицы компетенций.

Ключевые слова: бакалавриат; компетенции; образовательная программа; проек-тирование; цифровая лингвистика; цифровизация.

CASE STUDY OF DEVELOPING THE NEW BACHELOR'S PROGRAMME 6-05-0231-04 "DIGITAL LINGUISTICS (WITH THE INDICATION OF THE FOREIGN LANGUAGE)" IN MINSK STATE LINGUISTIC UNIVERSITY

The article presents the experience of developing the new Bachelor's degree programme 6-05-0231-04 "Digital Linguistics (with the indication of the foreign language)" in Minsk State Linguistic University. The programme is aimed at training new generation of linguists capable of combining complex linguistic competencies with technical competencies for effective linguistic support for innovative product development in the field of Artificial Intelligence, Big Data processing and analysis. This interdisciplinary approach has necessitated the awarding of two qualifications at the end of the educational programme – "Linguist. Information Technology Specialist". The article discusses the reasons for launching a new bachelor's programme "Digital Linguistics (with the indication of the foreign language)" in the Republic of Belarus and the stages of its development, in particular, the principles and process of curriculum development and competency matrix creation.

Key words: Bachelor's degree; competences; educational programme; design; digital linguistics; digitalization.

Цифровая трансформация является общепризнанным стратегическим направлением развития современного общества в целом и системы образования в частности. Информационно-коммуникационные технологии (ИКТ) как основной инструмент цифровизации активно модернизируются и в значительной степени трансформируются в технологии искусственного интеллекта, успешное внедрение и применение которых требует наличия подготовленных для этого кадров. Задача подготовки, точнее, формирования, кадров нового – цифрового – поколения может решаться и решается различными путями: точечно и оперативно путем реализации программ по обмену знаниями на уровне Совета Министров [1, с. 9] и организации обучающих курсов для школьников и молодежи на базе технопарков (например, Национальный детский технопарк и образовательный центр при Парке высоких технологий), а также более масштабно, но и более «времязатратно» путем обновления существующих и внедрения новых образовательных программ в университетах.

Таким образом, в условиях цифровой трансформации для системы высшего образования в качестве одной из приоритетных задач выступает разработка перспективных образовательных программ, «стимулирующих критическое мышление, творчество и эмоциональное восприятие, а также умение быстро осваивать цифровые технологии. <...> В ближайшие годы для цифровой трансформации и преобразования экономики Беларуси в полноправного участника мирового рынка потребуются квалифицированные специалисты в области цифровых технологий, точных наук и традиционных сфер, такие, как «цифровой инженер-механик» и «аналитик данных бизнес-операций» [1, с. 6].

Как правило, продвижение цифровой трансформации и разработку «цифровизированных» образовательных программ связывают с развитием так называемых технических или естественнонаучных сфер знания (математики, информатики и программирования и т. п.). Однако сегодня данные процессы в значительной степени затрагивают и лингвистику. По мнению профессора В. Д. Соловьева, современная лингвистика, начиная со второй половины 20-го века, «стремительно эволюционирует в направлении естественных и математических дисциплин» [2]. Он апеллирует к выдающемуся лингвисту Н. Хомскому, который показал, что «несмотря на все разнообразие языков, в их основе лежат фундаментальные законы, строго выполняющиеся во всех без исключения языках. Эти законы в той же мере универсальны, как и законы естественных наук – законы механики, гравитации в физике или дарвиновские законы эволюции в биологии. Это сблизило лингвистику с естественными науками» [Цит. по: 2]. Благодаря формализации лингвистических теорий, язык может изучаться математическими методами, в том числе с использованием технологий программирования. В настоящее время лингвисты оперируют огромным объемом языковых данных, ставших доступными с помощью различных поисковых систем, а для анализа и обработки используются компьютерные программы статистического анализа. Поэтому нельзя не согласиться с В. Д. Соловьевым в том, что «для лингвистики, как и для физики очень важным является научное направление “Big Data” (Большие данные), разрабатываемое в компьютерных науках» [2]. Эти и другие изменения приводят не только к расширению границ современной лингвистики, но и к увеличению числа людей, которым необходимо такое «расширенное» лингвистическое образование.

Представляется, что профессия «цифровой лингвист» и образовательная программа по цифровой лингвистике правомерно могут занять свое место в данном ряду инновационных направлений деятельности. Подтверждением этому служат аналитические материалы авторитетных информационных источников.

Так, по данным исследования информационного канала «РБК Тренды», цифровая лингвистика входит в список профессий будущего [3].

«Атлас новых профессий 3.0.», изданный с дополнениями в 2022 г. в Москве под редакцией Д. Варламовой, Д. Судакова, включает профессию «цифровой лингвист» в число самых востребованных в ближайшие 30 лет

[4, с. 34] и дает ей следующее определение: «Цифровой лингвист – это профессионал, разрабатывающий лингвистические системы семантического перевода (перевода с учетом контекста и смысла), системы обработки текстовой информации (в том числе семантический поиск в интернете) и новые интерфейсы общения между человеком и компьютером на естественных языках» [4, с. 191].

Обзор тематики докладов, представленных на авторитетных международных конференциях по компьютерной лингвистике и интеллектуальным технологиям (Диалог [5], СПЕКОМ [6]) свидетельствует о том, что современная работа с лингвистическими (языковыми) данными может подразумевать разную степень цифровизации. Спектр обсуждаемых на конференциях актуальных вопросов включает как задачи NLP, или обработки естественного языка (область, где требуются знания по математике, программированию и машинному обучению), так и задачи традиционного языкознания (морфологии, синтаксиса, семантики), решаемые при помощи современных методов исследования (корпусного анализа, статистических методов анализа данных). На наш взгляд, анализ материалов докладов красноречиво подтверждает тот факт, что цифровая лингвистика является одним из локомотивов развития искусственного интеллекта и цифровизации современного общества.

Анализ образовательных программ в Российской Федерации показал, что цифровая лингвистика стала уже «образовательной реальностью» на уровне магистерских программ [7]. За последние 5 лет, наряду со ставшими уже традиционными программами магистратуры по компьютерной и прикладной лингвистике (СПбГУ, ВШЭ, РГГУ), появились новые междисциплинарные магистерские программы, которые позволяют студенту сформировать ключевые компетенции в области цифровой лингвистики: 45.03.01 «Прикладная цифровая филология» (РУДН); 45.03.03 «Фундаментальная и прикладная лингвистика. Лингвистика в информационно-коммуникационной цифровой среде» (ДВФУ); 45.03.04 «Интеллектуальные системы в гуманитарной сфере. Цифровая лингвистика» (ЮФУ); 45.04.02 «Цифровая лингвистика и локализация» (НИТУ МИСИС); 45.04.02 «Цифровые технологии в лингвистических проектах» (СПбГУ); 45.04.02 «Цифровая лингвистика» (РГПУ им. Герцена); 45.04.04_02 «Цифровая лингвистика» (Гуманитарный институт СПбПУ); 45.04.04 «Цифровая гуманитаристика» (УрФУ); 09.04.01 «Обработка естественного языка» (НИТУ МИСИС) и др. [7].

Специфика вышеуказанных магистерских программ заключается в их междисциплинарности, а именно: подготовка специалистов, имеющих профессиональные компетенции в области лингвистики, прикладной математики и информатики. Принципиальным отличием данных программ является их практикоориентированность и реальный (а не декларируемый) учет запросов конкретного сегмента рынка труда в том или ином регионе. Так, например, в НИТУ МИСИС в магистерской программе 09.04.01 «Обработка естественного языка» акцент сделан на компьютерной обработке текста и речи путем применения алгоритмов искусственного интеллекта и машинного

обучения [8], а в программе 45.04.02 «Цифровая лингвистика и локализация» в фокусе внимания находится работа с системами синтеза и распознавания речи, машинного перевода, а также с различными типами нейросетей [9]. В программе магистратуры 45.04.04_02 «Цифровая лингвистика» в Гуманитарном институте СПбПУ объединены два профессиональных ядра – прикладная лингвистика и информационные технологии, которые позволяют студенту сформировать ключевые компетенции в рамках двух основных образовательных треков – «Машинного перевода» и «Лингвистических технологий в бизнесе» [10]. Последний трек предполагает изучение технологии автоматизированной обработки данных, систем предиктивной аналитики данных и др.

В Республике Беларусь существующие образовательные программы, как правило, были направлены на подготовку специалистов либо сугубо технического профиля (как правило, инженеров-программистов), либо специалистов-гуманитариев, обладающих базовыми знаниями информационно-коммуникационных технологий. Однако, учитывая запросы современного рынка труда, а также учитывая перспективы его развития и прогнозы по профессиям будущего, Минский государственный лингвистический университет (МГЛУ) осуществлял проектирование инновационной специальности бакалавриата «Цифровая лингвистика (с указанием языка)», которое завершил в 2023 году – юбилейном 75-ом году деятельности университета. Основные этапы разработки новой образовательной программы включали: 1) определение предметной области и методологической основы, на которой базируется новая специальность, 2) определение перечня профессиональных компетенций специалиста и формулирование на этой основе присваиваемой квалификации, 3) проектирование учебного плана и матрицы компетенций. Далее представим краткое описание проблематики каждого из этапов.

1. Предметная область и методологическая основа специальности «Цифровая лингвистика».

Несмотря на тот факт, что существует дефиниция понятия «цифровой лингвист», о которой мы упоминали ранее (см. «Атлас новых профессий» [4]), определение предметной области цифровой лингвистики представляло определенную трудность, поскольку она являет собой новую, в значительной степени зарождающуюся, междисциплинарную область знаний, которая находится на стыке лингвистических наук и информатики.

В научной литературе пока не существует единого подхода к определению понятия «цифровая лингвистика». Имеют место факты отождествления как синонимичных понятий «цифровая лингвистика», «компьютерная лингвистика», «прикладная лингвистика», «математическая лингвистика», что обусловлено тем, что данные термины, в конечном счете, служат для описания области, которая объединяет лингвистику и информационные технологии для анализа и обработки естественного языка.

Однако «цифровая лингвистика» и «компьютерная лингвистика» не являются тождественными понятиями. Понятие «цифровая лингвистика», которым широко оперируют в настоящее время, является, на наш взгляд, более широким и базируется на следующих трех компетентностных «китах» (рисунки):



Рис. Предметная область цифровой лингвистики

Во-первых, на комплексных *лингвистических знаниях*, таких, как профессиональное владение иностранным языком, глубоких теоретических знаниях о структуре и механизмах функционирования естественного языка и способах его формализации, во-вторых, на *знаниях в области программирования и искусственного интеллекта* и, в-третьих, на *знаниях технологий и инструментов обработки больших данных*.

Вместе с тем уже устоявшаяся в качестве области научных знаний прикладная и математическая лингвистика служит методологической основой для новой специальности «Цифровая лингвистика (с указанием языка)».

2. Профессиональные компетенции и квалификация специалиста.

При определении перечня компетенций были изучены потребности рынка труда, а также опыт распределения выпускников МГЛУ, обучавшихся по специализации (профилизации) «Компьютерная лингвистика». Кроме того, для определения востребованности образовательной программы были проведены консультации со специалистами Белорусского института стратегических исследований (БИСИ), представителями компаний IT-сектора и промышленных предприятий, а также с другими организациями-заказчиками кадров, от которых было получено около 30 заявок на подготовку «цифровых лингвистов» и их востребованность в 5-летний период, т. е. после 2028 года. В результате был сформирован перечень профессиональных компетенций, которые далее послужили основой и целевыми ориентирами для разработки учебного плана специальности. Итак, сфера компетенций специалиста-выпускника образовательной программы по цифровой лингвистике включает:

1) осуществление межкультурной коммуникации на иностранном языке с учетом страновой и региональной специфики;

2) создание, перевод и локализация технической документации на разрабатываемые продукты с учетом лингвокультурных особенностей регионов и рынков;

3) анализ больших объемов структурированных и неструктурированных данных (Big Data) с использованием цифровых инструментов;

4) владение языками программирования (Python, SQL) и способами создания систем искусственного интеллекта для решения профессиональных задач;

5) разработка речевых технологий: преобразование звучащей речи в текст и текста в звучащую речь, голосовая биометрия (идентификация личности по голосу) и др.

Поскольку образовательная программа направлена на формирование компетенций в трех сферах (лингвистика, иностранный язык и информационные технологии), целесообразным представляется присвоение двух квалификаций: лингвист, специалист по информационным технологиям.

Таким образом, цель открытия новой специальности «Цифровая лингвистика (с указанием языка)» – подготовка специалистов на основе междисциплинарного подхода, способных сочетать комплексные лингвистические компетенции с техническими компетенциями для лингвистического сопровождения разработки инновационных продуктов в сфере применения искусственного интеллекта, анализа и обработки больших данных, в том числе на иностранном языке. Иными словами, специальность направлена, во-первых, на изучение теоретических основ и практики создания формальных моделей естественного языка, во-вторых, на изучение методов использования и разработки систем интеллектуальной обработки текстовой и речевой информации, а также инструментов интеллектуальной обработки больших данных (Big Data).

3. Учебный план и матрица компетенций.

На основе разработанного перечня профессиональных компетенций, которые описаны в предыдущем разделе, были спроектированы блоки и модули учебных дисциплин, а также матрица компетенций, на формирование которых направлено освоение заявленных учебных дисциплин. При этом учитывался российский опыт подготовки цифровых (компьютерных) лингвистов (научные школы МФТИ, ВШЭ, СПбГУ, МГУ) и достижения собственных авторитетных научно-педагогических школ – школы компьютерной и корпусной лингвистики, а также Минской фонетической школы. В результате была разработана инновационная образовательная программа междисциплинарного характера «Цифровая лингвистика (с указанием языка)», которая предполагает освоение следующего спектра учебных курсов:

– модуль социально-гуманитарных дисциплин (философия искусственного интеллекта, логика, лингвокибернетика и др.), нацеленный на формирование критического и системного мышления в контексте межотраслевого общения, в том числе на иностранных языках;

– модули учебных дисциплин, направленные на овладение иностранным языком на профессиональном уровне (практический курс иностранного языка, функциональная грамматика, практикум речевого общения, техническое письмо, специальный перевод и др.);

– модули лингвистических дисциплин (введение в языкознание, структурная лингвистика английского языка, когнитивная лингвистика, корпусная лингвистика, формализация семантики и др.), дающих глубокое понимание закономерностей функционирования языка и лингвистических процессов и закладывающих теоретические основы и методологическую базу моделирования языка;

– модули прикладных дисциплин, направленных на формирование компетенций в области информационных технологий и программирования (математика и информатика, алгоритмизация и основы программирования, базы данных и SQL, профессионально-ориентированное программирование и др.);

– модули учебных дисциплин, направленные на обработку данных (прикладная статистика, основы веб-аналитики, обработка больших данных).

В настоящее время в реализуемой образовательной программе заложена профилизация «Речевые технологии». Однако в будущем набор профилизаций может быть расширен в целях обеспечения индивидуализации профессионального развития студента с учетом места будущего трудоустройства и/или актуальных запросов рынка труда. При этом, как нам кажется, более глубокая профилизация должна развиваться в рамках магистерских программ.

Таким образом, в основе разработки учебного плана лежит модульный подход, учитывающий принципы вариативности, метапредметности и индивидуализации. В виду этого, с одной стороны, в учебный план включены учебные дисциплины (модули), целенаправленно формирующие заданные компетенции, с другой стороны, данный план обеспечивает возможность оперативно реагировать на конкретные запросы рынка труда посредством дисциплин профилизации и дисциплин по выбору студента. Отметим, что при разработке учебного плана также принимался во внимание тот факт, что не все студенты, обучающиеся по данной программе, будут иметь глубокие знания математики за курс средней школы. Этот факт будет учтен и далее – при разработке содержания учебных программ соответствующих учебных дисциплин.

Следует особо подчеркнуть, что специальность «Цифровая лингвистика (с указанием языка)» не ставит своей целью обеспечить овладение выпускниками навыками программирования на высоком профессиональном уровне и в том объеме, в котором ими обладают успешные выпускники собственно ИТ-специальностей, для которых эти компетенции основные. Программирование для цифрового лингвиста является инструментом для решения профессиональных задач, связанных с лингвистическим сопровождением инновационных продуктов в сфере искусственного интеллекта, например, таких,

как: повышение производительности алгоритмов NLP, использование (настройка) систем искусственного интеллекта под выбранные человеком задачи, создание алгоритмов систематизации текстов по разным признакам, сопровождение процесса программирования генераторов текстов по заданным параметрам, разработка моделей/механизмов оценки качества функционирования систем и методов интеллектуальной обработки информации, а также оценка и совершенствование функционирования речевых/голосовых ассистентов, чат-ботов, виртуальных помощников и др.

Завершая рефлексию опыта проектирования новой специальности, нельзя не отметить тот факт, что каждый из этапов представлял определенный вызов для разработчиков и в целом ряде случаев требовал преодоления традиционных подходов и трактовок этих подходов, что обусловлено спецификой проектируемой специальности, ее междисциплинарным и интегративным характером.

Безусловно, процесс реализации данной образовательной программы будет также сопряжен с вызовами и решением вопросов, которые поставит практика преподавания – начиная от соотнесения результатов обучения с формируемыми компетенциями и завершая анализом уровней овладения заявленным набором компетенций и валидностью данного набора с учетом скорости развития технологий искусственного интеллекта. Но это естественный процесс в ходе внедрения новых специальностей в систему высшего образования.

Таким образом, новая специальность «Цифровая лингвистика (с указанием языка)», спроектированная и открытая МГЛУ, внесет значительный вклад в формирование лингвиста нового поколения – специалиста, который профессионально владеет иностранным языком и как минимум одним языком программирования, владеет цифровыми компетенциями, знаниями закономерностей функционирования естественного языка и глубоким пониманием его роли и места в эпоху искусственного интеллекта.

ЛИТЕРАТУРА

1. Цифровая трансформация. Основные понятия и терминология : сборник статей / Национальная академия наук Беларуси, Объединенный институт проблем информатики ; А. В. Тузиков [и др.]. Минск : Беларуская навука, 2020. 266 с.

2. Соловьев В. Д. Три этапа цифровизации лингвистики [Электронный ресурс] // Казанский федеральный университет. 2021. URL: <https://krfu.ru/philology-culture/39tri-etapa-cifrovizacii-lingvistiki39-404921.html> (дата обращения: 01.07.2024).

3. 150 профессий будущего [Электронный ресурс] // РБК Тренды. 2024. URL: <https://trends.rbc.ru/trends/education/5d6e48529a794777002717b> (дата обращения: 01.07.2024).

4. *Атлас новых профессий 3. 0.* (дополненное издание) / под ред. Д. Варламовой, Д. Судакова. М. : Альпина PRO, 2022. 472 с.

5. *Коллекция сборников «Компьютерная лингвистика и интеллектуальные технологии» [Электронный ресурс]* // Dialogue : [сайт]. URL: <https://www.dialog-21.ru/digest/> (дата обращения: 01.07.2024).

6. СПЕКОМ [Электронный ресурс] : [сайт]. URL: <https://specom.nw.ru/2024/> (дата обращения: 01.07.2024).

7. Цифровой лингвист: в каких вузах учат, куда поступать в России [Электронный ресурс] // Vuzopedia : [сайт]. URL: <https://vuzopedia.ru/professii/2221/vuzu> (дата обращения: 01.07.2024).

8. Обработка естественного языка [Электронный ресурс] // МИСИС : [сайт]. URL: <https://misis.ru/applicants/admission/magistracy/faculties/informatikaivtmag/obestyaz> (дата обращения: 01.07.2024).

9. Цифровая лингвистика и локализация [Электронный ресурс] // МИСИС : [сайт]. URL: <https://misis.ru/applicants/admission/magistracy/faculties/lingmg/digling/> (дата обращения: 01.07.2024).

10. 45.04.04_02 «Цифровая лингвистика» [Электронный ресурс] // Политех : [сайт]. URL: https://hum.spbstu.ru/edu/45.04.04/45.04.04_02/ (дата обращения: 01.07.2024)

Благодарности: авторы статьи выражают благодарность членам рабочей группы по проектированию образовательной программы, а также коллективу Минского государственного лингвистического университета, поддерживающему и реализующему новые идеи содержания программы подготовки цифровых лингвистов.