

Научная статья

УДК 372.881.111.1

doi: 10.17223/19986645/94/3

Формирование лексического минимума с помощью корпусных инструментов (на материале кейс-исследования по интегрированному курсу Tangible and Intangible Assets)

**Ольга Григорьевна Горина¹, Лариса Эдуардовна Камнева²,
Светлана Николаевна Кучеренко³, Дарья Алексеевна Куганова⁴**

*^{1,2,3,4} Национальный исследовательский университет «Высшая школа экономики»,
Санкт-Петербург, Россия*

¹ gorina@bk.ru

² kamneva.larisa.2706@gmail.com

³ sunflowersnk@gmail.com

⁴ dakuganova@edu.hse.ru

Аннотация. Представлены результаты теоретического анализа и практического использования корпусных методов в отборе лексического минимума для узкоспециализированного курса CLIL в вузе. Освещен дидактически значимый материал по отбору лексики для курса и компиляции авторского корпуса; статистическая процедура сравнения с использованием справочного (БНК)¹ и учебного корпусов. Оценка лингвостатистических показателей проводилась с помощью программы Word Smith 6.0 и специально разработанных программных модулей.

Ключевые слова: корпусный отбор, частотность лексики, принципы отбора, ключевые слова, лексический минимум, корпусный анализ, ESP, CLIL

Благодарности: публикация подготовлена в ходе проведения исследования (№ 23-00-005) в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2022–2023 гг.

Для цитирования: Горина О.Г., Камнева Л.Э., Кучеренко С.Н., Куганова Д.А. Формирование лексического минимума с помощью корпусных инструментов (на материале кейс-исследования по интегрированному курсу Tangible and Intangible Assets) // Вестник Томского государственного университета. Филология. 2025. № 94. С. 43–66. doi: 10.17223/19986645/94/3

¹ БНК – Британский национальный корпус. URL: <https://www.english-corpora.org/bnc/>

Original article

doi: 10.17223/19986645/94/3

The compilation of a lexical minimum for the curriculum subject "Economics of Tangible and Intangible Assets"

**Olga G. Gorina¹, Larisa E. Kamneva², Svetlana N. Kucherenko³,
Darya A. Kuganova⁴**

*^{1, 2, 3, 4} National Research University Higher School of Economics,
St. Petersburg, Russian Federation*

¹ gorina@bk.ru

² kamneva.larisa.2706@gmail.com

³ sunflowersnk@gmail.com

⁴ dakuganova@edu.hse.ru

Abstract. Since the issue of vocabulary build up accompanies the entire process of teaching a foreign language, the proposed research focuses on the prospects of the automatic professional lexis compilation for the "Economics of Tangible and Intangible Assets" course, which is taught in English to second-year undergraduate students of HSE University. The purpose of the corpus-based statistical compilation of lexis is to determine words that are crucial in understanding texts on a given topic. Within this research a corpus of texts relevant to the subject area is compiled; therefore, the texts selected are either professionally-oriented or doubled as syllabus materials for the discipline in question. In order to enhance learners' vocabulary and academic skills, academic periodicals on the topics studied as well as macroeconomic reports are also included in the corpus under research. The British National Corpus (BNC) that consists of 100 million running words acted as a benchmark corpus. The methodology section elaborates on the algorithm's comparison method that defines keywords as unexpectedly frequent in professionally-oriented texts in comparison with their frequency in the BNC. The essence of the keyword extraction procedure, which is carried out by corpus manager Word Smith Tools 6.0, is as follows. Wordlists ordered by frequency are produced with the help of WS Tools 6.0. Then, word frequencies in the benchmark and target corpora are compared to compute keywords. It should be noted that the comparison is based on two statistical procedures implemented in the WS Tools 6.0 algorithm. The article also examines both the reasons for the emergence and the main distinctive features of ESP courses. In addition, the methods and principles of lexis compilation are considered from a diachronic perspective. The article reviews the evolution of approaches to lexical means of communication as well as corpus-driven qualitative and quantitative parameters to compile necessary and sufficient learner's vocabulary. The research also discusses the expediency of investing the efforts of both teachers and students in "random" vocabulary expansion which is not specifically compiled in accordance with the professional interests of university students. It is also noteworthy that apart from lexis compilation the research objectives included devising a set of lexical exercises; the methodologically optimal and justified lexical unit in terms of its presentation to students is determined. Thus, this article examines the linguistic and methodological potential of keywords and frequent content words as a statistical support for the compilation of learner lexical minimum, as well as the optimal ways to present them in the framework of a professionally focused course.

Keywords: corpus-based content, selection principles, keyword computation, lexical minimum, corpus analysis, ESP, CLIL

Acknowledgments: The publication was prepared within the framework of the Academic Fund Program at HSE University in 2022– 2023 (grant No. 23-00-005).

For citation: Gorina, O.G., Kamneva, L.E., Kucherenko, S.N. & Kuganova, D.A. (2025) The compilation of a lexical minimum for the curriculum subject "Economics of Tangible and Intangible Assets". *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*. 94. pp. 43–66. (In Russian). doi: 10.17223/19986645/94/3

1. Введение

Предлагаемое исследование велось в русле разработки и внедрения информационных компьютерных технологий в учебный процесс, что созвучно идеям национальной доктрины развития российского образования, подчеркивающей необходимость регулярного обновления всех аспектов образования, отражающих изменения в культуре, экономике, науке, технике и технологиях.

Современную эпоху исследований часто называют временем «дефисных наук», в рамках которых ведутся гибридные исследования, что обуславливает необходимость пересечения многих областей знаний в постановке задач исследования. В контексте данной работы решения поставленных целей предполагают рассмотрение комбинированных методов корпусной лингвистики и лингводидактики через призму профессионального дискурса.

Теоретическая и методическая значимость работы заключается в том, что с лингводидактической точки зрения исследование проводится в рамках еще пока только складывающегося направления, описываемого не вполне устойчивой коллокацией *корпусная лингводидактика*. Совершенно очевидно, что направление, во-первых, опирается на существующее методологическое и лингвистическое знание, а во-вторых, на аналитический инструментарий корпусной лингвистики, потенциал которого на стыке указанных направлений не реализован полностью. По этой причине нам видится, что «Введение в корпусную лингводидактику» должно стать одним из курсов, читаемых специалистам по филологии и лингвистике. Более того, действуя в рамках этого направления, специалист по языку в неязыковом вузе может не замыкаться на тьюторской позиции, контролируя выполнение упражнений по готовым учебникам, а продолжать исследовательскую деятельность. Вместе с тем, безусловно, с ростом квалификации специалиста, ведущего профессионально направленный курс иностранного языка, улучшается качество отбора компонентов содержания обучения: в частности, научно обоснованно отбираются минимумы – грамматический, лексический, поскольку при конструировании содержания практического курса языка непосредственному отбору подлежат именно минимумы, призванные обслужить тот или иной вид вербальной деятельности [1].

Отметим, что практика применения корпуса в лингводидактике имеет два вектора ее развития. Первое направление связано с информированием

составителей учебников, словарей и другой учебной продукции о результатах корпусных исследований. В рамках второго направления ведется разработка аудиторных приемов работы с опорой на корпус, т.е. непосредственное использование корпусных материалов на занятиях. Как правило, крупные издательства и научные коллективы, реализующие большие проекты, подобные созданию Лонгмановской корпусной грамматики английского языка (Longman Grammar of Spoken and Written English), работают в рамках первого направления.

Однако существуют важные ниши и для маленьких университетских лабораторий, которые ведут более узкоспециализированные исследования. В частности, наше исследование обратило внимание на важную задачу по отбору лексического минимума по узкопрофессиональному курсу, который читается студентам второго курса вуза на английском языке. Направление, в котором ведется преподавание, в широком смысле, может относиться к области ESP (*English for Specific Purposes, английский для специальных целей*); вместе с тем по признаку наполненности профессиональным содержанием на иностранном языке данный курс читается и в русле CLIL (*Content and Language Integrated Learning, интегрированное обучение содержанию и языку*). Возникшее в ответ на потребности в обучении языку, которое одновременно способствует усвоению предметных знаний, направление предоставляет студентам возможность изучать предметы на иностранном языке, т.е. понимать материал и осваивать его на двух уровнях – языковом и предметном.

Перед специалистом, читающим подобный курс, стоят две основные задачи: (i) разработка учебных материалов, так как преподавателю CLIL часто приходится разрабатывать или адаптировать учебные материалы, чтобы они соответствовали обоим аспектам – содержанию предмета и языковому уровню студентов; (ii) интеграция учебных областей: специалист работает над интеграцией языкового обучения и обучения предметам, помогая студентам углубленно изучать и понимать иностранный язык через предметный контекст. В этих условиях как для наполнения содержания обучения, так и для совершенствования в преподаваемой предметной области специалист университета, занимающий «двойную» позицию специалиста по языку и специалиста в профессиональном контексте, особенно нуждается в инструментах исследования, подобных корпусным.

Корпусный материал и инструментарий могут использоваться как источник для создания учебных материалов, отбора языковых фактов, уточнения знаний об изучаемом языке в заданной профессиональной области. В частности, выделение ключевых слов в их статистическом понимании с помощью коллекции профессионально ориентированных текстов [2] помогает отобрать необходимые и достаточные лексические или более емкие лексикограмматические единицы для понимания профессионального контекста и получения ожидаемого качества речи студента. Несмотря на то, что в силу особенностей формата публикации в научном журнале данная статья преимущественно отразила опыт отбора лексического минимума, собранный лингвистический ресурс может выступать в качестве учебных материалов.

Таким образом, тематика, затронутая в статье, может представлять интерес для профессионального круга читателей как с точки зрения знакомства с практически реализованным корпусным проектом по отбору лексического минимума для профессионального курса, так и с точки зрения дальнейшего теоретико-практического развития направления.

1.1. К вопросу о компиляции лексического минимума в контексте ESP и CLIL

Диахронически отбор профессионально ориентированного словарного запаса под конкретный курс обучения связан с возникновением практики преподавания иностранного языка для специальных целей (English for Specific Purposes, ESP). Авторы этой идеи в западной методической традиции А. Уотерс и Т. Хатчинсон [3] отмечают три основных фактора, благодаря которым произошло выделение языка для специальных целей. Во-первых, в результате расширения коммерческих и производственных международных связей впервые образовалась категория таких обучающихся, которые четко представляли, с какой целью они изучают английский язык. Как следствие, возникла необходимость в создании сжатых по времени, высокоэффективных курсов обучения с четко определенными целями. Во-вторых, произошел сдвиг направления исследований в самой лингвистике от изучения черт языковой системы к изучению употребления языка в аутентичной коммуникации. Вместе с тем обнаружились существенные расхождения в письменной и устной языковой норме и влияние контекста, что привело к мысли о том, что и язык разных профессий может существенно варьировать. Решение лежало в области выделения лингвистических особенностей ситуаций профессионального общения и построения специализированных курсов с учетом этих особенностей. В-третьих, развитие педагогической психологии привело к осознанию необходимости сдвига фокуса внимания на обучающегося и его интересы, что усилило осознание необходимости профессиональной направленности учебных материалов, которые должны были отражать сферу интересов обучающихся, а именно профессиональную сферу [3].

CLIL или интегрирование контента специальности в курс иностранного языка является дальнейшим развитием идеи переноса фокуса на профессиональные нужды студентов. Преподавателю в русле CLIL необходимо хорошо понимать сам предмет, который он преподает, и уметь связывать содержание предмета с языковыми навыками. Он должен уметь выбирать подходящие учебные материалы, а также иметь компетенции по их разработке. Следует подчеркнуть, что нечасто преподаватель иностранного языка параллельно является еще и специалистом в той области, для которой ведется преподавание языка, или состоит в команде опытных ESP/CLIL-экспертов. Большинству приходится работать достаточно автономно и в конкурентной среде, самостоятельно исследовать идеи для разработки курсов и учебных материалов. Одним из эффективных путей решения проблем обучения профессиональному иностранному языку и контенту профессии может быть

сбор релевантного для курса лингвистического материала, компиляция корпуса, его статистическое исследование и использование.

Что касается самих обучаемых, то сегодня в университетской среде на уровне бакалавриата студенты, как правило, относятся к не имеющим опыта производственных отношений и опыта работы по специальности потенциальным специалистам в определенной области, которым необходимо сочетание английского языка для академических и профессиональных целей на узкоспециализированном контексте. На этом фоне специалисты видят свою задачу в том, чтобы обеспечить отбор релевантного содержания обучения, выраженного в том числе в языковых минимумах.

При обсуждении проблем отбора материала для курса профессионально ориентированного иностранного языка неизбежно возникает необходимость уточнения таких понятий, как языковой минимум, а также принципы и критерии его отбора, которые в отечественной методике были определены довольно давно¹. Под минимумом предлагается понимать «достаточную для обеспечения задач курса в области данного вида речевой деятельности, особым образом отбираемую и обязательную для усвоения совокупности языковых явлений (или паралингвистических, в широком смысле, явлений, или системных знаний об изучаемом языке, или фоновых знаний), каждое из которых предназначено для использования в предусмотренном программой общении и требует либо введения и отработки в упражнениях, либо только введения (демонстрации)» [1. С. 23]. На наш взгляд, данная формулировка обладает необходимой емкостью и эластичностью, отразив основные черты эффективного минимума: (i) учет специфики минимумов в области обучения продуцированию речи и в области рецепции речи, т.е. пассивной и активной лексики; (ii) исключение из минимума промежуточных навыков, умений и сведений, которые не предусмотрены целями обучения; (iii) специальный отбор минимума, т.е. не случайный характер в выборе необходимого и достаточного словарного запаса; (iv) обязательность минимума как набора наиболее важных в коммуникативном отношении единиц и необходимость отработки материала, входящего в минимум [1].

Отсюда вытекает на первый взгляд спорное требование, чтобы отбор минимума был строгим и отличался скорее недостатком слов, чем избытком. Однако не следует рассматривать такое требование как ограничение стремления обучающегося знать больше. По этой причине всячески поощряется возможность слушателя дисциплины развиваться автономно, обогащать

¹ Справедливости ради, надо отметить, что в трудах Л.В. Щербы было предложено различение активного и пассивного словаря, активной и пассивной грамматики. Методически оправданная идея различения активного и пассивного языкового материала звучала уже в работах Г. Пальмера и М. Уэста. Идея разграничения обязательного сектора и свободного сектора при говорении высказывалась в работах М. Уэста. В трудах этих ученых была также сформулирована необходимость научного отбора лексики и сформированы первые научно обоснованные лексические минимумы.

лексический запас; для этой цели у обучаемого должен быть свободный доступ к факультативным учебным материалам (в нашем случае к корпусу текстов), расширенному лексическому списку, коллокациям, коллигациям и упражнениям.

Ряд вопросов отбора минимумов для любых целей, в частности для таких новых и еще не устоявшихся, а иногда и часто обновляющихся специальных курсов, относятся к числу современных проблем лингводидактического обеспечения университетского обучения. Итак, при отборе лексических минимумов возникает целый ряд переменных параметров, которые необходимо учитывать. В первую очередь это вопросы посильного количества, которое можно отнести к общедидактическому требованию, а также определения исходных положений отбора, его принципов и критериев. Также необходима надежная, выверенная процедура их реализации, отвечающая современному состоянию как лингвистики и лингвотехнологий, так и быстро меняющемуся контенту дисциплины, которую усваивает будущий профессионал. Следует также принимать во внимание, что во многих современных курсах обучения существует опора на научную периодику, иными словами, минимум по дисциплине должен подготовить обучаемого к восприятию самых современных публикаций на иностранном языке, издаваемых в ведущих академических журналах. Кроме того, существует запрос на продуцирование академических текстов на английском языке для дальнейшей публикации в международных рецензируемых изданиях, индексируемых в таких наукометрических базах, как Scopus, Web of Science.

Обратившись к истории вопроса об отборе вокабуляра, интересно отметить, что английский методист и лексикограф Г. Пальмер одним из первых поставил вопрос о научном отборе словаря, подразделив учебный словарь на две большие группы: «строго отобранный материал, который он называл микрокосм (the microcosm), и стихийный» [4. С. 51]. Первый изучался систематически на начальной и средней ступенях обучения, а второй накапливался стихийно на продвинутом уровне [4]. Подчеркнем, что слово «microcosm» означает «что-то в миниатюре», «отражение большого в малом», притом что это малое имеет все основные свойства большого, например microcosm of society – отражение общества, общество в миниатюре. Сказанное означает, что британский методист уже тогда ставил задачи сегодняшней корпусной лингвистики – репрезентации, «адекватного отражения, адаптации или интеграции больших массивов текстов или некоторых иных фрагментов речевой деятельности в существенно меньший по объему корпус текстов» [5]. По замыслу Г. Пальмера, первая группа словаря под названием «microcosm» изучалась систематически на ранних стадиях, а вторая – накапливалась стихийно уже на продвинутом уровне. Автор идеи пользовался термином «лексическая единица», поскольку считал термин «слово» слишком неопределенным, и рекомендовал следующие принципы отбора лексических единиц: (i) опора на частотность (frequency), при этом разные значения одной и той же единицы должны учитываться отдельно; (ii) структурная сочетаемость (ergonic combination), т.е. способность лексической

единицы (эргона) сочетаться с другими единицами; (iii) конкретность (concreteness), или приоритет словам конкретного значения; (iv) пропорциональность (proportion), т.е. соблюдение известной пропорциональности, как между отдельными частями речи, так и в отношении равномерного изучения различных аспектов языка; (v) целесообразность (general evidence), в рамках которой могут быть нарушены первые четыре основных принципа и в микрокосм могут включаться слова, принадлежащие к одной семантической группе [6].

В отечественной методике ранее всего лингвистические принципы отбора были разработаны Л.В. Щербой, И.В. Рахмановым [7]. На их основе были составлены словари наиболее употребительных слов (4 500 слов), в которых учитывались семь принципов: сочетаемости, стилистической неограниченности, семантической и словообразовательной ценности, многозначности, строевой способности и частотности [7]. Исследователи часто обращаются к принципам отбора активной лексики, сформулированным С.Ф. Шатиловым, которые уточняются в ходе конкретных исследований и с учетом задач обучения. С.Ф. Шатилов предлагал руководствоваться следующими принципами: (i) сочетаемости, согласно которому слова с большей сочетаемостью предпочтительнее слов с редкой сочетаемостью; (ii) семантики, согласно которому отбираемые слова должны выражать наиболее важные понятия по той тематике (устной и письменной речи), с которой встречается учащийся, изучая иностранный язык; (iii) стилистической неограниченности, т.е. принадлежности слова к нейтральному, литературному, разговорному и книжно-письменному стилям языка; (iv) частотности, в соответствии с которым в словарь-минимум включаются наиболее употребительные и литературно-разговорные слова; (v) исключения синонимов, означающий, что в словарь-минимум из синонимического ряда включается только одно наиболее употребительное и нейтральное слово синонимического ряда; (vi) словообразовательной ценности, предполагающий включение только наиболее употребительных в словообразовательном отношении слов; (vii) исключения интернациональных слов, полностью совпадающих в родном и иностранном языке [8].

Несмотря на то, что корпусные технологии существенно упростили определение абсолютной частотности и позволяют обработать не только большие массивы текстов, но и, как будет рассмотрено далее, учесть более сложные производные частотных характеристик слов, изложенные основоположниками принципы отбора лексического минимума не утратили своей актуальности и разумности. Говоря о более современных лингвотехнологиях, отметим, что корпусные процедуры помогают значительно быстрее трансформировать тексты по профессиональной тематике в компонент содержания обучения, который будет подлежать контролю.

Завершая обзор истории развития принципов отбора лексики, обратимся к одной из масштабных инициатив последних лет – созданию лексического минимума для всех уровней (A1–C2) общеупотребительной лексики ан-

глийского языка команды британских исследователей. По результатам работы, во-первых, словарный минимум был определен количественно для всех указанных уровней. Так, для уровня владения иностранным языком B2 в классификации европейских требований установлен объем необходимого словаря в количестве 4 000 отдельных слов и, соответственно, 8 200 слов с учетом словообразовательных моделей. Во-вторых, перечень отобранных словарных единиц представлен в виде списка словарных семей (гнезд) по алфавиту и находится в свободном доступе. Авторы проекта [9] указывают, что они остановились на пороге профессионально направленного обучения английскому языку, что представляется хорошим стартом и ориентиром для всех преподавателей, которые занимаются далее развитием лингвистической составляющей профессионально ориентированной иноязычной компетенции. Вместе с тем исходные положения, принципы и критерии отбора, разработанные в рамках проекта, представляют собой самый современный подход к отбору лексического минимума как с идеологической, так и с практической точек зрения.

1.2. Количественные корпусные ориентиры лексического минимума в русле корпусных исследований

Несмотря на то, что на первый взгляд корпусные данные поменяли лишь техническую часть отбора лексического минимума в части нетрудоемкого определения частотности, на практике корпусные исследования оказали влияние как на качественные, так и на количественные параметры словарного запаса обучаемого. Следует отметить, что количественные характеристики словарных минимумов по разным корпусно-ориентированным источникам согласуются и останавливаются на пороге в 5–6 тысяч слов общеупотребительной лексики. Начиная с этого количества слов дальнейшее инвестирование усилий по наращиванию запаса непрофессиональной лексики не рекомендуется; усилия преподавателей и обучаемых следует направить в русло специализированной лексики [10, 11].

Говоря о том количестве слов, которое способен усвоить обучаемый, можно обратиться к данным по серии известных лексических учебных пособий «Vocabulary in Use», созданных авторами М. Маккарти и Ф. О'Делл. Учебники рассчитаны на усвоение 2 000 слов на каждом уровне начиная с Elementary до Upper Intermediate. Если обучающийся находится в среде носителей, в стране изучаемого языка, то вполне реалистичным признается усвоение 2 500 слов в год [12]. Таким образом, исходя из корпусных данных, пилотных проектов, оценок специалистов и обучаемых, цель овладения лексиконом в 2 000 слов на каждом уровне определяется как достижимая.

Еще одна особенность наращивания словарного запаса связана с тем фактом, что язык часто определяют как «набор редких событий». Действительно, корпусная статистика и теоретические выводы Gorina et al. [13] доказывают, что 50% словаря любого корпуса текстов имеют единичное вхождение. Другими словами, половина слов любого корпуса составляют группу

гапаксов («лишь раз названное»), т. е. являются сравнительно редкими, или низкочастотными словами. В этой связи следует привести такие данные: увеличение словарного запаса обучаемого уровня В2 с 6 до 10 тысяч увеличивает шансы встретить знакомое слово лишь на 3%, с 10 тысяч. слов до 16 тысяч – на 2%.

Что касается профессионально направленного минимума, то к сопоставимым количественным показателям пришли составители минимума лексических единиц русского языка. Власова и др. занимались решением вопроса о минимальном объеме вокабуляра, необходимого для понимания дискурса предметной области. В своем исследовании авторы также применяют лингвостатистический подход, собрав корпус по политологии на русском языке. Исследователи сгенерировали частотные списки, которые состоят из основного перечня знаменательных слов, служебных слов, редких слов и имен собственных, сделав вывод о том, что оптимальная длина лексического минимума не должна превышать 4 500 слов. Отмечается, что именно знание знаменательных слов определяет уровень понимания текстов в исследуемой предметной области (1 000–2 000 лексемы необходимо знать, чтобы понимать 50–60% материала профессионального корпуса) [14. С. 73].

Таким образом, процедура отбора словаря, которая позволяет определить типичные для профессионального дискурса лексические единицы и увеличить шансы обучаемого встретить изученное слово, необходима и оправданна. Однако в процессе составления лексических минимумов, или списков слов, возникает сложность в определении объема вокабуляра, уже известного студенту, и какой объем необходимо изучить для того, чтобы понимать академические и профессионально ориентированные тексты.

1.3. К вопросу об отборе текстов в корпус предметной области

Важной особенностью совокупности языковых данных, организованной в корпус, является идея, в соответствии с которой собираются эти данные. Так, в названии «Британский национальный корпус» слово «национальный» означает в первую очередь «характеризующий британский национальный вариант английского языка», т.е. идея составления корпуса заключалась в максимально полной репрезентации этого варианта языка [15].

Следует еще раз обратить внимание на то, что компиляция корпуса, во-первых, подчиняется определенной концепции. Таким образом, при компиляции корпуса нам следует ответить на вопрос: что будет отражать или репрезентировать наш корпус? Во-вторых, для того чтобы просто коллекция текстов стала корпусом, необходим филологически компетентный корпусный менеджер, который ответит на вопрос, как будет происходить подсчет и автоматический вывод результатов. По этим двум причинам тексты и систему поиска данных, например, в Интернете нельзя признать профессионально ориентированным корпусом. Тексты Интернета ориентированы на интересы пользователей Интернета, а не на профессиональные интересы

обучаемых. Система поиска и вывода данных тоже не создавалась для решения лингвистических и методических задач. Таким образом, мы подошли к определению методически-ориентированного корпуса предметной области: это такой корпус, тексты которого должны адекватно отражать предметную область, а автоматический подсчет (лингвостатистика), поиск и вывод информации должны позволять решать задачи изучения профессионального дискурса в целях обучения иностранному языку. Обратимся к принципам отбора текстов в методически-ориентированный корпус предметной области.

2. Создание собственного методически-ориентированного корпуса

В данной части работы рассматриваются принципы отбора текстов в методически-ориентированный корпус предметной области «Tangible and Intangible Economy» (Материальная и нематериальная экономика). Подбор текстов осуществлялся на основе четко сформулированных критериев: общенаучных критериев отбора исследовательского материала [16], общих критериев составления корпусов и узконаправленных критериев, соответствующих задачам данного исследования [17]. Для наших исследовательских задач при составлении корпуса наиболее ценными мы признаем следующие критерии:

- 1) критерий репрезентативности;
- 2) критерий аутентичности;
- 3) критерий соответствия проблемной области;
- 4) критерий лексической насыщенности;
- 5) критерий новизны.

Ниже приводится описание методики создания корпуса в соответствии с каждым из этих критериев.

Основное условие, регулирующее нашу деятельность в рамках отбора текстов в корпус, это критерий репрезентативности корпуса, справедливо рассматриваемый В.П. Захаровым в качестве основной характеристики любого корпуса, независимо от других свойств корпуса [18]. Репрезентативность, по мнению В.В. Рыкова – это уникальное свойство корпуса, которое отличает корпус от любой другой коллекции текстов, доступной на электронном или ином носителе [19]. Несмотря на невозможность (по крайней мере на данный момент) точно измерить репрезентативность того или иного корпуса, под репрезентативностью принято понимать необходимое и пропорциональное распределение максимального количества свойств изучаемого явления для того, чтобы отдельно взятый корпус можно было рассматривать в качестве модели предмета исследования [18]. Очевидно, что необходимое количество свойств будет отличаться в зависимости от задач исследования, однако требование максимально объективно представить в корпусе свойства предмета исследования остается неизменным и в случае создания крупных национальных корпусов и специализированных корпусов.

Для достижения репрезентативности большие национальные корпуса языков, как правило, включают в себя от 22 млн до 1 млрд лемматизированных словоформ. Так, по статистическим данным М.В. Копотева [18], корпуса наиболее распространенных романо-германских языков (английский, немецкий, французский, итальянский и испанский языки) варьируются в объеме от 100 млн токенов (для корпусов английского и испанского языков) до 8,7 млрд для корпуса немецкого языка. Кроме того, для достижения максимальной степени репрезентативности в них включены тексты разных жанров и типов. Например, общеизвестный факт, что Брауновский корпус письменных текстов включает тексты 15 жанров [20], в то время как Лонгмановский корпус (the LSWE Corpus) был призван репрезентировать письменную и устную английскую речь и служил аутентичным материалом для примеров в принципиально новой книге по грамматике устного и письменного английского языка [21].

Специализированные корпуса, создаваемые для решения узких исследовательских задач, составлялись и составляются на основе гораздо меньшего количества токенов. Например, корпус ошибок обучаемых, собранный в РГПУ им А.И. Герцена, включал 38 122 словоупотребления на момент его издания [22]. Очевидно, что в случае специализированных корпусов критерий репрезентативности трактуется несколько иначе в силу их небольшого объема. Репрезентативность в данном случае – это «необходимо-достаточное количество текстов, обеспечивающих решение исследовательских задач» [23. С. 164]. Таким образом, соблюдение критерия репрезентативности при составлении нашего специализированного корпуса достигалась за счет решения исследовательских задач, описываемых ниже.

Прежде всего следует отметить, что наш корпус, наряду с другими исследовательскими и прикладными задачами, создавался для обеспечения эффективности учебного процесса в ходе преподавания дисциплины «Tangible and Intangible Economy» (Материальная и нематериальная экономика) в рамках учебного плана дисциплин студентов второго курса факультета ШЭМ НИУ ВШЭ Санкт-Петербург. С одной стороны, преподавание дисциплины обусловлено такой целью, как развитие умений профессиональной речи на английском языке за счет расширения понятийного аппарата по специальности и развития всех видов деятельности на английском языке [24]; с другой стороны, преподавание дисциплины осуществляется в соответствии с «Концепцией развития англоязычной коммуникативной компетенции студентов в НИУ ВШЭ», согласно которой преподавание английского языка для специальных целей направлено на «формирование и развитие академических умений <...> с учетом специфики конкретной предметной области» и ведется на основе специализированного текстового материала», носящего аутентичный характер и отражающего последние достижения предметной области [25. С. 4]. Необходимость привлечения корпусной лингвистики для обеспечения эффективности учебного процесса в частности и самого учебного процесса в целом обусловлена отсутствием до-

статочного количества доступных аутентичных учебных материалов, отражающих новые дискурсивные и нарративные явления в языке современной экономической науки, такие как дарообмен, шэринг, экономика идей, экономика контента, креативная экономика [26]. Соответственно, некоторые из нижерассматриваемых критериев изначально работают тем или иным образом, в зависимости от ограничений, накладываемых содержанием Концепции и наполнением рабочего учебного плана.

Критерий сбалансированности часто рассматривается вторым ведущим критерием при составлении специализированных корпусов [27] и достигается за счет жанрового разнообразия и объема текстового материала. В нашем случае специфика исследовательской задачи накладывает определенные ограничения на фактор жанрового разнообразия, так как жанр текстового материала определяется требованиями программы учебной дисциплины и Концепции, предполагающими формирование лексических навыков на основе профессиональных академических текстов. Источниками таких текстов являются журналы экономической направленности первого, второго и третьего квартилей, доступ к которым обеспечивает электронная подписка НИУ ВШЭ. Среди журналов можно назвать следующие: *Review of Economic Studies*, *American Economic Review*, *Journal of Public Economics*, *Journal of International Economics*, *Research in Economics*.

Немаловажным при составлении специализированных корпусов, ориентированных на определенную предметную область, является принцип аутентичности. Во-первых, центральная идея создания и обработки корпусов текстов заключается в обработке именно аутентичных массивов речевой продукции [28. С. 26]. Во-вторых, формирование уровня коммуникативной компетенции по английскому языку, обозначенного в Концепции НИУ ВШЭ, невозможно без аутентичных материалов. Термин «аутентичность» не имеет четких терминологических границ ни в отечественной, ни в зарубежной науке [29]. В своей работе мы опираемся на ставшие классическими подходы к пониманию аутентичных текстов отечественных лингвистов Р.П. Мильруда и Г.И. Ворониной как текстов, чье языковое наполнение не было специально придумано для прикладных целей изучения языка [30] и чье содержание было заимствовано из коммуникативной практики носителей языка [31]. Именно поэтому, наш корпус составлялся исключительно из академических статей, опубликованных в ведущих экономических журналах, гарантирующих высокое качество материалов с точки зрения английского языка.

Помимо универсальных принципов репрезентативности, сбалансированности, аутентичности необходимо учитывать также принципы соответствия проблемной области. Непосредственно в контексте отбора специализированного корпуса по дисциплине «Tangible and Intangible Economy» (Материальная и нематериальная экономика) суть этого принципа состоит в «определении границ языкового пространства проблемной области, которое планируется отобразить в корпусе» [32. С. 63]. Языковое пространство обу-

словлено содержанием дисциплины, а именно такими подтемами, как экономика обмена, экономика контента, цифровая экономика, нематериальная экономика, устойчивая экономика, экспериментальная и поведенческая экономика, роль государства в экономике, общественные блага, налоговая система, экономика неравенства, экономика образования, экономика здравоохранения, новейшие экономические тренды.

Принципы аутентичности и соответствия проблемной области в определенной степени обуславливают необходимость учитывать принцип лексической насыщенности: профессионально ориентированный языковой материал, созданный носителями английского языка, должен быть не только заимствован из коммуникативной практики английского языка, он должен адекватно отражать профессиональный аспект такой коммуникативной практики в терминах лексического разнообразия, лексической сложности и лексической плотности [33]. Понятие лексической насыщенности аутентичных материалов само по себе является объектом многочисленных исследований и требует уточнения и разработки инструментария для измерения ее составляющих. Соблюдение принципа лексической насыщенности достигалось за счет опоры на экспертные рекомендации специалистов НИУ ВШЭ, занимающихся экономикой материальных и нематериальных активов, которые рекомендовали академические статьи и конкретных авторов для профессионального обогащения языкового материала.

Таблица 1

Характеристика корпуса

Состав	Состав корпуса
Цель	Специализированный
Назначение	Методически-ориентированный
Объем	1 100 000 словоупотреблений
Сбалансированность	Несбалансированный
Тип языковых данных	Письменные тексты
Жанр	Научный
Тематика	Экономика материальных и нематериальных активов
Новизна	2005–2022
Количество текстов	159 файлов
Разметка	Неразмеченный
Доступность	Ограничен доступ [материалы находятся в электронной библиотеке НИУ ВШЭ и на базах данных / репозиториях]

Одним из формальных конструирующих принципов нашего корпуса является принцип новизны, который также ограничен рекомендациями рабочей учебной программы включать в учебный процесс материалы не старше пяти лет. В чистом виде данное ограничение неприемлемо для предметной области нашего корпуса, так как оно повлечет изменение степени репрезентативности корпуса. Соответственно, было принято исследовательское решение углубить новизну корпуса до 2005 г. Такой выбор временной границы обусловлен тем, что примерно с этого года начинается постепенное

смещение мировой экономики на принципы совместного потребления и рост доли нематериальных активов, которые определяют характер современной экономики [34]. Новая экономика, основанная на нематериальных активах, и является той узкой предметной областью, на которой построена дисциплина «Tangible and Intangible Economy» (Материальная и нематериальная экономика).

На основании вышеперечисленных принципов нами был составлен корпус, репрезентирующий предметную область «Tangible and Intangible Economy» (Материальная и нематериальная экономика) для выделения содержательного лексического ядра путем статистической обработки (см. табл. 1).

3. Методика обработки корпуса

В соответствии с нашей задачей – отбор статистически значимой лексики – необходимо описать процедуру и принципы такого отбора. Из всего многообразия корпусных инструментов, используемых для данной исследовательской задачи, мы выбрали технологию выделения ключевых слов, встроенную в корпусный менеджер WordSmith 6.0. Ключевые слова вычисляются путем сравнения частотности слова в исследуемом корпусе (небольшом) с частотностью слова в корпусе справочном (большом). На основании данных справочного корпуса вычисляется математическое ожидание – как часто слово вообще в языке, поэтому справочный корпус должен быть максимально большим. Если слово в исследовательском корпусе попадает чаще, оно становится кандидатом на статус ключевого: его частотность не совпадает с ожидаемой. В нашем случае исследуемый корпус – это специальный корпус по предметной области Экономика материальных и нематериальных активов, характеристики которого были описаны выше, и Британский Национальный Корпус, который имеется в нашем распоряжении.

Перед обработкой нашего корпуса программой WordSmith 6.0 все тексты были переведены из формата .doc в формат .txt блокнот. Далее текстовые файлы были загружены в интерфейс программы WordSmith 6.0 для вычисления. Для подсчета величины «ключевого характера» (keyness) программа обрабатывает четыре значения: количество вхождений (частотность) искомого слова в исследуемом тексте (корпусе), количество вхождений (частотность) исследуемого слова в опорном корпусе, количество всех слов в исследуемом тексте (корпусе) и, наконец, количество всех слов в опорном корпусе. Результаты были представлены в виде таблиц с указанием частотности слов и сопутствующих вычислениям параметров. Приводим фрагмент рабочей таблицы, в которой отражен процесс вычислений и даны следующие значения: (i) ранговый номер слова в списке ключевых слов, упорядоченном по свойству быть ключевым; (ii) ключевое слово; (iii) абсолютная частота слова в исследуемом корпусе; (iv) относительная частота слова в исследуемом корпусе; (v) «RC. Freq.» – абсолютная частота вхождения слова в справочный корпус, в нашем случае БНК; (vi) «RC. %» – относительная частота

вхождения слова в справочный корпус в процентах¹; (vii) значение Keynes, т.е. параметра, по которому слово признается ключевым, а именно результат вычисления параметра в зависимости от метода, который используется в соответствии с алгоритмом: хи-квадрат или логарифмическое подобие (chi-square/log likelihood statistic).

Таблица 2

Результаты применения алгоритма WordSmith по определению ключевых слов

N (i)	Key word (ii)	Freq. (iii)	% (iv)	RC. Freq. (v)	RC. % (vi)	Keyness (vii)
3	INTANGIBLE	698	0,830151856	180		8993,644531
5	CAPITAL	574	0,682675064	0		8127,345215
8	ASSETS	445	0,529251575	0		6300,131348
20	SHARE	184	0,218836591	0		2604,426025
21	ACCOUNTING	162	0,192671359	0		2292,984863
30	INTANGIBLES	141	0,167695433	37		1813,81665
35	PERFORMANCE	302	0,35917747	13218	0,013295413	1404,041382
63	VARIABLES	59	0,070170432	0		835,0271606
68	GIG	93	0,110607632	585		775,1405029
73	APPROVAL	52	0,061845124	0		735,9517822
79	ENTITY	96	0,114175618	1196		676,9440918
87	COMPETENCIES	58	0,068981104	83		630,0023193
92	EXPENDITURES	42	0,049951833	0		594,4176025
94	MILESTONE	41	0,048762504	0		580,2643433
110	STAKEHOLDERS	36	0,042815857	0		509,4982605
112	HERITAGE	85	0,101092994	1888		505,3440552
115	BOARD	35	0,041626528	0		495,3450928
126	COMPETITIVE	98	0,116554275	3786		478,715271
134	IMMATERIAL	49	0,058277138	190		451,3303223
140	ACCOUNT	31	0,036869209	0		438,7327576
141	INTERNALLY	59	0,070170432	603		438,1759338
153	OUTCOMES	29	0,034490552	0		410,4267273
167	AWARENESS	26	0,030922562	0		367,967865
176	EXPENSES	25	0,029733233	0		353,8149719
189	MEASUREMENT	61	0,07254909	1668		338,3108215
197	CAUSALITY	34	0,040437199	103		327,8379822
200	ACQUIRED	23	0,027354574	0		325,5092163
238	OBJECTIVES	20	0,023786588	0		283,0507813
242	PROPERTIES	67	0,079685062	4061		270,0170898
248	MANAGERIAL	19	0,022597257	0		268,8980103
264	CAPACITY	18	0,021407928	0		254,7452698
269	SUSTAINABILITY	28	0,033301223	150		241,6063843
274	ACQUISITION	17	0,0202186	0		240,5925598
276	STRENGTHEN	17	0,0202186	0		240,5925598
278	VARIANCE	17	0,0202186	0		240,5925598

¹ В нашем случае это значение нередко нулевое, т.е. слово очень низкочастотно в справочном корпусе; в приведенном фрагменте таблицы исключение составляет слово PERFORMANCE, которое встречается в БНК 13 тысяч раз.

N (i)	Key word (ii)	Freq. (iii)	% (iv)	RC. Freq. (v)	RC. % (vi)	Keyness (vii)
287	IMPLICATIONS	16	0,019029269	0		226,4398651
289	COMPARABILITY	16	0,019029269	0		226,4398651
304	GRANGER	20	0,023786588	29		216,8338623
309	ASSESS	15	0,01783994	0		212,2872009
317	AMORTIZED	15	0,01783994	0		212,2872009
318	REGULATIONS	15	0,01783994	0		212,2872009
348	RESPONSES	14	0,016650612	0		198,134552
349	COMPETITIVENESS	31	0,036869209	560		196,5691071
396	INTENSITY	12	0,014271952	0		169,8293304
402	IMPROVEMENTS	12	0,014271952	0		169,8293304
412	REGRESSION	28	0,033301223	666		162,7696381
414	GOODWILL	29	0,034490552	793		160,8246155
416	PATENTS	24	0,028543904	373		159,0912323
432	SHAREHOLDERS	11	0,013082623	0		155,6767578
439	ACCUMULATED	11	0,013082623	0		155,6767578
449	SUBSIDIES	29	0,034490552	1116		141,8536377
451	AGENDA	10	0,011893294	0		141,5242004
453	IDENTITY	10	0,011893294	0		141,5242004
456	CAPITA	10	0,011893294	0		141,5242004
459	ASSESSMENT	10	0,011893294	0		141,5242004
464	BREAKPOINTS	10	0,011893294	0		141,5242004
470	REPUTATIONAL	10	0,011893294	0		141,5242004
472	CONCERNING	10	0,011893294	0		141,5242004
481	CONTRIBUTION	48	0,057087809	5268		140,7761688
485	DOMESTIC	52	0,061845124	6900		134,8648682

4. Результаты

В рамках данного исследования был проведен корпусный отбор лексического материала, который важен для обучаемого с точки зрения понимания текстов по дисциплине ESP экономического направления. В ходе разработки курса первая порция лексических единиц отбиралась вручную, так как при разработке дисциплины каждая статья пособия сопровождалась списком слов и словосочетаний с переводом, которые авторы курса отбирали самостоятельно в соответствии с их ощущениями и интуицией.

На втором этапе в результате обработки скомпилированного нами корпуса с помощью пакета программ WordSmith 6.0 были получены ключевые слова по каждой из тем дисциплины и дополнительный список лексики для расширения словарного запаса, связанного с профессиональным дискурсом, в частности для совершенствования навыков чтения академических публикаций. Как уже указывалось, экспериментальная часть данной работы основана на использовании оригинального корпуса профессиональных текстов объемом 1 млн 100 словоупотреблений, включающего лингводидактические материалы и академическую периодику по релевантным темам на английском языке, а также серию макроэкономических обзоров.

Дополнительно был выполнен второй проверочный автоматический вариант отбора с помощью созданных нами алгоритмов, в основе которых лежит генерирование частотного словаря по темам курса. Дело в том, что ряд исследователей, включая М. Скотта, М. Хоуи, Р. Картера и др. [35], говоря о лексическом корпусном анализе, вводят не только понятие ключевых слов, но и понятие содержательных слов (content words). Помимо ключевых, смысловая функция, элементы связности текста и раскрытия его тематического содержания обеспечиваются и содержательными словами, которые также могут быть типичными для данной предметной области и профессионального жанра. Таким образом, проводя это разграничение, мы должны учитывать данную категорию лексики в отбираемом минимуме.

Таблица 3

Ключевые и содержательные слова

Ключевые слова	Высокочастотные содержательные слова
INTANGIBLE	COMMODITY
ASSETS	VULNERABLE
CAPITAL	STAGNATION

Отметим, что ключевые и высокочастотные содержательные слова выполняют роль указки, т.е. направляют внимание преподавателя на дальнейшую работу с этой лексикой, включающую произношение, профессионально релевантные коллокации, коллигации или наиболее грамматически вероятное оформление лексических единиц. Поэтому планируется включить отобранные лексические единицы в пособие по курсу с комплексом лексических заданий.

5. Выводы

Подводя итоги исследования, можно сделать выводы относительно (i) процесса составления и обработки корпусов, (ii) особенностей автоматически отобранной статистически значимой лексики и (iii) интегрирования полученных данных в процесс преподавания английского языка для специальных целей.

В процессе составления и обработки специального корпуса было выявлено два существенных обстоятельства: (i) значительный объем подготовительных мероприятий по отбору текстов и организации их в корпус, (ii) избыточность списка по количеству слов даже при опоре на показатели частотности (не менее 15 000 лемм без учета слов с единичной частотностью, в том числе порядка 1000 ключевых слов). Объем предварительной работы при составлении корпуса, вероятно, является одним из факторов, сдерживающих активное использование корпусов при разработке и внедрении дисциплин на основе профессионально ориентированного английского языка. Бурное развитие методов цифровой обработки текстов позволит в будущем сократить трудоемкость предварительного этапа составления корпуса.

Что касается объема списка ключевых слов, то при наличии четких критериев сокращения автоматически отобранного списка допустимо утверждать, что преподаватель вполне может положиться на корпусный отбор ключевых слов как предварительный список слов с опорой на частотность и показатель величины «ключевого характера», т. е. роль слова в формировании смыслов. В дальнейшем необходимо корректировать полученный список ключевых слов «вручную», (i) соотнося его с потребностями конкретных обучаемых (ряд слов, не представляющих интереса для конкретной группы обучаемых, по результатам входных тестов может быть исключен), (ii) исключая из него те слова, которые относятся к так называемому «общеанглийскому» (General English) пласту лексики (см. список Уэста), (iii) основываясь на экспертных мнениях специалистов в данной области.

В отношении сгенерированных списков ключевых слов необходимо отметить, что несмотря на инициальные трудозатраты, изучать отобранный компьютером список ключевых слов значительно легче, чем отбирать их из всего текста или коллекции текстов в силу того, что список ключевых слов, подготовленный компьютером, существенно меньше по количеству слов. Следует также подчеркнуть, что выделить важные слова в тексте объемом в страницу для преподавателя не составит труда. Получить же статистически значимый список ключевых слов при подготовке профессионально ориентированного курса иностранного языка на материале в несколько тысяч или миллионов словоупотреблений вручную уже не представляется возможным. А именно в этом и заключается преимущество корпусных методов – в обработке больших массивов данных для получения достоверного результата.

На наш взгляд, на данном этапе развития алгоритмов отбора слова с единичной частотностью представляют нерешенную проблему отбора по двум причинам: (i) алгоритмы не могут отобрать гапаксы в качестве ключевых в силу низкой частотности, (ii) слова с единичной частотностью составляют половину словаря любого корпуса текстов. Тем не менее среди слов с единичной частотностью могут встречаться критически важные для понимания смысла единицы. Таким образом, на данный момент такие единицы должны просеиваться «вручную», а решение о включении в словарный минимум принимают эксперт-экономист и преподаватель английского языка.

Особого внимания заслуживают методический потенциал полученного списка ключевых слов и дальнейшее его использование в практике преподавания дисциплины «Tangible and Intangible Economy» (Материальная и нематериальная экономика) и профильно ориентированного английского языка в целом. Прежде всего, практическая работа с корпусом предполагает полное перестроение учебного процесса по дисциплине таким образом, чтобы формирование лексической базы, а также развитие всех видов речевой деятельности осуществлялись полностью на основе сгенерированного программой WordSmith 6.0. списка частотности. Построение лексического минимума на основе автоматически сгенерированного списка ключевых слов имеет ряд преимуществ перед традиционным принципом составления «вручную» с опорой на интуицию преподавателя. Во-первых, исключен

принцип субъективности и прескриптивности; во-вторых, достигается эмпирическая адекватность и аутентичность английского языка, который студенты используют для решения своих профессиональных задач.

В целом речь идет о разработке комплексной модели обучения с использованием корпусной технологии. Основными компонентами такой модели являются: принципы обучения, методы обучения, средства обучения, средства контроля [36]. Спорадическое применение отдельных учебных приемов (например, использование готовых корпусов или упражнения на основе конкорданса) не представляются нам более достаточно эффективными для развития всех видов речевой деятельности, а остаются лишь на уровне моды на корпусные технологии.

Создание модели обучения не менее трудоемкий процесс, чем составление и обработка корпуса, но он столь же необходим, если наша конечная цель – «доказательное» обучение (data driven learning) [37]. Модель обучения на основе технологии корпусной лингвистики, очевидно, потребует определенных усилий как со стороны преподавателей, так и со стороны обучающихся [38]. Со стороны обучающихся наблюдается некая субъективная привязанность к традиционным лексическим упражнениям на заполнение пропусков и раскрытие скобок. Со стороны преподавателей может наблюдаться невладение методикой составления упражнений на основе корпусной технологии [39].

В контексте преподавания дисциплины «Tangible and Intangible Economy» (Материальная и нематериальная экономика) опора на список ключевых слов открывает перспективы по созданию учебного пособия нового типа, которое помимо системы упражнений, построенной на основе корпуса, может содержать рекомендации по самостоятельной работе с профессиональным корпусом по дисциплине и иметь мультиплатформенный и мультимодальный характер.

Список источников

1. Ланидус Б.А. Проблемы содержания обучения языку в языковом вузе. М. : Высш. шк., 1986. 144 с.
2. Scott M. Wordsmith Tools: Software. Oxford : Oxford University Press, 2012.
3. Hutchinson T., Waters A. English for specific purposes: A learning-centered Approach. S.l. : Cambridge Univ. Press, 1991. 183 p.
4. Гез Н.И., Фролова Г.М. История зарубежной методики преподавания иностранных языков : учеб. пособие для студ. лингв. ун-тов и ф-тов ин. яз. высш. пед. учеб. заведений. М. : Академия, 2008. 256 с.
5. Рыков В.В. Корпус текстов как реализация объектно-ориентированной парадигмы // Компьютерная лингвистика и интеллектуальные технологии: тр. междунар. семинара «Диалог–2002», Протвино, 6–11 июня 2002 г. : в 2 т. Т. 1. М., 2002. С. 390–393.
6. Palmer G. Second Interim Report on English Collocations. Tokyo : Kaitakusha, 1933.
7. Гальскова Н.Д., Гез Н.И. Теория и методика обучения иностранным языкам: лингводидактика и методика : учеб. пособие для студ. лингв. ун-тов ф-тов ин. яз. высш. пед. учеб. заведений. 6-е изд., стереотип. М : Академия, 2009. 336 с.
8. Шатилов С.Ф. Методика обучения немецкому языку в средней школе : учеб. пособие для студ. пед. ин-тов по спец. № 2103 «Иностр. яз.». 2-е изд., дораб. М. : Просвещение, 1986. 223 с.

9. West R. et al. The Word Family Framework (WFF). 2013. URL: <http://www.teaching-english.org.uk/article/word-family> (дата обращения: 15.02.2014).
10. O'Keeffe A., McCarthy M., Carter R. From Corpus to Classroom: language use and language teaching. Cambridge : Cambridge Univ. Press, 2007. 315 p.
11. Simpson R., Mendis D. A corpus-based study of idioms in academic speech // TESOL Quarterly. 2003. № 37. P. 419–441.
12. Milton J., Meara P. How periods abroad affect vocabulary growth in a foreign language // ITL: Review of Applied Linguistics. 1995. Vol. 107–108. P. 17–34.
13. Gorina O., Tsarakova N., Tsarakov S. Study of Optimal Text Size Phenomenon in Zipf–Mandelbrot's Distribution on the Bases of Full and Distorted Texts. Author's Frequency Characteristics and derivation of Hapax Legomena // Journal of Quantitative Linguistics. 2019. doi: 10.1080/09296174.2018.1559460
14. Власова Е.А., Карпова Е.Л., Ольшевская М.Ю. Лексический минимум по языку специальности: сколько слов достаточно?: Разработка принципов минимизации // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 4. С. 63–77. doi: 10.25205/1818-7935-2019-17-4-63-77
15. Плунгян В.А. Корпус как инструмент и как идеология: О некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении. 2008. № 2 (16). С. 7–20.
16. Арнольд И.В. Основы научных исследований в лингвистике : учеб. пособие. М. : Высш. шк., 1991. 140 с.
17. Копотев М.В. Введение в корпусную лингвистику : учеб. пособие для студентов филологических и лингвистических специальностей университетов. Praha : Animedia, 2014.
18. Захаров В.П. Корпусная лингвистика : учеб.-метод. пособие. СПб. : Изд-во СПбГУ, 2005. 48 с.
19. Рыков В.В. Корпусная лингвистика: (Научно-аналитический обзор) // РЖ: Социальные и гуманитарные науки: Зарубежная литература. М., 1996.
20. The research infrastructure for language as social and cultural data. URL: www.clarin.eu (дата обращения: 28.04.2023).
21. Biber D., Johansson S., Leech G., Conrad S., Finegan E. Longman Grammar of Spoken and Written English. S.l. : Pearson Education Limited, 1999. 1204 p.
22. Камишилова О.Н. Исследовательский потенциал корпуса английских текстов петербургских школьников: анализ интерязыка // Известия РГПУ им. А.И. Герцена. 2009. № 104. С. 114–123.
23. Богоявленская Ю.В. Репрезентативность лингвистического корпуса: метод верификации достоверности полученных данных // Политическая лингвистика. 2016. Вып. 4 (58). С. 163–166.
24. Английский для специальных академических целей: Экономика материальных и нематериальных активов-3 // Программа учебной дисциплины. URL: <https://www.hse.ru/edu/courses/647339099> (дата обращения: 28.04.2023).
25. Концепция развития англоязычной коммуникативной компетенции студентов НИУ ВШЭ. URL: <https://www.hse.ru/docs/381549301.html> (дата обращения: 28.04.2023).
26. Рогожникова В.Н. Язык современной экономической науки: подходы к изучению // Научные исследования экономического факультета. Электронный журнал. 2022. Т. 14, вып. 4. С. 7–23. doi: 10.38050/2078-3809-2022-14-4-7-23
27. Шилихина К.М. Использование корпусов в исследованиях дискурса // Вестник Воронеж. гос. ун-та. Сер.: Лингвистика и межкультурная коммуникация. 2014. № 3. С. 21–26.
28. Азимов Э.Г., Щукин А.Н. Новый словарь методических терминов и понятий. М. : ИКАР, 2009. 448 с.
29. Казакова М.А., Евтюгина А.А. Аутентичные текстовые материалы в обучении иностранному языку // Вестник Бурятского государственного университета. Образование. Личность. Общество. 2016. № 4. С. 50–59.

30. Мильруд П.П. Компетентность учителя иностранного языка // Иностранные языки в школе. 2012. № 1. С. 11–17.
31. Воронина Г.И. Организация работы с аутентичными текстами молодежной прессы в старших классах школ с углубленным изучением иностранного языка // Иностранные языки в школе. 1999. № 2. С. 23–25.
32. Карамнов А.С. Модель создания корпуса учебника английского языка // Научный диалог. 2013. № 2 (14): Педагогика. С. 59–69.
33. Ахмади Л. Лексическая насыщенность как лингво- и социокультурная характеристика англоязычного научного текста // Иностранные языки в высшей школе. 2022. № 4 (62). С. 5–17. doi: 10.37724/RSU.2022.62.3.001
34. Haskel J., Westlake S. Capitalism without Capital: The Rise of the Intangible Economy. Princeton ; Oxford : Princeton University Press, 2018.
35. Hovey M. Lexical Priming: A New Theory of Words and Language. Routledge, 2005. 202 p.
36. Чернякова Т.А. Методика формирования лексических навыков студентов на основе лингвистического корпуса (английский язык, языковой вуз) : дис. ... канд. пед. наук. М., 2012.
37. Котюрова И.А. DDL (Data Driven Learning) в условиях российских вузов // Актуальные проблемы современной лингвистики и гуманитарных наук : сборник статей XIII Всероссийской научно-методической конференции с международным участием. М., 2021. С. 487–501. doi: 10.22363/10784-2021-487-501
38. Grigaliūnienė J. Corpora in the classroom. Vilnius : Vilnius University, 2013. 81 p.
39. Слободян Е.А. Обучение методике составления упражнений с помощью национального корпуса русского языка // Гуманитарные исследования в Восточной Сибири и на Дальнем Востоке. 2010. № 1. С. 71–73.

References

1. Lapidus, B.A. (1986) *Problemy sodержaniya obucheniya yazyku v yazykovom vuze* [Problems of language teaching content in language universities]. Moscow: Vyssh. shk.
2. Scott, M. (2012) *Wordsmith Tools: Software*. Oxford: Oxford University Press.
3. Hutchinson, T. & Waters, A. (1991) *English for specific purposes: A learning-centered approach*. Cambridge Univ. Press.
4. Gez, N.I. & Frolova, G.M. (2008) *Istoriya zarubezhnoy metodiki prepodavaniya inostrannykh yazykov* [History of foreign language teaching methodology]. Moscow: Akademiya.
5. Rykov, V.V. (2002) [Text corpus as an implementation of object-oriented paradigm]. *Kompyuternaya lingvistika i intellektual'nye tekhnologii* [Computational linguistics and intelligent technologies]. Proceedings of the International Seminar "Dialog-2002". Protvino. 6–11 June 2002. Vol. 1. Moscow. pp. 390–393. (In Russian).
6. Palmer, G. (1933) *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.
7. Gal'skova, N.D. & Gez, N.I. (2009) *Teoriya i metodika obucheniya inostrannym yazykam: lingvodidaktika i metodika* [Theory and methodology of foreign language teaching]. Moscow: Akademiya.
8. Shatilov, S.F. (1986) *Metodika obucheniya nemetskomu yazyku v sredney shkole* [Methods of teaching German in secondary school]. Moscow: Prosveshchenie.
9. West, R. et al. (2013) *The Word Family Framework (WFF)*. [Online] Available from: <http://www.teachingenglish.org.uk/article/word-family>
10. O'Keeffe, A., McCarthy, M. & Carter, R. (2007) *From Corpus to Classroom: language use and language teaching*. Cambridge: Cambridge Univ. Press.
11. Simpson, R. & Mendis, D. (2003) A corpus-based study of idioms in academic speech. *TESOL Quarterly*. 37. pp. 419–441.

12. Millton, J. & Meara, P. (1995) How periods abroad affect vocabulary growth in a foreign language. *ITL: Review of Applied Linguistics*. 107–108. pp. 17–34.
13. Gorina, O., Tsarakova, N. & Tsarakov, S. (2019) Study of Optimal Text Size Phenomenon in Zipf–Mandelbrot's Distribution on the Bases of Full and Distorted Texts. *Journal of Quantitative Linguistics*. doi: 10.1080/09296174.2018.1559460
14. Vlasova, E.A., Karpova, E.L. & Ol'shevskaya, M.Yu. (2019) Vocabulary: how many words are enough? Principles of minimizing learners' vocabulary. *Vestnik NGU. Seriya: Lingvistika i mezhkul'turnaya kommunikatsiya*. 17 (4). pp. 63–77. (In Russian). doi: 10.25205/1818-7935-2019-17-4-63-77
15. Plungyan, V.A. (2008) Korpus kak instrument i kak ideologiya [Corpus as a tool and as ideology]. *Russkiy yazyk v nauchnom osveshchenii*. 2 (16). pp. 7–20.
16. Arnold, I.V. (1991) *Osnovy nauchnykh issledovaniy v lingvistike* [Fundamentals of linguistic research]. Moscow: Vyssh. shk.
17. Kopotev, M.V. (2014) *Vvedenie v korpusnuyu lingvistiku* [Introduction to corpus linguistics]. Praha: Animedia.
18. Zakharov, V.P. (2005) *Korpusnaya lingvistika* [Corpus linguistics]. Saint Petersburg: Saint Petersburg State University.
19. Rykov, V.V. (1996) *Korpusnaya lingvistika* [Corpus linguistics]. In: *RZh: Sotsial'nye i gumanitarnye nauki: Zarubezhnaya literatura* [RZh: Social and Humanitarian Sciences: Foreign Literature]. Moscow.
20. Clarin.eu. (n.d.) *The research infrastructure for language as social and cultural data*. [Online] Available from: www.clarin.eu
21. Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. Pearson Education Limited.
22. Kamshilova, O.N. (2009) Issledovatel'skiy potentsial korpusa angliyskikh tekstov peterburgskikh shkol'nikov [Research potential of a corpus of English texts by St. Petersburg schoolchildren]. *Izvestiya RGPU im. A.I. Gertsena*. 104. pp. 114–123.
23. Bogoyavlenskaya, Yu.V. (2016) Reprezentativnost' lingvisticheskogo korpusa [Representativeness of a linguistic corpus]. *Politicheskaya lingvistika*. 4 (58). pp. 163–166.
24. HSE. (n.d.) *Angliyskiy dlya spetsial'nykh akademicheskikh tseley: Ekonomika material'nykh i nematerial'nykh aktivov-3* [English for specific academic purposes: Economics of tangible and intangible assets-3]. [Online] Available from: <https://www.hse.ru/edu/courses/647339099>
25. HSE. (n.d.) *Kontseptsiya razvitiya angloyazychnoy kommunikativnoy kompetentsii studentov NIU VShE* [Concept of developing English communicative competence of HSE students]. [Online] Available from: <https://www.hse.ru/docs/381549301.html>
26. Rogozhnikova, V.N. (2022) Yazyk sovremennoy ekonomicheskoy nauki [Language of modern economic science]. *Nauchnye issledovaniya ekonomicheskogo fakul'teta. Elektronnyy zhurnal*. 14 (4). pp. 7–23. doi: 10.38050/2078-3809-2022-14-4-7-23
27. Shilikhina, K.M. (2014) Ispol'zovanie korpusov v issledovaniyakh diskursa [Using corpora in discourse studies]. *Vestnik Voronezh. gos. un-ta. Ser.: Lingvistika i mezhkul'turnaya kommunikatsiya*. 3. pp. 21–26.
28. Azimov, E.G. & Shchukin, A.N. (2009) *Novyy slovar' metodicheskikh terminov i ponyatiy* [New dictionary of methodological terms]. Moscow: IKAR.
29. Kazakova, M.A. & Evtugina, A.A. (2016) Autentichnye tekstovye materialy v obuchenii inostrannomu yazyku [Authentic texts in foreign language teaching]. *Vestnik Buryatskogo gosudarstvennogo universiteta*. 4. pp. 50–59.
30. Mil'rud, R.P. (2012) Kompetentnost' uchitelya inostrannogo yazyka [Foreign language teacher competence]. *Inostrannye yazyki v shkole*. 1. pp. 11–17.
31. Voronina, G.I. (1999) Organizatsiya raboty s autentichnymi tekstami molodezhnoy pressy [Working with authentic youth press texts]. *Inostrannye yazyki v shkole*. 2. pp. 23–25.
32. Karanov, A.S. (2013) Model' sozdaniya korpusa uchebnika angliyskogo yazyka [Model for creating an English textbook corpus]. *Nauchnyy dialog*. 2 (14). pp. 59–69.

33. Akhmedi, L. (2022) Leksicheskaya nasyschennost' kak lingvo- i sotsiokul'turnaya kharakteristika angloyazychnogo nauchnogo teksta [Lexical density as a linguistic and sociocultural feature]. *Inostrannyye yazyki v vysshey shkole*. 4 (62). pp. 5–17. doi: 10.37724/RSU.2022.62.3.001
34. Haskel, J. & Westlake, S. (2018) *Capitalism without Capital: The Rise of the Intangible Economy*. Princeton; Oxford: Princeton University Press.
35. Hoey, M. (2005) *Lexical Priming: A New Theory of Words and Language*. Routledge.
36. Chernyakova, T.A. (2012) *Metodika formirovaniya leksicheskikh navykov studentov na osnove lingvisticheskogo korpusa* [Methods of developing lexical skills using linguistic corpora]. Pedagogics Cand. Diss. Moscow.
37. Kotyurova, I.A. (2021) [DDL in Russian universities]. *Aktual'nye problemy sovremennoy lingvistiki i gumanitarnykh nauk* [Current issues of modern linguistics and humanities]. Conference Proceedings. Moscow. pp. 487–501. (In Russian). doi: 10.22363/10784-2021-487-501
38. Grigaliūnienė, J. (2013) *Corpora in the classroom*. Vilnius: Vilnius University.
39. Slobodyan, E.A. (2010) Obuchenie metodike sostavleniya uprazhneniy s pomoshch'yu natsional'nogo korpusa russkogo yazyka [Teaching exercise design using the Russian National Corpus]. *Gumanitarnye issledovaniya v Vostochnoy Sibiri i na Dal'nem Vostoke*. 1. pp. 71–73.

Информация об авторах:

Горина О.Г. – канд. пед. наук, доцент департамента иностранных языков Санкт-Петербургской школы гуманитарных наук и искусств Национального исследовательского университета «Высшая школа экономики» (Санкт-Петербург, Россия). E-mail: gorina@bk.ru

Камнева Л.Э. – старший преподаватель департамента иностранных языков Санкт-Петербургской школы гуманитарных наук и искусств Национального исследовательского университета «Высшая школа экономики» (Санкт-Петербург, Россия). E-mail: kamneva.larisa.2706@gmail.com

Кучеренко С.Н. – старший преподаватель департамента иностранных языков Санкт-Петербургской школы гуманитарных наук и искусств Национального исследовательского университета «Высшая школа экономики» (Санкт-Петербург, Россия). E-mail: sunflowersnk@gmail.com

Куганова Д.А. – студент Национального исследовательского университета «Высшая школа экономики» (Санкт-Петербург, Россия). E-mail: dakuganova@edu.hse.ru

Авторы заявляют об отсутствии конфликта интересов.

Information about the authors:

O.G. Gorina, Cand. Sci. (Pedagogics), associate professor, National Research University Higher School of Economics (St. Petersburg, Russian Federation). E-mail: gorina@bk.ru

L.E. Kamneva, senior lecturer, National Research University Higher School of Economics (St. Petersburg, Russian Federation). E-mail: kamneva.larisa.2706@gmail.com

S.N. Kucherenko, senior lecturer, National Research University Higher School of Economics (St. Petersburg, Russian Federation). E-mail: sunflowersnk@gmail.com

D.A. Kuganova, student, National Research University Higher School of Economics (St. Petersburg, Russian Federation). E-mail: dakuganova@edu.hse.ru

The authors declare no conflicts of interests.

Статья поступила в редакцию 04.05.2023;
одобрена после рецензирования 12.11.2024; принята к публикации 26.03.2025.

The article was submitted 04.05.2023;
approved after reviewing 12.11.2024; accepted for publication 26.03.2025.