

Корпусная лингводидактика в контексте парадигмы открытых образовательных ресурсов

М.С. Коган

Санкт-Петербургский политехнический университет Петра Великого

m_kogan@inbox.ru

Аннотация

Развитие интернета и информационного общества привело к появлению важного тренда на открытое образование, науку, информацию. Эта тенденция поддержана документами ООН и UNESCO, решениями двух всемирных международных конгрессов по развитию открытых образовательных ресурсов (ООР). Одним из препятствий для широкого использования корпусной лингводидактики для изучения ИЯ является тот факт, что многие корпуса, описанные в научных статьях, создаются исключительно для исследовательских целей и недоступны научной и педагогической общественности. Проблемой также является недостаточная компетенция преподавателей иностранного языка в области корпусной лингвистики. Находящиеся в открытом доступе корпуса являются ценным ресурсом для лингвистического анализа, но разработка дидактических материалов на их основе для учебных целей при изучении родного и иностранного языков является весьма трудной и трудоемкой задачей для большинства преподавателей иностранных языков в школе и вузе. Решением проблемы может стать разработка набора заданий/упражнений, привязанных к конкретному доступному корпусу. В статье обсуждаются имеющиеся передовые практики, реализующие этот подход, и обсуждаются варианты повышения корпусной компетенции преподавателей ИЯ за счет имеющихся открытых интернет-ресурсов.

Ключевые слова: ООР, онлайн корпуса, корпусная лингводидактика, дидактические материалы, повышение квалификации

Библиографическая ссылка: Коган М.С. Корпусная лингводидактика в контексте парадигмы открытых образовательных ресурсов // Компьютерная лингвистика и вычислительные онтологии. Выпуск 6 (Труды XXV Международной объединенной научной конференции «Интернет и современное общество», IMS-2022, Санкт-Петербург, 23 – 24 июня 2022 г. Сборник научных статей). — СПб.: Университет ИТМО, 2022. С. 63-75. DOI: 10.17586/2541-9781-2022-6-63-75

Введение

Развитие интернета и информационного общества привело к появлению важного тренда на открытое образование, науку, информацию. Инициатором этого движения считается Майкл Стерн Харт, изобретатель электронных книг и основатель проекта «Гутенберг» в 60-е годы прошлого века [1]. В XXI веке этот тренд задала Будапештская Инициатива «Открытый Доступ»¹, к которой с тех присоединились тысячи организаций. В этом документе Открытый доступ определялся как «открытые для всех публикации

¹ The Budapest Open Access Initiative (BOAI) 14. 02. 2002.
<https://www.budapestopenaccessinitiative.org/read/>
<https://www.budapestopenaccessinitiative.org/read/russian-translation/>

в Интернете, которые можно читать, разгружать, копировать, распространять, распечатывать, находить или присоединять к полным текстам соответствующих статей, использовать для составления указателей, вводить их как данные в программное обеспечение или использовать для других законных целей при отсутствии финансовых, правовых и технических преград» ... при соблюдении целостности работы и обязательной ссылке на имя автора при использовании и цитировании. В последующие годы Будапештская инициатива получила развитие в многочисленных документах ООН и ЮНЕСКО, например, в рекомендациях по развитию и продвижению открытого доступа², рекомендациях по открытой науке³, по работе с открытыми образовательными ресурсами (ООР) в сфере высшего образования⁴, об открытых образовательных ресурсах⁵, а также рекомендациях по развитию практик использования ООР и т.д. Исследования, касающиеся влияния открытого образования на изучение и преподавание иностранных языков, собраны в томе [2], находящемся в открытом доступе.

Открытые образовательные ресурсы обладают рядом характеристик, важнейшими из которых являются бесплатный доступ, право на использование, адаптацию и распространение третьими лицами без ограничений или с незначительными ограничениями. Некоторые исследователи дополняют список характеристик понятия открытость до 12 пунктов. Обзор исследований принципов и характеристик ООР можно найти в статье Космас с коллегами [3].

Одна из характеристик открытой науки: «прозрачность, контроль, критический анализ и воспроизводимость результатов ... в целях усиления надежности и точности научных результатов, а также воздействия науки на общество» [4, с. 12]. Специалисты в области прикладной и корпусной лингвистики говорят о кризисе воспроизводимости (replication crisis) (см., например, [5]). Анализируя тенденции развития исследований в области обучения ИЯ с использованием подходов корпусной лингвистики (DDL) за последние 30 лет на материале 480 публикаций в авторитетных рецензируемых англоязычных журналах, глав монографий и материалах авторитетных конференций, А. Болтон и Н. Вяткина обнаружили полное отсутствие исследований, направленных на частичное воспроизведение/ повторение «классических» работ в новых условиях с другими категориями обучающихся [6, р. 83.]. Следует заметить, что авторы в соответствии с принципами открытой науки разместили информацию о всех отобранных для анализа публикациях в открытом доступе в онлайн репозитории инструментов и материалов для исследований в области иностранных языков IRIS⁶.

1. Важность концепции ООР для более широкого использования корпусной лингводидактики в обучении ИЯ

Несмотря на эффективность использования подходов корпусной лингвистики в обучении ИЯ как показал мета-анализ, проведенный Болтоном и Коббом [7] и давно предсказанную революцию в методах обучения ИЯ благодаря широкому использованию корпусов [8], этого до сих пор не произошло и А. Чамбес в статье 2019 г. обратила внимание на необходимость преодоления разрыва между исследованиями в области КЛ

² Policy guidelines for the development and promotion of open access 2012. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000215863>

³ ЮНЕСКО Проект рекомендации по открытой науке 2021. URL: https://unesdoc.unesco.org/ark:/48223/pf0000378841_rus

⁴ ЮНЕСКО Рекомендации по работе с открытыми образовательными ресурсами (ООР) в сфере высшего образования 2011, 2015. URL: https://unesdoc.unesco.org/ark:/48223/pf0000213605_rus

⁵ Guidelines on the development of open educational resources policies // UNESCO and COMMONWEALTH OF LEARNING, 2019 The Global Education 2030 Agenda URL: <https://unesdoc.unesco.org/ark:/48223/pf0000371129>

⁶ IRIS. URL: <https://www.iris-database.org/iris/app/home/detail?id=york%3a939061&ref=search>

и методикой преподавания ИЯ [9]. Разные исследователи называют разные причины сложившейся ситуации. Наша попытка проанализировать их привела к списку из 9 причин [10]. В контексте парадигмы ООР мы остановимся на трех проблемах, а именно: 1) недоступность корпусов, описанных в статьях; 2) недостаточная подготовка преподавателей ИЯ по корпусной лингводидактике; 3) отсутствие общедоступных заданий, созданных на базе находящихся в свободном доступе онлайн корпусов, для непосредственного использования в учебном процессе.

Одним из препятствий к более широкому распространению подходов, описанных в статьях, является то, что многие исследования выполнены на корпусах, составленных специально для целей исследования и потому недоступных для научной и педагогической общественности [9]. В свободном доступе находится достаточно большое количество больших корпусов, в первую очередь национальные корпуса: НКРЯ, BNC, COCA, DWDS (представительный список онлайн корпусов доступен в разделе «Другие корпуса» сайта НКРИ⁷ с указанием режима обращения: «доступен для онлайн-поиска, требуется регистрация, доступен для скачивания»).

Проблемой, однако, является отсутствие доступного банка упражнений, которые преподаватель мог бы использовать при изучении той или иной темы. Многочисленные примеры использования корпусов в обучении ИЯ опубликованы в научных статьях и материалах конференций, но доступ к ним обычному преподавателю вуза, не говоря уже об учителях средней школы, затруднен: эти публикации нигде не каталогизированы и не описаны с точки зрения имеющихся в них корпусных заданий определенного типа. В качестве примера можно сослаться на недавние публикации на материале Национального корпуса русского языка [11, 12, 13]. Изданные немногочисленные пособия с поисковыми заданиями, как правило, не относятся к открытым ресурсам, например, [14, 15, 16], а находящиеся в открытом доступе меняют свой адрес. Так, магистерская диссертация Е.М. Шоу с сайта COCA⁸, где она находилась вплоть до весны 2020 г., переместилась в архив университета Бригама Янга⁹. Если преподаватель – «в теме», то он вынужден метаться между корпусом и разными источниками с упражнениями. По мнению Н. Вяткиной, доступные онлайн корпуса английского языка и других языков не содержат как инструкций по их использованию для обучения ИЯ, так и разработанных заданий/системы упражнений, привязанных к конкретному корпусу [17, р. 364].

Проблемой является и недостаточная подготовка преподавателей ИЯ в области корпусной лингвистики и корпусной лингводидактики. Проблема серьезна и простых путей ее решения, кажется, не существует. В нашем исследовании, имевшем целью оценить влияние корпусной лингвистики на систему высшего образования в России, мы пришли к выводу, что на современном этапе в соответствии с требованиями ФГОС все студенты филологических специальностей получают базовые знания о корпусной лингвистике, некоторые имеют публикации по теме использования подходов корпусной лингвистики в обучении ИЯ, зарегистрированные наукометрических базах данных разного ранга [18]. Однако открытым остается вопрос, насколько они готовы регулярно использовать эти знания в практической деятельности после окончания обучения. А. Ленко-Шиманска считает, что элективного односеместрового специального курса по использованию корпусов и подходов КЛ в обучении ИЯ недостаточно для того, чтобы будущие преподаватели ИЯ могли успешно интегрировать подходы КЛ в собственную педагогическую практику [19, р. 234]. Исследователь предполагает, что ситуация улучшится, если студенты будут изучать возможности корпусной лингвистики наряду

7 Информация о других корпусах на сайте НКРЯ. URL: <https://ruscorpora.ru/page/corpora-other/>

8 Недействующий адрес магистерской диссертации Shaw E.M. URL: Teaching vocabulary through data driven learning https://corpus.byu.edu/coca/files/Teaching_Vocabulary_Through_DDL.pdf

9 Действующий адрес магистерской диссертации Шоу ЕМ. URL: Shaw E.M. Teaching vocabulary through data-driven learning <https://scholarsarchive.byu.edu/etd/3024/>

с другими подходами в курсе методики обучения ИЯ и использовать их при изучении других теоретических и практических лингвистических дисциплин. Для мотивированных будущих и действующих преподавателей ИЯ, желающих освоить основы КЛ и способы ее применения в обучении ИЯ самостоятельно, доступны разные способы, которые мы рассмотрим в следующем разделе.

2. Обзор открытых образовательных ресурсов для повышения квалификации по КЛ

Есть целый ряд образовательных ресурсов, находящихся в свободном/открытом доступе, которые можно использовать для самообразования в области корпусной лингвистики. Мы рассмотрим коллекции YouTube, MOOK курсы и записи вебинаров, организованные ведущими центрами по корпусной лингвистике.

Начнем с обзора **видеоресурсов** по корпусной лингвистике доступных на YouTube. По запросу «Corpus linguistics» YouTube выдает целый ряд ссылок на отдельные ролики и на целые серии, в которых последовательно объясняются основы этой дисциплины. Например, Яссин Иабдуан разместил в своем канале за последние 2 года уже 11 роликов продолжительностью от 6:57 до 24:21 мин. каждый¹⁰.

Элен Ле Фолл выложила 10 роликов¹¹. Наиболее популярными, набравшими за год более 1000 просмотров, являются ролики по освоению Sketch Engine и по созданию собственного специального корпуса в корпусе COCA. Разработчики крупных корпусов также имеют свои YouTube каналы. Так, Sketch Engine разместил 24 ролика. Продолжительность обучающих видео колеблется от 0:29 до 5:25. Л. Антони в своем канале разместил 11 видео роликов, объясняющих как использовать основные функции программы AntConc Версии 3.4.0¹², широкоизвестного корпусного менеджера, который он разработал. Длительность видео колеблется от 4:15 до 14:38.

Э. Ле Фолл призывает активно создавать небольшие видео, делаясь в них собственным опытом применения подходов корпусной лингвистики в обучении ИЯ и скрытыми хитростями работы в доступных онлайн корпусах или применения корпусных менеджеров. В русскоязычном сегменте интернета такие материалы практически отсутствуют.

Массовые открытые онлайн курсы в массовом сознании являются примером открытых образовательных ресурсов, хотя специально проведенный анализ сравнения MOOK курсов и ООР позволил выявить важные отличия между этими категориями ресурсов [20]. Тем не менее, их доступность делает их очень ценным ресурсом для самообразования. Наиболее популярные MOOK курсы по корпусной лингвистике это:

— *Corpus Linguistics: Method, Analysis, Interpretation*, разработанный в Ланкастерском университете Т. МакЭнери и размещенный на платформе FutureLearn¹³. Этот MOOK является введением в КЛ, затрагивает вопросы обучения ИЯ с использованием КЛ, знакомит с крупными онлайн корпусами и новым корпусным менеджером LancsBox., разработанным В. Брезиной.

10 Ссылка на YouTube канал Yassine Iabdounane URL: https://www.youtube.com/c/YassineIabdounane/videos?view=0&sort=dd&shelf_id=0

11 Ссылка на YouTube канал Elen Le Foll. URL: https://www.youtube.com/channel/UCRV6et-PUdiS9WZht3y7WcQ/videos?view=0&sort=dd&shelf_id=0

12 Ссылка на серию видео по освоению программы AntConc Версии 3.4.0., созданных разработчиком L. Anthony. URL: https://www.youtube.com/playlist?list=PLiRIDpYmiCOTa0-Hdvc1D7hG6dmiS_TZj

13 Ссылка на MOOK Corpus Linguistics: Method, Analysis, Interpretation. URL: <https://www.futurelearn.com/courses/corpus-linguistics>

- *Введение в корпусную лингвистику*¹⁴ - курс ВШЭ, представленный А.И. Левинзон. Он размещен на Национальной платформе «Открытое образование». Курс дает представление о практическом применении достижений современной корпусной лингвистики, учит пользоваться существующими корпусами русского и английского языков, позволяет освоить базовые методы корпусного преподавания в повседневной работе.
- Курс *Improving writing through corpora: Data-driven learning* относится к категории SPOC (small private online course: маленький частный онлайн курс) Разработан П. Кроссвейтом в университете Квинсленда и размещен на платформе EdEx¹⁵. Его целью является развитие навыков академического письма с прямым обращением к корпусу. Задания разработаны на основе открытого корпуса BAWE из коллекции корпусов Sketch Engine. Апробация курса описана в статье [21].
- В эту же группу ресурсов можно включить курс, разработанный А. Ленко-Шиманска *Corpora in Language Teaching* для студентов варшавского университета, который находится в свободном доступе на сайте университета¹⁶. Целью курса является знакомство участников с базовыми концепциями: корпус, анализ корпуса, применение корпусных подходов в обучении иностранным и родному языку и развитие навыка создания учебных материалов на основе корпусного подхода. Важно, что изучаются только общедоступные корпуса и корпусные инструменты, находящиеся в открытом доступе. Апробация курса описана А. Ленко-Шиманска в статьях [19, 22].

Серии вебинаров

В честь 50-летия основания магистерской программы по прикладной лингвистике университета Сан-Паолу (Бразилия) руководители программы организовали серию вебинаров по корпусной лингвистике, пригласив для участия всемирно известных ученых: Д. Байбера, М. Скотта, Т. МакЭнери, Ю. Ромер, М. Дэвис, Р. Реппен и др. Вебинары проходили еженедельно с апреля по сентябрь 2020 г., включая 2 недели интенсива в августе и сентября, в течение которых проводилось несколько мастер-классов по использованию новых программных продуктов или новых версий уже известных программ. Например, М. Скотт представил последнюю версию корпусного менеджера WordSmith 8.0, которая на тот момент была в свободном доступе. Записи вебинаров доступны на канале YouTube LAEL Webinars¹⁷ и на странице кафедры LAEL 50 years, где также размещены презентации докладчиков¹⁸.

Осенью 2021 г. эту инициативу продолжили вебинары, организованные на базе Школы языков и культур университета Квинсленда (Австралия), по теме *International Perspectives on Corpus Technology for Language Learning*. Вебинары проходили с октября 2021 по апрель 2022 г. Всего состоялся 21 вебинар. Докладчиками также были очень крупные исследователи в области корпусной лингвистики и разработки программного обеспечения А. Франкенберг-Гарсия, А. Ленко-Шиманска, А. Болтон, П. Кроссвейт, Л. Антони. Последний познакомил с новой версией своего корпусного менеджера AntConc4, функционал которого существенно расширен по сравнению с версией AntConc3.4.0.

14 Ссылка на курс «Корпусная лингвистика». URL: https://openedu.ru/course/hse/CORPUS/?session=fall_2020

15 Ссылка на курс *Improving writing through corpora: Data-driven learning*. URL: <https://edge.edx.org/courses/course-v1:UQx+SLATx+2019/about>

16 Ссылка на курс *Corpora in Language Teaching*. <https://moodle.ils.uw.edu.pl/course/view.php?id=101>, созданный А. Ленко-Шиманска

17 Ссылка на запись вебинаров в честь 50-летия кафедры LAEL. URL: https://www.youtube.com/channel/UC9-sGRAJFwaPj014U3aR0jA/videos?view=0&sort=dd&shelf_id=0

18 Ссылка на записи вебинаров в честь 50-летия кафедры LAEL с презентациями докладчиков. URL: http://corpuslg.org/lael_english/webinars/

Записи вебинаров доступны на канале YouTube П. Кроссвейта¹⁹, где также собраны презентации автора на ряде международных конференций по корпусной лингвистике. (Например, TaLC 2020, EuroCALL 2021 и др.). Развернутое описание вебинаров и презентации докладчиков размещены на сайте университета Квинсленда²⁰. Анонсы предстоящих событий публикуются в социальных сетях, таких группах, как *The Eurocall Corpus SIG* на фейсбуке, *Applied Linguistics* в LinkedIn.

В русскоязычном сегменте интернета корпусной лингвистике посвящены лекции и беседы В.А. Плунгяна, Д.О. Добровольского в Постнауке, Д. Бузаджи, В.П. Захарова и некоторых других известных исследователей, а также видеозаписи заседаний конференции «Корпусная лингвистика»²¹, которая проходила в дистанционном режиме на кафедре Математической лингвистики СПбГУ в 2021г. Обновленная в мае 2022 г версия НКРЯ содержит существенно расширенный «Образовательный портал Национального корпуса русского языка»²².

3. Обзор готовых ООР для использования подходов корпусной лингводидактики в обучении ИЯ

Рассмотрим предложение Н. Вяткиной об эффективности привязки планов уроков/заданий по работе с корпусом к конкретным корпусам [17, р. 364-365]. На первый взгляд может показаться, что для больших онлайн корпусов, таких как НКРЯ или СОСА, проблема неактуальна, т.к. они снабжены инструкциями пользователя внушительного объема: инструкция по НКРЯ содержит 87 стр., инструкция по СОСА – 27 стр. с дополнительными инструкциями по созданию собственного специального корпуса, дополнительным инструментам для режима просмотра KWIC и т.д. Н. Вяткина указывает на то, что эти руководства пользователя очень далеки от знакомых учителю планов уроков с использованием корпусов [23]. По ее мнению, инструкции для преподавателей должны базироваться на педагогических принципах, подразумевающих введение материала, сопровождение, вовлечение, и идти от простого к сложному, от большей поддержки со стороны преподавателя к заданиям, выполняемым самостоятельно. В своих статьях она описывает примеры заданий обращений к корпусу немецкого языка DWDS для учащихся с разным уровнем владения немецким языком (от начинающих до продвинутых), предполагающих разный формат работы (индивидуальную работу, работу в парах и малых группах, введение материала преподавателем). Основным фокусом является расширение словарного запаса, и задания могут включать освоение новых лексико-грамматических характеристик слов с помощью примеров из корпуса, верификация примеров употребления слов/словосочетаний в учебниках примерами из корпуса. Примеры заданий, приведенные в статье, соответствуют разным инструментам корпуса DWDS: базовый поиск, временная ось, визуализация коллокаций через облако слов и др. [23].

Другим блестящим примером создания ООР для преподавания ИЯ на основе конкретного корпуса является проект CORPUS FOR SCHOOLS: Teaching English Language with Corpus Linguistics, создаваемый в Ланкастерском университете под руководством Д. Габласовой. Учебные материалы привязаны к большим выборкам устной речи BNC 1990-х и 2014, представляющих неформальные беседы носителей британского варианта

19 Ссылка на записи вебинаров International Perspectives on Corpus Technology for Language Learning URL:

https://www.youtube.com/channel/UC4rCKjMDRFs2ZQzv8ID6NwQ/videos?view=0&sort=dd&shelf_id=0

20 Ссылка на записи вебинаров с описанием и презентациями докладчиков. URL: <https://languages-cultures.uq.edu.au/event/7334/international-perspectives-corpus-technology-language-learning-seminar-series>

21 Ссылка на записи заседаний конференции Corpora 2021. URL: https://www.youtube.com/playlist?list=PLvN-HgiIJTKTVM29etGSGf9nrAHAiUT8_

22 Образовательный портал НКРЯ. URL: <https://studiorum.ruscorpora.ru/>

английского языка. Уроки разработаны как для носителей английского языка, готовящихся сдать итоговый экзамен по курсу средней школы (A-level exam) так и для изучающих английский язык как второй или иностранный. Выборки представлены на платформе BNClab²³. Для выпускников английских школ задания представлены по следующим рубрикам/категориям *Language Change, Language and Gender, Language and Region, Language and Age, Spoken Communication*. Для изучающих английский язык как второй или иностранный разработаны уроки по рубрикам *Current English Use, Spoken Communication, Communicative Skills*, в рамках которых изучаются реальные явления современного английского языка, не отраженные в учебниках английской грамматики: например, изменение функции местоимения 3 лица множественного числа *they* или способы выражения множественного числа во втором лице. Учебные материалы по каждой теме включают раздаточные упражнения (4 задания: чтение абзаца, иллюстрирующего правило, анализ конкордансных строк, трансформация, обсуждение предлагаемых тем с обязательным использованием изучаемого грамматического явления (в данном случае, *they* в функции 3 лица ед. числа.) По паролю доступен файл с ответами и комментариями для преподавателя.

Еще одним примером привязки планов уроков к корпусу является мультимедийный ресурс *EnglishCentral*. Он создан с использованием корпусных технологий и содержит более 14000 видео и 250 млн строк аудиозаписей для развития навыков аудирования и говорения; каждую неделю добавляется по 50 новых видео. Контент содержит тысячи отдельных видео с указанным уровнем сложности и более 100 мини-курсов в категориях: *Academic English, Travel English, Business English, Career English, Useful Expressions* и др. Каждый видеофрагмент длится 1-2 минуты. Показ сопровождается субтитрами. Видео можно остановить в любом месте, чтобы прослушать незнакомое слово несколько раз и увидеть примеры его употребления в всплывающем окне, а затем вернуться к просмотру видео урока. Пользователю предлагают посмотреть 5 тематически близких видео выбранного уровня сложности, которые подбираются системой автоматически. В систему интегрирована передовая технология записи речи IntelliSpeechSM, которая автоматически сравнивает произнесенный фрагмент с спектрограммой речи диктора, оценивая близость попытки в процентах. Учащийся может неоднократно повторять попытки, пытаясь приблизить свое произношение к речи диктора. В бесплатной версии такая возможность представляется только для двух видео в месяц. Другая возможность платной версии касается тестирования незнакомой лексики. В контексте статьи важно отметить, что для каждого видео урока имеется интерактивный План урока (Lesson Plan). Он состоит из 4-х заданий: Video transcript с пропусками нескольких (4-6) слов, которые можно заполнить по памяти, пользуясь контекстом, получить подсказку в виде определения из словаря, впечатанного в пробел или прослушать пропущенное слово, кликнув по пробелу, получить правильный ответ, выбрать отдельные предложения для последующей практики. В задании 2 учащиеся должны составить собственные предложения со словами, которые тренировались в задании 1. Предпочтительными являются предложения из видеосюжета. Именно они приводятся в качестве возможных ответов при активации функции Sentence, которая по умолчанию деактивирована. Задание 3 содержит 3-4 вопроса на понимание видеосюжета в формате множественного выбора. Ответы даны ниже. Задание 4 предлагает 4 вопроса для обсуждения. 1-й просит кратко сформулировать основную идею видеосюжета (*Please explain what this video is about in a few sentences*), 3 других вопроса побуждают высказать собственное мнение на основе информации видео урока.

Еще одной успешной практикой по интегрированию корпусной лингвистики в обучение ИЯ является платформа *The Corpus-Aided Platform for Language Teachers (CAP)*, созданная на кафедре лингвистики и современных языков в университете Гонконга

23 BNClab tutorial 2018. URL: <https://www.youtube.com/watch?v=28EFVcak99Q&t=18s>

под руководством Ma Qing Angel²⁴. На платформе представлены материалы для повышения квалификации/самообразования преподавателей в области корпусной лингвистики и лингводидактики, небольшие учебные видеоролики продолжительностью 5-7 минут для грамматики, произношения, развития письменных навыков, работе с параллельным корпусом и др. В разделе Teaching Activities помещены примеры учебных материалов /детально разработанных планов уроков продолжительностью 80-120 мин. по категориям: обучение лексике, грамматике, произношению, задания на основе параллельного англо-китайского корпуса. (В каждой категории есть подкатегории, указывающие на категорию обучающихся: младшие школьники, учащиеся средних классов или вузов). Например, целью урока по развитию письменных навыков в формате IELTS является расширение словарного запаса и использование близких синонимов, чтобы уменьшить употребление избыточно используемой лексики (в данном случае слова *people*) с использованием корпуса BNC в коллекции англоязычных корпусов М. Дэвиса. План урока занимает 22 страницы.

Элен Ле Фол вместе со своими студентами-магистрантами, будущими преподавателями иностранного языка, разработала учебные материалы и задания на основе материалов открытых онлайн корпусов по разным изучаемым темам и разместила их на своем сайте²⁵ как ООР. Ресурс создавался в течение 3-х семестров. Важно, что в начале студентам не удалось избежать типичной ошибки: разработанные материалы представляли четкие инструкции по поиску в разных корпусах, а не методически обоснованные планы уроков по использованию корпусов в классе. Эту проблему они решали по ходу своей работы. Материалы разработаны для конкретной категории учащихся: в 1-й части для начальной и средней школы, во 2-й – для старших классов средней школы, в 3-й – для использования разработанных материалов, в 4-й – для английского для специальных целей и профессионально-ориентированного языка. Каждая глава содержит подробную инструкцию поиска материалов в конкретном корпусе для составления конкретного задания, план урока, идеи для последующего развития и применения, выявленные трудности и ограничения [24]. Например, зарегистрированные, но не имеющие подписки пользователи платформы СОСА часто сталкиваются с задержкой в выполнении запроса. Есть свидетельства, что этот факт, может быть, демотивирующим для последующей работы с корпусом у неподготовленных к такому повороту событий учащихся [25]. Мы также сталкивались с этим эффектом и со своей стороны предложили способ его обойти, сократив заложенное время ожидания [26]. Все материалы созданы на основе доступных онлайн корпусов и корпусных инструментов, список которых приведен в приложении пособия²⁶. По мнению разработчиков, это должно способствовать более широкому применению подходов КЛ в повседневной педагогической практике преподавателей ИЯ.

Некоторым аналогом этого проекта является размещенные в разделе сайта НКРИ Studorium результаты проекта «Своевольные смыслы»²⁷, направленного на изучение и описание изменений в значении и употреблении 30 лексем русского языка на протяжении 19-21 вв., выполненный студентами ВШЭ под руководством Н.Р. Добрушиной и М.А. Даниель в 2010 – 2012 гг. Однако формат представления ближе к формату презентаций/научных докладов на конференции, а не к плану урока. НКРЯ содержит «детский» раздел «Верные слова»²⁸ с путеводителем по корпусу «Инструкция по НКРЯ для детей», бесплатным курсом «Как работать в НКРЯ», разнообразными заданиями от игровых для младших школьников до серьезных олимпиадных задач для 13-15-летних

24 The Corpus-Aided Platform for Language Teachers (CAP). URL: <https://corpus.eduhk.hk/cap/>

25 Сайт Элен Ле Фол URL: <https://elenlefol.pressbooks.com/>

26 Приложение пособия Эле Фол URL: <https://elenlefol.pressbooks.com/back-matter/appendix>

27 Своевольные смыслы: опыт микроисторического исследования лексики 19 – 21 веков URL: <https://studiorum.ruscorpora.ru/meanings/>

28 Верные слова URL: <https://vsrussian.com/>

подростков, серий игр для старших школьников и студентов, и очень информативным блогом.

Выводы

Проведенный анализ показал, что тренд на открытое образование, науку и информацию, всемерно поддерживаемый ООН и ЮНЕСКО, может помочь более широкому внедрению методов корпусной лингводидактики в повседневную педагогическую практику преподавателей иностранных языков.

Не все доступные онлайн корпуса и другие ресурсы являются открытыми образовательными ресурсами в строгом смысле этого слова, т.е. обладают базовыми характеристиками ООР, описанными в литературе, но свободный доступ к ним делает их ценными инструментами для самообразования преподавателей ИЯ, желающих повысить свой профессиональный уровень и понять, как можно использовать подходы корпусной лингвистики в своей практике, а также как разрабатывать учебные материалы на основе корпусных данных.

Анализ учебных материалов/планов уроков, созданных как ООР, показал, что ряд вопросов /идей/высказанных предложений требует дальнейшего исследования. Так очень привлекательная идея Н. Вяткиной о разработке учебных материалов в привязке к определенному корпусу, реализована в 4-х из 6 рассмотренных ресурсах: в материалах Вяткиной для DWDS на сайте проекта языкового ресурсного центра канзасского университета <https://corpora.ku.edu> ; в материалах Габласовой по устному компоненту BNC на сайте BNClab, в планах мини-уроков к каждому видеофрагменту на сайте English Central и в разделе «Верные слова» НКРЯ. В двух других случаях – ООР, созданных под руководством Э. Ле Фолл и Ма, содержатся учебные материалы, созданные на базе разных корпусов для решения определенной педагогической задачи с определенной категорией обучающихся.

Требует дальнейшего исследования вопрос о том, что эффективнее и методически целесообразнее: двух страничный план мини-урока с заданиями или план полноценного урока продолжительностью 80-120 минут (как у Ма) объемом в десятки страниц?

Думается, что в созданных ООР недооценивается основной учебник как ресурс/источник для разработки заданий на основе подходов корпусной лингвистики. Авторы анализируют используемые ими основные учебники, доказательно критикуют их, например, за отсутствие примеров аутентичных разговорных речевых образцов и возможностей для исследовательской самостоятельной работы обучающихся [24], и далее дополняют их заданиями без указания (по крайней мере, явно) на то, в каком виде, посредством каких заданий эти темы раскрываются /изучаются в учебнике. Думается, что разработка заданий с использованием подходов корпусной лингвистики для изучения конкретной темы на основе определенного учебника, также может быть весьма эффективной и привести к ряду планов мини-уроков, в качестве дополнения к конкретному учебнику (по модели, реализованной создателями English Central). Эти мини-уроки легко могут быть интегрированы в виде «корпусных пятиминуток» при прохождении самых разных тем, обеспечивая тем самым необходимую гибкость и увеличивая вероятность того, что преподаватели, не знакомые с подходами корпусной лингвистики, начнут их использовать в собственной практике.

Наконец, совершенно очевидно, что создание качественных ООР с использованием корпусов – очень трудоемкий процесс. Идея вовлечения студентов, будущих педагогов, к их разработке в рамках проектной деятельности заслуживает всяческого внимания, поддержки и внедрения в разных учебных заведениях, особенно в условиях, когда Российские стандарты уделяют первостепенное внимание проектной деятельности студентов.

Литература

- [1] Madalli D.P. UNESCO Concepts of Openness and Open Access. Paris. 2015. 67 p.
- [2] Blyth C.S., Thoms J.J. Introduction: Second Language Education as an Open Knowledge Ecology // *Open Education and Second Language Learning and Teaching: The Rise of a New Knowledge Ecology* (C.S. Blyth, J.J. Thomas – eds.). Bristol, Blue Ridge Summit: Multilingual Matters. 2021. P. 1-22. URL: <https://www.degruyter.com/document/doi/10.21832/9781800411005-002/html> (дата обращения: 22.05.2022).
- [3] Kosmas P., Parmaxi A., Perifanou M., Economides A. A. Open Educational Resources for Language Education: Towards the Development of an e-Toolkit // *Learning and Collaboration Technologies: New Challenges and Learning Experiences. HCI 2021. Lecture Notes in Computer Science* (P. Zaphiris, A. Ioannou – eds.). 2021. Vol. 12784. P. 65 – 79. https://doi.org/10.1007/978-3-030-77889-7_5.
- [4] Larsson T. Has ‘the replication crisis’ reached corpus linguistics? // *Blog at WordPress.com*. 2021. URL: <https://linguisticswithacorpus.wordpress.com/2021/11/30/has-the-replication-crisis-reached-corpus-linguistics/> (дата обращения: 22.05.2022).
- [5] Boulton A., Vyatkina N. Thirty years of data-driven learning: Taking stock and charting new directions over time // *Language Learning & Technology*. 2021. Vol. 25. №3. P. 66–89. URL: <https://www.lltjournal.org/item/10125-73450/> (дата обращения: 22.05.2022).
- [6] Boulton A., Cobb T. Corpus use in language learning: A meta-analysis // *Language learning*. 2017. Vol.67, № 2. P.348–393.
- [7] Conrad S. Will corpus linguistics revolutionize grammar teaching in the 21st century? // *TESOL Quarterly*. 2000. Vol.34. № 3. P. 548–560.
- [8] Chambers A. Towards the corpus revolution? Bridging the research–practice gap // *Language Teaching*. 2019. Vol. 52, № 4. P. 460–475. doi:10.1017/S0261444819000089.
- [9] Kogan M., Zakharov V. A way of corpora to secondary school through fun, interest and project work / *Jazykovedný časopis (special issue)*. 2021. Vol.4. (In press).
- [10] Ольховская А.И., Парамонова М.К. Корпус в преподавании русского языка и литературы // *Rhema. Рема*. 2019. № 1. С. 93–123.
- [11] Семенова Н.В. Корпусные технологии изучения русской лингвокультуры в аспекте РКИ // *Мир науки, культуры, образования*. 2022. № 2(93). С. 419-423.
- [12] Semenova N.V., Pulyayevskaya A. M., Polezhayeva Zh. Y. The Role of Corpus technologies in the Study of Russian linguistic and Cultural Vocabulary // *Вестник Пятигорского государственного университета*. 2022. № 1. С. 9-14.
- [13] Reppen R. *Using corpora in the language classroom*. Cambridge, UK: Cambridge University Press. 2010.
- [14] Poole R.A *Guide to Using Corpora For English Language Learners*. Edinburgh University Press Ltd. 2018. 101 p.
- [15] Friginal E. *Corpus Linguistics for English Teachers: New Tools, Online Resources, and Classroom Activities*. New York: Routledge. 2019.
- [16] Shaw E.M. *Teaching vocabulary through data-driven learning*. Provo, UT: Brigham Young University. Master thesis. 2011 URL: <https://scholarsarchive.byu.edu/etd/3024/> (дата обращения: 22.05.2022).
- [17] Vyatkina N. Corpora as open educational resources for language teaching // *Foreign Language Annals*. 2020. Vol. 53. P. 359–370.
- [18] Kogan M., Zakharov V., Popova N., Almazova N. The Impact of Corpus Linguistics on Language Teaching in Russia’s Educational Context: Systematic Literature Review // *LCT. Designing, Developing and Deploying Learning Experiences. HCI 2020. Lecture Notes in Computer Science* (P. Zaphiris, A. Ioannou – eds.). V. 12205. Switzerland: Springer, Cham., 2020. P. 339–355.

- [19]Leńko-Szymańska A. Training teachers in data driven learning: Tackling the challenge // *Language Learning & Technology*. 2017. Vol. 21. № 3. P. 217–241.
- [20]Уайндстейн Ев.В., Коган М.С. Альтернативы массовым открытым он-лайн курсам при интегрировании их в учебный процесс // *Вопросы методики преподавания в вузе*. 2017. Том 6. № 20. С. 19–28.
- [21]Crosthwaite P. Taking DDL online: Designing, implementing and evaluating a SPOC on data-driven learning for tertiary L2 writing // *Australian Review of Applied Linguistics*. 2020. Vol.43. №2. P. 169-195.
- [22]Leńko-Szymańska A. Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education // *ReCALL* 2014. Vol. 26. № 2. P. 260–278.
- [23]Vyatkina N. Language corpora for L2 vocabulary learning: Data-driven learning across the curriculum // *Understanding vocabulary learning and teaching: Implications for language program development* (P. Ecke, S. Rott – eds.). Boston, MA: Cengage Learning. 2018. P. 121 – 145.
- [24]Le Foll E (ed.). *Creating Corpus-Informed Materials for the English as a Foreign Language Classroom: A step-by-step guide for teachers using online resources*. Osnabrück: Pressbooks. 2021. URL: <https://elenlefol.pressbooks.com> (дата обращения: 22.05.2022).
- [25]Ma Q., Yuan R. (E.), Cheung L.M.E., Yang J. Teacher paths for developing corpus-based language pedagogy: a case study // *Computer Assisted Language Learning*. 2022. P. 2-32.
- [26]Кузнецова И.В., Коган М.С. О возможности использования корпуса NOW в курсе английского для специальных целей для студентов специальности «Биотехнология»// *Компьютерная лингвистика и вычислительные онтологии. Выпуск 4 (Труды XXIII Международной объединенной научной конференции «Интернет и современное общество», IMS-2020, Санкт-Петербург, 17 – 20 июня 2020 г. Сборник научных статей)*. СПб.: Университет ИТМО, 2020. С. 29 – 43.

Corpus Linguodidactics in the Context of Open Educational Resources Paradigm

M.S. Kogan

Peter the Great Saint Petersburg Polytechnic University

The development of the Internet and information society led to the trend of open education, science, data and information, which is becoming increasingly popular nowadays. This trend is supported and partially defined by UN and UNESCO documents as well as two World OER Congresses. One of many obstacles preventing widespread using Corpus linguistics approaches in language teaching is the fact that many corpora described in research papers are not available for research and pedagogical community because they are developed for a particular research project. On the other hand, large open online corpora are a valuable resource for linguistic analysis and research but they lack a didactic supplement. But for many language teachers at secondary and tertiary level of education the development of corpus-based teaching materials is a hard, labour- and time- consuming task. A possible solution may be in development of corpus-driven or corpus-based tasks / lessons plans based on and ‘tied to’ a particular corpus. The paper discusses current advanced practices of realization of the approach and discusses the ways of corpus competence of language teachers development based on available open educational resources.

Keywords: open educational resources, OER, online corpora, corpus linguodidactics, didactic materials, pre-service-/in-service- language teachers training

Reference for citation: Kogan M.S. Corpus Linguodidactics in the context of Open Educational Resources Paradigm // Computational Linguistics and Computational Ontologies. Vol.6 ((Proceedings of the XXV International Joint Scientific Conference «Internet and Modern Society», IMS-2022, St. Petersburg, June 23-24, 2022)). - St. Petersburg: ITMO University, 2022. P. 63 – 75. DOI: 10.17586/2541-9781-2022-6-63-75

Reference

- [1] Madalli D. P. UNESCO Concepts of Openness and Open Access. Paris. 2015. 67p.
- [2] Blyth C. S., Thoms J. J. Introduction: Second Language Education as an Open Knowledge Ecology // Open Education and Second Language Learning and Teaching: The Rise of a New Knowledge Ecology (C. S. Blyth, J. J. Thomas – eds.). Bristol, Blue Ridge Summit: Multilingual Matters. 2021. P. 1-22. URL: <https://www.degruyter.com/document/doi/10.21832/9781800411005-002/html> (access date: 22.05.2022).
- [3] Kosmas P., Parmaxi A., Perifanou M., Economides A. A. Open Educational Resources for Language Education: Towards the Development of an e-Toolkit // Learning and Collaboration Technologies: New Challenges and Learning Experiences. HCII 2021. Lecture Notes in Computer Science (P. Zaphiris, A. Ioannou – eds.). 2021. Vol 12784. P. 65 – 79. https://doi.org/10.1007/978-3-030-77889-7_5.
- [4] Draft Recommendation on Open Science. 2021 URL: https://unesdoc.unesco.org/ark:/48223/pf0000378841_rus (access date 22.05.2022).
- [5] Larsson T. Has ‘the replication crisis’ reached corpus linguistics? // Blog at WordPress.com. 2021. URL: <https://linguisticswithacorpus.wordpress.com/2021/11/30/has-the-replication-crisis-reached-corpus-linguistics/> (access date: 22.05.2022).
- [6] Boulton A., Vyatkina N. Thirty years of data-driven learning: Taking stock and charting new directions over time // Language Learning & Technology. 2021. Vol. 25. № 3. P. 66–89. URL: <https://www.lltjournal.org/item/10125-73450/> (access date: 22.05.2022).
- [7] Boulton A., Cobb T. Corpus use in language learning: A meta-analysis // Language learning. 2017. Vol. 67. № 2. P. 348–393.
- [8] Conrad S. Will corpus linguistics revolutionize grammar teaching in the 21st century? // TESOL Quarterly. 2000. Vol. 34. № 3. P. 548–560.
- [9] Chambers A. Towards the corpus revolution? Bridging the research–practice gap // Language Teaching. 2019. Vol. 52. № 4. P. 460–475. doi:10.1017/S0261444819000089.
- [10] Kogan M., Zakharov V. A way of corpora to secondary school through fun, interest and project work / Jazykovedný časopis (special issue). 2021. Vol. 4.
- [11] Olkhovskaya A.I., Paramonova M.K. Corpus in teaching Russian language and literature // Rhema. Pema. 2019. № 1. P. 93–123. (In Russian).
- [12] Semenova N.V. Corpus technologies for studying Russian linguistic culture in the aspect of RFL // Mir nauki, kul'tury, obrazovaniya. 2022. № 2(93). S. 419-423 (In Russian).
- [13] Semenova N. V., Pulyayevskaya A. M., Polezhayeva Zh. Y. The Role of Corpus technologies in the Study of Russian linguistic and Cultural Vocabulary // Pyatigorsk State University Bulletin. 2022. № 1. P. 9-14.
- [14] Reppen R. Using corpora in the language classroom. Cambridge, UK: Cambridge University Press. 2010. 101 p.
- [15] Poole R. A Guide to Using Corpora For English Language Learners. 2018. Edinburgh University Press Ltd. 180 p.
- [16] Friginal E. Corpus Linguistics for English Teachers: New Tools, Online Resources, and Classroom Activities. New York: Routledge. 2019.
- [17] Vyatkina N. Corpora as open educational resources for language teaching // Foreign Language Annals. 2020. Vol. 53. P. 359–370.

- [18]Kogan M., Zakharov V., Popova N., Almazova N. The Impact of Corpus Linguistics on Language Teaching in Russia's Educational Context: Systematic Literature Review // LCT. Designing, Developing and Deploying Learning Experiences. HCII 2020. Lecture Notes in Computer Science (P. Zaphiris, A. Ioannou – eds.). Vol. 12205. Switzerland: Springer, Cham. 2020. P. 339–355.
- [19]Leńko-Szymańska, A. Training teachers in data driven learning: Tackling the challenge // Language Learning & Technology. 2017. Vol. 21. № 3. P. 217–241.
- [20]Kogan M.S., Windstein Ev.V. Alternatives to Massive Open Online Courses from the Perspective of Their Integration into University Programs curricula // Teaching Methodology in Higher Education. 2017. Vol. 6. №. 20. P. 19–28. (In Russian).
- [21]Crosthwaite P. Taking DDL online: Designing, implementing and evaluating a SPOC on data-driven learning for tertiary L2 writing //Australian Review of Applied Linguistics. 2020. Vol.43. №2. P. 169-195.
- [22]Leńko-Szymańska A. Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education // ReCALL 2014. Vol.26. №2. P. 260–278.
- [23]Vyatkina N. Language corpora for L2 vocabulary learning: Data-driven learning across the curriculum // Understanding vocabulary learning and teaching: Implications for language program development (P. Ecke, S. Rott – eds.). Boston, MA: Cengage Learning. 2018. P. 121 – 145.
- [24]Le Foll E (ed.). Creating Corpus-Informed Materials for the English as a Foreign Language Classroom: A step-by-step guide for teachers using online resources. Osnabrück: Pressbooks. 2021. URL: <https://elenlefol.pressbooks.com> (access date: 22.05.2022).
- [25]Ma Q., Yuan R. (E.), Cheung L.M.E., Yang J. Teacher paths for developing corpus-based language pedagogy: a case study // Computer Assisted Language Learning. 2022.
- [26]Kuznetsova I.V., Kogan M.S. On prospects of using the NOW corpus in ESP course for undergraduates majoring in biotechnology // Computational Linguistics and Computational Ontologies.Vol.4 (Proceedings of the XXIII International Joint Scientific Conference «Internet and Modern Society», IMS-2020, St. Petersburg, June 17-20, 2020). St.Petersburg: ITMO University, 2020. P. 29– 43. (In Russian).