https://doi.org/10.52944/PORT.2023.52.1.005
Научная статья

Формирование коммуникативной компетенции студентов-документоведов с помощью методов корпусных исследований

И. Г. Гончаренко ✉
Российский государственный профессионально-педагогический университет, Екатеринбург, Российская Федерация
✉ ilya.goncharenko@rsvpu.ru

Аннотация

Введение. Использование педагогических технологий, предполагающих обучение через исследование, получает все более широкое распространение в практике подготовки специалистов для системы среднего профессионального образования (СПО). В работе показана возможность применения одной из таких технологий, связанных с использованием корпусных исследований для формирования профессионально значимых компетенций.

Цель. Представить описание и продемонстрировать преимущества технологии использования корпусных данных в профессиональной подготовке студентов-документоведов.

Методы. При проведении исследования был выполнен анализ нормативно-правовой документации, а также учебно-методической литературы. В качестве средств обучения были использованы инструменты корпусной лингвистики, а именно: дистрибутивный анализ при характеристике сочетаемостных возможностей слов, количественный анализ при определении частотности слов и словосочетаний в целом и в конкретном типе текстов в частности. Кроме того, был произведен контент-анализ выполненных студентами исследовательских работ.

Результаты. Приведены доказательства эффективности использования технологии корпусных исследований при формировании у студентов не только языковых, но и исследовательских компетенций, что позволяет расширить диапазон формируемых у студентов умений и знаний, стимулировать внутридеятельностную мотивацию учения.

Научная новизна. В работе впервые описывается технология применения методов корпусной лингвистики в практике преподавания документной лингвистики и, шире, в практике формирования коммуникативной компетенции у студентов-документоведов — будущих преподавателей системы СПО.

Практическая значимость. Результаты работы внедрены в практику подготовки документоведов — преподавателей системы СПО, изложенные в работе идеи и практические рекомендации могут быть использованы при проведении занятий по документной лингвистике.

Ключевые слова: коммуникативные компетенции, профессиональная коммуникация, профессионально-педагогическое образование, исследовательские компетенции, документная лингвистика, корпусная лингвистика, преподаватель СПО

Для цитирования: Гончаренко И. Г. Формирование коммуникативной компетенции студентов-документоведов с помощью методов корпусных исследований // Профессиональное образование и рынок труда. 2023. Т. 11, № 1. С. 61–76. https://doi.org/10.52944/PORT.2023.52.1.005

© И. Г. Гончаренко, 2023

Статья поступила в редакцию 31 января 2023 г; поступила после рецензирования 21 февраля 2023 г; принята к публикации 24 февраля 2023 г.

Original article

Applying corpus-based learning to develop communicative competence among documentary linguistics students

Ilia G. Goncharenko ✉
Russian State Vocational Pedagogical University,
Yekaterinburg, Russian Federation
✉ ilya.goncharenko@rsvpu.ru

Abstract
Introduction. At present, research-based learning is one of the most popular contemporary training methods for secondary vocational education teachers. The paper shows the possibility of using one of these technologies, namely corpus-based learning, for the formation of professionally significant competencies.
Aim. To describe the technology of using corpus data to train documentary linguistics students.
Methods. The analysis of regulatory and legal documentation, as well as educational and methodological literature was carried out. The tools of corpus linguistics were used as teaching tools, namely: distributive analysis for characterizing the combinability of words, quantitative analysis for determining the frequency of words and phrases in general and in a particular type of texts. In addition, a content analysis of students' research works was carried out.
Results. The article provides evidence of the effectiveness of corpus-based learning for the formation of students' research and language competencies. This technology also allows one to expand the repertoire of students' skills and knowledge, and stimulate intra-activity motivation of learning.
Scientific novelty. The article is the first one to describe the technology of applying methods of corpus linguistics to teaching documentary linguistics, and even to form communicative competence among documentary linguistics students — future teachers of the secondary vocational education system.
Practical significance. The results of the study can be used to train vocational education teachers and to teach documentary linguistics.

Keywords: communicative competence, research competence, research based learning, VET-teacher, vocational pedagogical education, document manager, corpus linguistics, documentary linguistics

For citation: Goncharenko, I. G. (2023). Applying corpus-based learning to develop communicative competence among documentary linguistics students. *Vocational Education and Labour Market*, *11*(1), 61–76. (In Russ.) https://doi.org/10.52944/PORT.2023.52.1.005

Received January 31, 2023; revised February 21, 2023; accepted February 24, 2023.

Введение

Стремительное развитие информационных технологий, проникновение их в большинство сфер профессиональной деятельности и в то же время консервативность некоторых из них, проявляющая себя в высокой степени регламентированности, требуют выработки новых норм профессионального взаимодействия внутри этих сфер, новых подходов к разработке регулирующей их нормативно-правовой документации.

Одной из таких консервативных сфер является документоведение, не уступает ему в этом и образование. Когда же речь заходит об образовании, связанном с подготовкой документоведов, то острота противоречий, возникающих между консервативностью содержания данной предметной области и приходящими в нее новыми информационными технологиями, кратно возрастает.

Особую роль в подготовке документоведов играет цикл дисциплин, связанных с изучением коммуникации и ориентированных, согласно ФГОС высшего образования, на формирование универсальной компетенции, позволяющей «осуществлять деловую коммуникацию в устной и письменной формах на государственном языке Российской Федерации и иностранном(ых) языке(ах)»[1].

Согласно профессиональному стандарту «Специалист по организационному и документационному обеспечению управления организацией»[2], все указанные в нем обобщенные трудовые функции, которым соответствуют определенные трудовые действия, необходимые умения и знания, в том или ином виде подразумевают наличие коммуникативной компетенции у специалиста. Это требование логично и обосновано тем, что делопроизводство, документооборот, нормотворчество, как правило, осуществляются на государственном языке, на всей территории Российской Федерации таким является русский язык (РЯ).

Знания и умения, которыми должен обладать выпускник-документовед, можно разделить на несколько групп: знания и умения в области психологии общения; знания и умения, связанные с этикетным поведением; знания и умения в сфере современных информационных технологий; знания и умения в сфере устной и письменной коммуникации. Перечисленные в профстандарте коммуникативные знания и умения должны систематически формироваться на протяжении всего периода обучения совокупностью речеведческих дисциплин, таких как РЯ и деловая коммуникация, речевая коммуникация, риторика, редактирование, документная лингвистика, культура научной речи и др. Отметим, что на практике далеко не всегда составители учебных планов учитывают заданную профессиональным стандартом траекторию

[1] Об утверждении федерального государственного образовательного стандарта высшего образования — бакалавриат по направлению подготовки 44.03.04 Профессиональное обучение (по отраслям): Приказ Министерства науки и высшего образования Российской Федерации от 22.02.2018 № 124. https://fgos.ru/fgos/fgos-44-03-04-professionalnoe-obuchenie-po-otraslyam-124/

[2] Об утверждении профессионального стандарта «Специалист по организационному и документационному обеспечению управления организацией»: Приказ Министерства труда и социальной защиты РФ от 15.06.2020 №333н. http://publication.pravo.gov.ru/Document/View/0001202007160002

формирования коммуникативной компетенции (Анохина и др., 2019), при том, что необходимость ее формирования в профессиональном сообществе не вызывает сомнений (Дрешер, 2014).

Не ставя перед собой глобальную задачу разработать модель коммуникативной подготовки документоведов в течение всего периода их обучения, сосредоточимся лишь на одном, но ключевом ее аспекте, — нормативном. Для того чтобы успешно осуществлять устную и письменную коммуникацию, документовед должен иметь ясное представление о нормах литературного языка. Требование о неукоснительном соблюдении норм литературного языка в сфере делопроизводства декларируется в п. 6 статьи 1 закона о государственном языке: нарушение норм литературного языка при использовании РЯ как государственного признается недопустимым[1].

На нормах литературного языка при подготовке документоведов внимание сосредоточивается прежде всего в ходе изучения документной лингвистики. В существующих учебных пособиях по документной лингвистике реализуется традиционный подход к интерпретации нормативного аспекта языка: норма рассматривается как наиболее распространенный среди образованной части общества и по этой причине рекомендуемый авторами нормативных справочников и словарей вариант употребления языковых выражений (слов, их форм, синтаксических конструкций). При такой интерпретации важно учитывать то обстоятельство, что экспертизой и утверждением перечня нормативных справочников и словарей занимается Правительственная комиссия по РЯ, в компетенцию которой входит разработка требований к составлению словарей, содержащих нормы литературного языка. В соответствии с таким подходом для разрешения спорного вопроса относительно нормативного употребления языкового выражения достаточно обратиться к официально утвержденным справочникам и словарям[2]. Именно такой подход реализуется в учебных пособиях по документной лингвистике, где для выполнения заданий, посвященных языковой норме в деловой речи, также необходимо обратиться к словарям и справочникам.

Не останавливаясь специально на теоретических аспектах утверждения литературных норм, отметим, что в определении нормы решающее значение всегда имел фактор количественного преобладания в речи образованной части общества того варианта, который признается нормативным и до недавнего времени исключительно эксперты (составители нормативных грамматик и словарей) устанавливали это количественное преобладание одного варианта над другим.

В настоящий момент в связи с развитием компьютерной лингвистики (в особенности такого ее направления, как корпусная лингвистика)

[1] О государственном языке Российской Федерации: ФЗ от 01.06.2005 №53-ФЗ. https://www.szrf.ru/szrf/doc.php?nb=100&issid=1002005023000&docid=62

[2] До настоящего момента сохраняет действие нормативный правовой акт «Об утверждении списка грамматик, словарей и справочников, содержащих нормы современного русского литературного языка при его использовании в качестве государственного языка Российской Федерации: Приказ Министерства образования и науки РФ от 08.06.2009 № 195». https://rg.ru/documents/2009/08/21/russkiy-slovari-dok.html

у любого человека, интересующегося языком, появилась возможность обращаться к большим банкам данных, или языковым корпусам, для определения того варианта языкового выражения, который будет наиболее частотным, а значит, наиболее правильным.

Среди рассмотренных учебных пособий по документной лингвистике, вышедших за последние несколько лет и считающихся наиболее авторитетными (Бортников, Михайлова, 2019; Кушнерук, 2013; Леонтьева, 2014; Муравьева, 2016), ни в одном не удалось обнаружить последовательно представленную информацию о применении методов корпусной лингвистики при изучении текстов официально-делового стиля, и только в пособии В. И. Бортникова и Ю. Н. Михайловой содержится всего два задания, предполагающие обращение к языковому корпусу.

Между тем корпусная лингвистика — не только активно развивающееся направление компьютерной лингвистики, но и новая методология и идеология исследования языка, которая стирает утвержденное структуралистской методологией и долгое время считавшееся фундаментальным противопоставление языка и речи (Копотев, 2021), позволяя переосмыслить все существующие в науке о языке и обучении ему понятия и категории. По этой причине представляется необходимым последовательно вводить в практику преподавания речеведческих дисциплин инструментарий корпусной лингвистики.

Методы корпусной лингвистики уже давно проникли в лингводидактику и прежде всего в преподавание иностранных языков (Sinclair, 2004), в том числе русского языка как иностранного (Лебедева, 2020), и в лексикографию (Kilgarriff, 2015; Kosem, 2016); в англоязычной лингводидактике широкое распространение получили термины «data-driven learning» («обучение на основе данных») и «corpus-based learning» («обучение на основе работы с корпусом») (Baker, 2006).

Однако в отечественной научной и учебно-методической литературе, если речь не идет об адресованных филологам систематических вузовских курсах, посвященных корпусной лингвистике, описание применения корпусных данных (в особенности при изучении родного языка) обычно имеет эпизодический, несистематический характер: корпусные данные используются как дополнительный инструмент для решения частных задач.

Актуальность и научная новизна настоящей работы определяется тем, что в ней впервые описывается технология применения методов корпусной лингвистики в практике преподавания документной лингвистики и, шире, в практике формирования коммуникативной компетенции у студентов-документоведов — будущих преподавателей системы СПО. Описанная технология была использована на занятиях по дисциплине «Документная лингвистика» в Российском государственном профессионально-педагогическом университете, где по направлению «Профессиональное обучение (по отраслям)» осуществляется подготовка педагогов СПО по программе бакалавриата «Управление документами в условиях цифровой экономики»; результаты ее применения показали высокую эффективность при развитии у студентов коммуникативной и исследовательской компетенций.

Методы

При проведении исследования был выполнен анализ нормативно-правовой документации и учебно-методической литературы. В качестве педагогических средств обучения в исследовании были использованы следующие инструменты корпусной лингвистики: дистрибутивный анализ для характеристики сочетаемостных возможностей слов, количественный анализ при определении частотности слов и словосочетаний вообще и в конкретном типе текстов в частности. Кроме того, был произведен контент-анализ исследовательских работ, выполненных студентами; для количественной и качественной оценки текстов использовались таблицы с занесением в них информации о наличии или отсутствии признаков, по которым оценивался текст; параметрами оценивания выступали применение в работе корпусных данных, использование двух и более инструментов корпусного менеджера, наличие следов академического мошенничества, выявляемых с помощью автоматических средств проверки оригинальности, новизна полученных данных.

Результаты и обсуждение

В ходе работы была выстроена следующая последовательность представления учебного материала.

1. **Знакомство с основными идеями корпусной лингвистики**. На данном этапе дается определение *лингвистического корпуса* как большого по объему, машиночитаемого, имеющего разметку массива языковых данных, подготовленного лингвистами и (или) специалистами в области информационных технологий; указывается основной признак корпуса — *репрезентативность* в отношении тех задач, для решения которых он создан (Захаров, 2020).

Целесообразным также представляется введение минимально необходимых для работы с корпусом понятий разметки, леммы, токена и тега. *Разметка* определяется как приписывание элементам корпуса (текстам, предложениям, словоформам) определенных признаков (меток, *тегов*) — языковых в случае, например, частеречной (part-of-speechtagging) или лексико-семантической разметки, и неязыковых или экстралингвистических (например, информации о времени создания текста, о его авторе, стиле, жанре и т. п.). *Лемма* определяется как начальная форма для конкретной словоформы, а *лемматизация* — процедура возведения словоформ к начальным формам. *Токеном* называется любой изолированный элемент текста (словоформа, знак препинания), а *токенизацией* — разделение символов на отдельные значимые единицы, осуществляемое, как правило, автоматическими средствами.

2. **Изучение на материале НКРЯ[1] — инструментов корпусного менеджера, представляющих собой графический интерфейс для формирования запросов.** Как известно, в состав НКРЯ входит множество корпусов — газетный, синтаксический, диалектный, поэтический, устный, акцентологический, параллельный, обучающий, корпус детской речи,

[1] Национальный корпус русского языка (НКРЯ) — наиболее разработанный русскоязычный корпус. https://ruscorpora.ru

мультимедийные и исторические корпусы, каждый из которых позволяет решать определенный круг задач. Для получения репрезентативной картины использования РЯ следует обращаться к основному корпусу НКРЯ как к наиболее сбалансированному, то есть состоящему из текстов различных стилей и жанров, воспроизводящему, по словам его создателей, языковой стандарт определенной эпохи, и именно на его данных демонстрировать возможности корпусного менеджера.

2.1. **Инструменты входного интерфейса корпусного менеджера.** К ним относятся поиск по словоформам и поиск по леммам (инструменты, позволяющие запрашивать либо конкретную грамматическую форму, либо все формы); установка расстояния между запрашиваемыми элементами, если этих элементов больше одного; поиск по грамматическим (частеречной принадлежности, падежу, одушевленности/неодушевленности, залогу, времени и т. п.), словообразовательным (типу морфемы) и семантическим признакам (таксономии, топологии и т. п.), а также другим доступным в НКРЯ атрибутам (наличию или отсутствию пунктуационных знаков, типу оборота и т. п.).

При характеристике инструментов входного интерфейса следует обратить внимание на возможности ограничения выдачи результатов заданным типом текстов, используя такие атрибуты, как год создания и сфера функционирования текста, его тематика, жанр, автор (имя, год рождения, пол автора).

В целях изучения особенностей официально-делового стиля выдачу результатов запросов следует ограничивать официально-деловой сферой функционирования, к которой в НКРЯ отнесены тексты деловых писем, законов, расписок, служебных записок, договоров, приказов, указов, уставов, должностных инструкций, резюме, характеристик, положений, судебных решений, ходатайств, заявлений, уведомлений, страховых актов, соглашений, протоколов заседаний, постановлений, доверенностей и т. п., а также публикации в периодических изданиях «Дипломатический вестник», «Адвокат», «Вопросы статистики», «Бухгалтерский учет» и некоторые монографические издания. Однако при исследовании языковых особенностей официально-делового стиля не всегда следует ограничиваться только подкорпусом официально-деловых текстов (включающим всего около 5 млн токенов, в то время как в целом основной корпус НКРЯ насчитывает почти 375 млн токенов), поскольку словообразовательные, морфологические, синтаксические, а также орфоэпические нормы, как правило, имеют надстилевой характер и поддерживаются всеми книжными стилями РЯ.

2.2. **Инструменты выходного интерфейса корпусного менеджера.** Они включают конкорданс, то есть совокупность примеров употреблений запрашиваемых элементов в относительно больших контекстах, которые можно сгруппировать по экстралингвистическим признакам (например, дате создания); KWIC (key word in context) — совокупность примеров употреблений, которые могут быть сгруппированы как по экстралингвистическим, так и по лингвистическим признакам (например, по правому и левому контекстам в алфавитном порядке); график по годам, показывающий частоту запрашиваемого элемента на оси времени;

статистику, позволяющую произвести анализ результатов по автору, сфере функционирования, типу, тематике и жанру текста; n-граммы, представляющие запрашиваемый элемент в сочетании с 2, 3, 4, 5 токенами с указанием их частоты.

Выполняя анализ распределения запрашиваемых элементов по сферам употребления языка, необходимо помнить о том, что в количественном отношении тексты различных сфер представлены в НКРЯ неравномерно: в частности, в подкорпусе текстов официально-деловой сферы 5 102 771 слов, учебно-научной — 39 828 978, обиходно-бытовой — 33 365 547, публицистической — 137 844 556; максимальным в НКРЯ является количество текстов художественной литературы — 150 940 718 слов (все данные приведены по состоянию на 01.01.2023). Это означает, что выводы относительно частотности того или иного грамматического или лексического явления при статистическом анализе[1] результатов запроса необходимо делать на основе сопоставимых данных, которые удобнее всего выразить в процентах. Так, во многих учебниках по стилистике РЯ отмечается, что в текстах научного стиля по сравнению с текстами других функциональных стилей чаще употребляются глаголы несовершенного вида, имеющие форму настоящего времени (так называемое настоящее постоянное) третьего лица. Для подтверждения этого наблюдения с помощью корпусных данных необходимо запросить глагольную форму (синтаксис запроса: V & praes & 3p & ipf), а затем распределенные по сферам функционирования текстов полученные данные разделить на число слов, составляющих соответствующие тексты. В результате будет получен процент глаголов несовершенного вида в форме настоящего времени третьего лица в текстах различных стилей: он равен 2,3 % в текстах официально-деловой сферы, 2,6% — в текстах обиходно-бытовой сферы, 2% — в художественных текстах, 2,6 % — в публицистике и 3,4 % (действительно, максимальное количество) — в текстах учебно-научной сферы. Опыт подготовки документоведов — преподавателей системы СПО к использованию языковых корпусов при изучении языка документов показал, что большинство студентов сначала пытаются сопоставить «сырые», не переведенные в проценты числа, вследствие чего полученные данные интерпретируются неверно. По этой причине следует сосредоточивать особое внимание на необходимости приведения результатов в сопоставимый вид.

3. **Выполнение демонстрационных исследований, связанных с анализом конкретных языковых явлений на основе корпусных данных.**

3.1. **Изучение некоторых вызывающих трудности случаев использования акцентологических норм.** Несмотря на то что в профессиональной деятельности документоведы прежде всего имеют дело с письменной речью, для будущих преподавателей системы СПО орфоэпический аспект изучения языка имеет особое значение, так как главным средством передачи информации на любых занятиях по-прежнему остается

[1] При решении практических задач не всегда возникает необходимость в использовании сложного статистического инструментария корпусной лингвистики; учебная литература, посвященная ему, регулярно обновляется (Brezina, 2018; Wallis 2021), и может быть использована в случае необходимости.

естественный язык. Для изучения особенностей русской акцентологии целесообразно использовать акцентологический корпус НКРЯ, при этом следует сосредоточить внимание на конкретных языковых фактах. Например, в некоторых орфоэпических словарях существительное *обеспечение* предписывается употреблять с ударением только на третьем слоге (Каленчук, 2018, с. 451), тогда как в других допустима постановка ударения на четвертый слог (Резниченко, 2003, с. 641). Для разрешения этого конфликта можно сделать запрос словоформы в акцентологическом корпусе НКРЯ: с ударением на третьем слоге словоформа *обеспечение* встретится в нем 31 раз, с ударением на четвертом — 17. Обладая данной информацией, свидетельствующей о допустимости обоих вариантов, носитель языка может сам выбрать, какой вариант предпочесть, и сделать это в определенном смысле аргументированно.

3.2. **Изучение стилистических компонентов значений слов.** Исследование употребления лексических единиц в НКРЯ позволяет устанавливать их стилистическую окраску без использования толковых словарей. Анализ употребления лексических единиц в стилистическом аспекте, как и остальные представленные в работе виды анализа, целесообразно выполнять на конкретных примерах. В основном корпусе НКРЯ лексема *оферта* встречается 71 раз в публицистических текстах, 48 — в официально-деловых, 7 — в научных. Учитывая, что публицистических текстов в корпусе в 27 раз больше, чем официально-деловых, можно посчитать, что в текстах официально-делового стиля лексема *оферта* встречается в 18 раз чаще, чем в публицистических. Таким образом, лексема тяготеет к употреблению в деловой речи. При анализе тематики текстов с этой лексемой можно увидеть, что 45,93 % из них связаны с темой бизнеса, коммерции, экономики и финансов, а 41,86 % — с темой права. Так, обратившись к корпусным данным, можно уточнить содержащуюся в «Малом академическом словаре русского языка (МАС)»[1] пометку *спец.*, сопровождающую толкование значения данного слова.

3.3. **Изучение семантики лексических единиц.** Наиболее эффективным методом исследования семантики с использованием корпусных данных следует считать анализ дистрибуции, то есть окружения изучаемых единиц. В качестве примера можно привести анализ употреблений паронимов *туристический*/*туристский* в НКРЯ. Словари трудностей[2], а вслед за ними и ГРАМОТА.РУ[3] рекомендуют употреблять *туристический* в значении «связанный с туризмом», а *туристский* в значении «имеющий отношение как к туристам, так и к туризму». В соответствие с данной рекомендацией правильными будут сочетания *туристическая*/*туристская фирма*, *туристическое*/*туристское агентство* и т. п., но будет неправильно употреблять сочетания *туристический рюкзак, туристическая палатка*. Однако результаты запроса лемм *туристский* и *туристический* показывают, что эти прилагательные способны

[1] Словарь русского языка: В 4-х т. / РАН, Ин-т лингв. исследований; Под ред. А. П. Евгеньевой. М.: Русский язык.; Полиграфресурсы, 1999.

[2] Например: Розенталь Д. Э., Теленкова М. А. Словарь трудностей русского языка. М.: Айрис-пресс, 1922.

[3] Справочно-информационный портал ГРАМОТА.РУ. http://new.gramota.ru/spravka/trudnosti?ysclid=lf0nzhnbu3945071175

употребляться в одних и тех же контекстах (можно встретить сочетания *туристический(ая)/туристский(ая) фирма, агентство, поход, поездка, палатка, рюкзак, автобус* и др.), что свидетельствует о почти полном совпадении их лексических значений. В то же время, если воспользоваться вебом как корпусом (Kilgarriff, Grefenstette, 2003) и сделать запрос этих же лексем, ограничив выдачу результатов только сайтом справочной правовой системы «КонсультантПлюс», можно убедиться, что прилагательное *туристический* чрезвычайно редко употребляется в нормативных правовых актах (в отличие от прилагательного *туристский*). Проведенный анализ позволяет продемонстрировать возможности поисковых систем при определении семантических и прагматических (связанных с употреблением) характеристик слов, вызывающих трудности при их использовании в речи, в том числе в речи официально-деловой.

3.4. **Изучение лексической парадигматики.** Лексика любого языка представляет собой довольно сложную систему, элементы которой связаны друг с другом антонимическими, синонимическими, гипер- и гипонимическими отношениями. Еще более сложными связями могут обладать элементы терминологических систем науки; в ряде случаев ясность в особенности употребления терминов может внести обращение к большому количеству примеров их употребления в речи, то есть использование корпусных данных. Такой случай может быть проиллюстрирован синонимическим рядом *правонарушение — преступление — деликт*. Исследование словообразовательных связей, особенностей сочетаемости данных лексем, выполненное не только на материале основного корпуса НКРЯ, но и на материале справочной правовой системы «КонсультантПлюс», — определенного рода концептуальный анализ — позволит продемонстрировать документоведам — преподавателям системы СПО практический смысл лингвистического анализа в различении смысла терминов, имеющих сходное значение.

3.5. **Изучение некоторых аспектов словообразования.** Помимо инструментов, позволяющих запрашивать элементы, совпадающие с аффиксами (приставками, суффиксами, флексиями), корпусный менеджер НКРЯ дает возможность формировать запросы относительно целых классов слов, имеющих сходные словообразовательные особенности, например диминутивов и аугментативов. Запрос диминутивов, выполненный на подкорпусе текстов официально-деловой сферы, может дать весьма любопытные результаты. Диминутивы в принципе не характерны для официально-деловой речи, однако в качестве диминутивов все же будут получены слова *рамка*, *карточка*, *ячейка*, *книжка*, *кружок, цепочка* и т. п. Очевидно, что в данном случае суффиксы *-к-*, *-очк-*, *-ок-* омонимичны суффиксам с уменьшительным значением. Данный результат может послужить основанием для вывода об особенностях эмоционально-экспрессивной тональности текстов официально-делового стиля.

3.6. **Изучение некоторых вызывающих трудности синтаксических правил.** Хотя в основном корпусе НКРЯ синтаксическая разметка отсутствует, корпусный менеджер позволяет создавать синтаксические (правильнее было бы их назвать «псевдосинтаксическими») запросы. В РЯ довольно много примеров колебания в определении синтаксической

нормы, например в глагольном управлении или в согласовании подлежащего со сказуемым. Важно, что эти колебания очень часто наблюдаются в официально-деловой речи. Например, до сих пор многие носители языка сомневаются в выборе формы слова, зависящего от прилагательного *заведующий* (*заведующий чего/чем*). Во всех справочниках по этому поводу можно найти информацию о том, что единственным правильным вариантом будет употребление творительного падежа (при этом не всегда обращается внимание на то, что грамматическое значение родительного падежа в этой конструкции является совсем иным, нежели значение творительного). Можно проверить данную информацию с помощью НКРЯ, оформив запрос следующим образом: 1) лемма *заведующий* (S) + неодушевленное существительное в творительном падеже (S & ins & inan) на расстоянии, равном единице без знаков между ним и предшествующим словом (-bmark); 2) лемма *заведующий* (S) + неодушевленное существительное в родительном падеже (S & gen & ins) на расстоянии, равном единице без знаков между ним и предшествующим словом (-bmark). Разумеется, в результате обработки запросов будет получена информация не о всех конструкциях подобного типа, встречающихся в НКРЯ, а только о тех, в которых представлен прямой порядок слов и отсутствуют распространители, стоящие между главным и зависимым элементом, однако и эта неполная информация будет свидетельствовать о преобладании управления творительным падежом более чем в 10 раз (даже если делать запрос по текстам со снятой грамматической омонимией). Опираясь на данный пример, в сходных ситуациях студенты-документоведы смогут самостоятельно формировать запросы и на основе их результатов принимать решение о том, какую грамматическую форму зависимого компонента словосочетания выбрать или какой тип согласования подлежащего и сказуемого — грамматический или семантический — использовать при составлении документов и обучении ему.

3.7. **Проведение диахронических исследований.** Временной охват текстов официально-деловой сферы в НКРЯ довольно широк — начиная с конца XVII в. и до начала XX в., что дает возможность проводить исследования, посвященные истории официально-делового стиля. Существует возможность проследить историю возникновения и развития определенных клише, например *понеже вышеизображенный*, *того ради повелели Мы* и т. п. Следует отметить, однако, что количество текстов официально-деловой сферы, имеющих историческую ценность, в НКРЯ относительно невелико и оно не позволяет получать сколько-нибудь репрезентативную картину развития официально-делового стиля.

4. **Выполнение студентами заданий, посвященных исследованию особенностей текстов официально-делового стиля.** В рамках курса «Документная лингвистика» студентам первого курса группы документоведов — будущих педагогов СПО, обучающихся по направлению «Профессиональное обучение (по отраслям)», — предлагалось выполнить задание, включающее самостоятельное изучение языковых особенностей текстов официально-делового стиля. Студентам необходимо было выбрать одну из предложенных тем, причем среди них были не только такие, которые подразумевали обращение к корпусным данным, но и такие, которые

можно считать вполне традиционными (например, темы, посвященные деловой переписке, в частности деловому этикету в ней; темы, связанные с изучением нормативно-правовой базы, регулирующей использование РЯ в качестве государственного и т. п.), и написать на эту тему самостоятельную научно-исследовательскую работу с заданной структурой: 1) характеристика проблемы, входящей в сферу документной лингвистики, 2) анализ информации, содержащейся в нормативно-правовой документации, нормативных справочниках и научной литературе для решения проблемы, 3) анализ языкового материала, извлеченного из корпусов или других источников, 4) вывод. Группа состояла из 30 человек. Результаты выполнения задания представлены в таблице.

Оценка результатов проведенных исследований
Evaluation of the research results

Критерии оценивания результата выполненной работы	Значение показателя	
	Соответствует требованиям	Не соответствует требованиям
Для всех отчетных работ		
Выполнение задания	26	4
Использование корпусных данных при выполнении работы	17	9
Для работ, выполненных с использованием корпусных данных		
Использование двух и более инструментов для создания поисковых запросов	14	3
Оригинальность и новизна выполненной работы	14	3
Возможность публикации результатов работы в сборниках и периодических изданиях	12	5

Из таблицы видно, что 4 человека не выполнили задание вообще; при этом количество рейтинговых баллов трех из четырех не выполнивших задание студентов было значительно ниже медианного, у одного равно медианному, что косвенно свидетельствует о связи между невыполнением задания и общей плохой успеваемостью студентов по предмету.

Среди тех, кто выполнил задание, 17 человек (65 % от выполнивших задание) использовали корпусные данные в работе, 9 (35 % от выполнивших задание) предпочли выбрать темы, не связанные с применением корпусных данных. Любопытно, что среди тех, кто предпочел не использовать языковые корпусы, только 5 человек (55 % от выбравших темы без использования корпусных данных) написали оригинальные, без обширных некорректных заимствований, тексты; у 4 человек (45 % от выбравших темы без использования корпусных данных) работы состояли в основном из явных заимствований; в то время как у применивших корпусные методы исследования 14 человек (82 %) работы были оригинальными и только у троих (18 %) работы содержали обширные некорректные заимствования. Это позволяет сделать вывод о том, что использование корпусных методов исследования способно в определенной степени снизить вероятность так называемого академического

мошенничества: если работа выполняется в рамках оригинальной темы, то получается относительно оригинальный результат, и в плагиате у студентов не возникает необходимости.

При оценке степени овладения инструментарием корпусной лингвистики следует отметить, что среди 17 работ, выполненных на материале корпусов, в 14 (82 %) прослеживалось использование двух и более инструментов входного и выходного интерфейса (запросов по подкорпусам, запросов по сайтам справочных систем «Гарант» и «КонсультантПлюс», синтаксических запросов и т. п.), в 3 работах (18 %) такого разнообразия примененных методов не наблюдалось. Данное обстоятельство свидетельствует о том, что большинство из выбравших тематику, связанную с корпусными исследованиями, овладело основным инструментарием корпусного менеджера.

Среди 17 выполненных с использованием корпусных методов работ определенной научной новизной, позволяющей считать их хоть и небольшими, но самостоятельно выполненными исследованиями, обладали 12, то есть 70 % от всех работ, выполненных на материале корпусов. Если определить отношение авторов этих работ к общему количеству студентов в группе, то можно заметить, что благодаря примененной технологии 40 % студентов оказались вовлечены в научный поиск. Таким образом, в рамках изучения дисциплины «документная лингвистика» может решаться задача устранения недостатка научно-исследовательских компетенций у преподавателей СПО (Коновалов и др., 2021).

Заключение

Применение технологии корпусных исследований при формировании языковой компетенции у студентов-документоведов позволило выявить ряд ее несомненных преимуществ перед традиционными технологиями.

1. Обучение через исследование позволяет эффективно формировать исследовательские компетенции у будущих преподавателей системы СПО. Нередко научное исследование рассматривается студентами (в особенности младших курсов) как возможность выдать результаты чужого интеллектуального труда за свои, однако использование технологий корпусных исследований способно уберечь их от такого соблазна — при работе с корпусными данными студенты оказываются в ситуации *вынужденной самостоятельности*.

2. Использование технологии корпусных исследований дает возможность расширить репертуар формируемых у студентов умений и знаний, при этом необходимость овладения навыками поиска источников информации, выполнения их обзора или реферирования, а также запроса информации в существующих базах данных, подразумеваемая выполнением заданий, соотносится с декларируемыми профессиональным стандартом требованиями к работникам (трудовые функции 3.3.13, 3.4.1, 3.4.2 и др.).

3. Описанная технология в максимальной степени позволяет охватить межпредметные связи: коммуникативная компетенция формируется при взаимодействии с профессионально значимыми, аутентичными

текстами, используются современные информационные технологии, причем такие, в которых реализуются ключевые принципы работы с любыми информационными базами, задействуются элементы индивидуальной проектной деятельности, — все это позволяет вызвать у студентов интерес к получению знания, стимулировать внутридеятельностную мотивацию учения.

4. Очевиден явно выраженный практико-ориентированный характер данной работы: ни составление документов, ни их редактирование невозможны без ясного представления о норме (а при его отсутствии в том или ином трудном случае — без четкого понимания, какими источниками следует воспользоваться для того, чтобы его составить) и ее реализации в узусе; описанная технология работы с источниками информации о норме и узусе ориентирована на то, чтобы помочь студентам в решении профессиональных задач в будущем.

В то же время нельзя не отметить ограничения в применении описанной технологии, связанные прежде всего с отсутствием возможности работать с завершенными текстами: структурные, композиционные, жанровые особенности текстов документов при таком подходе оказываются вне рассмотрения, однако эти недостатки могут быть компенсированы другими формами работы.

Список литературы

1. Анохина С. А., Козько Н. А., Позднякова Н. В. Проблемы лингвистической подготовки документоведов // Перспективы науки и образования. 2019. № 1 (37). С. 94–107. https://doi.org/10.32744/pse.2019.1.7

2. Бортников В. И., Михайлова Ю. Н. Документная лингвистика. М.: Флинта; Екатеринбург: Изд-во Урал. ун-та, 2019. 132 с.

3. Дрешер Ю. Н. Коммуникативная компетентность специалистов-документоведов. Проблемы и перспективы развития // Вестник РГГУ. Сер. Документоведение и архивоведение. Информатика. Защита информации и информационная безопасность. 2014. № 2 (124). С. 94–98.

4. Захаров В. П., Богданова С. Ю. Корпусная лингвистика. СПб.: Изд-во С.-Петерб. ун-та, 2020. 234 с.

5. Каленчук М. Л., Касаткина Л. Л., Касаткина Р. Ф. Большой орфоэпический словарь русского языка. Литературное произношение и ударение начала XXI века: норма и ее вариант. М.: АСТ-Пресс Школа, 2018. 1024 с.

6. Коновалов А. А., Щипанова Д. Е., Лыжин А. И., Чернышов Б. А. О дефиците исследовательских компетенций у педагогов СПО: результаты исследования // Профессиональное образование и рынок труда. 2021. № 2. С. 112–125. https://doi.org/10.52944/PORT.2021.45.2.009

7. Копотев М. В. О некоторых следствиях корпусной лингвистики для общей теории языка // Филологический класс. 2021. Т. 26. № 2. С. 90–102. https://doi.org/10.51762/1FK-2021-26-02-07

8. Кушнерук С. П. Документная лингвистика. М.: Флинта: Наука, 2013. 253 с.

9. Лебедева М. Ю. Дано мне тело — что мне делать с ним? Применение корпусных технологий в лингводидактике РКИ // Русский язык за рубежом. 2020. № 6. С. 4–13. https://doi.org/10.37632/PI.2020.283.6.001

10. Леонтьева Т. В. Документная лингвистика. М.: Флинта: Наука, 2014. 253 с.

11. Муравьева Н. В. Документная лингвистика. М.: Термика, 2016. 609 с.

12. Резниченко И. Л. Орфоэпический словарь русского языка: Произношение. Ударение. М.: АСТ: Астрель, 2007. 1182 с.

13. Baker P. Glossary of corpus linguistics. Edinburgh: Edinburgh University Press, 2006. 187 p.

14. Brezina V. Statistics in corpus linguistics: a practical guide. Cambridge: Cambridge University Press, 2018. 296 p. https://doi.org/10.1017/9781316410899

15. How to use corpora in language teaching / Ed. by J. M. Sinclair. Philadelphia: John Benjamins Publishing, 2004. 307 p.

16. Kilgarriff A. Using corpora as data sources for dictionaries // In H. Jackson (Ed.). The Bloomsbury companion to lexicography. London: Bloomsbury, 2015. P. 71–88.

17. Kilgarriff A., Grefenstette G. Introduction to the special issue on the web as corpus // Computational Linguistics. 2003. Vol. 29. No. 3. P. 333–347. https://doi.org/10.1162/089120103322711569

18. Kosem I. Interrogating a corpus // In P. Durkin (Ed.). The Oxford handbook of lexicography. Oxford: Oxford University Press, 2016. P. 76–93.

19. Wallis S. Statistics in corpus linguistics: a new approach. London: Routledge, 2020. 355 p.

References

Anokhina, S. A., Kozko, N. A., & Pozdnyakova, N. V. (2019). Problems of language training of document specialists. *Perspectives of Science & Education,* 1, 94–107. (In Russ.) https://doi.org/10.32744/pse.2019.1.7

Baker, P., Hardie, A., & McEnery, T. (2006). *Glossary of corpus linguistics*. Edinburgh University Press.

Bortnikov, V. I., & Mikhailova, Yu. N. (2019). *Dokumentnaia lingvistika* [Documentary linguistics]. Flinta; Ural University Press. (In Russ.)

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press. https://doi.org/10.1017/9781316410899

Drescher, Yu. N. (2014). Communicative competence of document specialists. Problems and development prospects. *Bulletin of the Russian State University for the Humanities,* 2, 94–98. (In Russ.)

Kalenchuk, M. L., Kasatkina, L. L., & Kasatkina, R. F. (2018). *Bol'shoi orfoepicheskii slovar' russkogo iazyka. Literaturnoe proiznoshenie i udarenie nachala XXI veka: Norma i ee variant* [Big dictionary of the Russian language pronunciation. Literary pronunciation and accentuation at the beginning of the XXI century: Standard of pronunciation and accentuation and its variant]. AST-Press Shkola. (In Russ.)

Kilgarriff, A. (2015). Using corpora as data sources for dictionaries. In H. Jackson (Ed.), *The Bloomsbury companion to lexicography* (pp. 71–88). Bloomsbury.

Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the special issue on the web as corpus. *Computational Linguistics, 29*(3), 333–347. https://doi.org/10.1162/089120103322711569

Konovalov, A. A., Shchipanova, D. Y., Lyzhin, A. I., & Chernyshov, B. N. (2021). On the lack of research competencies among teachers of secondary vocational

education: The results of the study. *Vocational Education and Labour Market*, 2, 112–125. (In Russ.) https://doi.org/10.52944/PORT.2021.45.2.009

Kopotev, M. V. (2021). Some thoughts on corpus and general linguistics. *Philological Class, 26*(2), 90–102. (In Russ.) https://doi.org/10.37632/PI.2020.283.6.001

Kosem, I. (2016). Interrogating a corpus. In P. Durkin (Ed.), *The Oxford handbook of lexicography* (pp. 76–93). Oxford University Press.

Kushneruk, S. P. (2013). *Dokumentnaia lingvistika* [Documentary linguistics]. Flinta: Nauka. (In Russ.)

Lebedeva, M. Yu. (2020). I'm given a corpus—what to do with it? Corpus technologies in Russian language teaching and learning. *Russian Language Abroad*, 6, 4–13. (In Russ.) https://doi.org/10.37632/PI.2020.283.6.001

Leontyeva, T. V. (2014). *Dokumentnaia lingvistika* [Documentary linguistics]. Flinta: Nauka. (In Russ.)

Muraviova, N. V. (2016). *Dokumentnaia lingvistika* [Documentary linguistics]. Termika. (In Russ.)

Reznichenko, I. L. (2007). *Orfoepicheskii slovar' russkogo iazyka: Proiznoshenie. Udarenie* [Russian pronunciation dictionary]. AST: Astrel'. (In Russ.)

Sinclair, J. M. (Ed.). (2004). *How to use corpora in language teaching*. John Benjamins Publishing.

Wallis, S. (2020). *Statistics in corpus linguistics: A new approach*. Routledge.

Zakharov, V. P., & Bogdanova, S. Yu. (2020). *Methods in corpus linguistics*. SPb University Press. (In Russ.)

Информация об авторе

Гончаренко Илья Георгиевич, канд. филол. наук, доцент, доцент кафедры документоведения, права, истории и русского языка, Российский государственный профессионально-педагогический университет, ORCID: https://orcid.org/0000-0001-5515-9332, ilya.goncharenko@rsvpu.ru

Конфликт интересов: автор заявляет об отсутствии конфликта интересов.

Автор прочитал и одобрил окончательный вариант рукописи.

Information about the author

Ilia G. Goncharenko, Cand. Sci. (Philology), Docent, Associate Professor of the Department of the Documentation, Law, History and Russian Language, Russian State Vocational Pedagogical University, ORCID: https://orcid.org/0000-0001-5515-9332, ilya.goncharenko@rsvpu.ru

Conflict of interests: the author declare no conflict of interest.

All author have read and approved the final manuscript.