

## **Корпусные инструменты, маршруты и эксперименты в современной лингводидактике**

**О. Г. Горина, Н. С. Царакова**

*Национальный исследовательский университет «Высшая школа экономики»  
Санкт-Петербург, Россия*

### *Аннотация*

Рассматриваются корпусные исследования как инструмент изучения и обучения английскому языку, его регистровым особенностям, лексике и грамматике. В круг задач исследования включены некоторые аспекты обучения иностранному языку в вузе с привлечением корпусной статистики, упоминаются когнитивные и практические предпосылки эффективности такого подхода. Дается обзор становления корпусной лингводидактики, приводится реферативный анализ влияния корпусных исследований на отбор содержания обучения. Детально обсуждаются результаты лингвостатистических исследований разнородных по жанрам материалов больших диверсифицированных корпусов английской речи, повлиявших на формирование качественных и количественных характеристик словарного запаса обучаемого. Дидактическое осмысление корпусных данных привело к тому, что при отборе вокабуляра обучаемого четче высвечены регистровые различия, введены новые дискурс-маркеры, отражены средства хеджинга. Обусловленное тем, что язык относится к ранговым распределениям, покрытие (понимание) текстов растет медленно по отношению к количеству новых выученных слов, поэтому начиная с запаса слов объемом 5–6 тыс. целесообразно инвестировать усилия в наращивание тщательно отобранной лексики по релевантным для обучаемого тематикам, в том числе профессиональной направленности. Приводится ряд примеров корпусных исследований, проводимых в лингводидактических целях.

Предлагаются способы трансформации этих исследований в аудиторные задания и деятельность, нацеленную на повышение осознанности (awareness-raising activities). Статья последовательно рассматривает вариативные способы развития заданий с использованием «сырых» корпусных данных. В выводах дается дидактическое осмысление целесообразности использования при обучении уникальных материалов с опорой на корпус. Публикация пропагандирует корпус как обязательный формат предъявления знания, обсуждает преимущества освоения корпусных инструментов как студентами, так и преподавателями. В выводах подчеркивается прогрессивность корпусного подхода, в рамках которого студент добывает знание с помощью корпусного инструментария и исследования, может верифицировать свои гипотезы, а не получает знание в готовом виде.

### *Ключевые слова*

корпусная лингводидактика, корпусно-ориентированный подход, регистровые особенности, объем словаря, корпусные аудиторные приемы

### *Благодарности*

Публикация подготовлена в ходе проведения исследования (№ 20-01-022) в рамках Программы «Научный фонд Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ)» в 2020–2021 гг.

### *Для цитирования*

Горина О. Г., Царакова Н. С. Корпусные инструменты, маршруты и эксперименты в современной лингводидактике // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2021. Т. 19, № 2. С. 36–53. DOI 10.25205/1818-7935-2021-19-2-36-53

## Corpus Routes and Experiments in Language Teaching

Olga G. Gorina, Natalya S. Tsarakova

HSE University  
St. Petersburg, Russian Federation

### Abstract

Corpus quantitative approach in teaching, which is of growing interest, entails some revision of the L2 vocabulary selection procedures and provides solutions for a wide range of practical problems. The focus throughout is on the discussion of research on the quantitative and qualitative characteristics of language which both teaching content and language acquisition practice could draw on.

This research regards human language as a rank distribution, which has serious implications on quantitative aspect of learner's vocabulary. We also looked into the ways to factor in the data from a small professional discourse corpus in order to target the units that have the greatest statistical prominence. Both BNC and our own collection of texts are explored. The study also elaborates on academic writing cohesive devices and grammar patterns introduction, which is approached through concordancing corpus strings to (i) provide frequency evidence and (ii) introduce a contrast in usage in various corpus genres. Striking differences that are evidenced by the frequency lists could be related not only to register, but also to the choice and instances of academic cohesive clusters which are favoured by the apprentice writers and the expert writers. With the aim of capitalizing on corpus approach a number of small-scale corpus research tasks were developed.

This study also uses corpus tools and data to give a seemingly subjective phenomenon of hedging some quantitative measurement. While experimenting on corpus in the classroom, the attention of learners was drawn to various means of hedging, such as lexical bundles or down-toners that manifest themselves as important communicative strategies.

Thus, corpus was used to inform both the language instructor and the student in the classroom to look in detail at differences in the use of lexical and grammatical units in different varieties of language, address contrasting register variations, and readily provide contemporary professionally relevant examples of actual language usage.

It has to be noted that university students have a tendency not to perceive register violations as language errors on a par with those of grammar, lexis or punctuation. Hence, corpus investigation as raising awareness tool also proved to be an effective teaching material generator. Nowadays syllabi have the opportunity to be rather sensitive to the quantitative evidence that corpora offer us; what is more, as a result of this study, we would conclude that university students are responsive to the small-scale investigation of register differences, lexico-grammatical frequency and patterning, which have been brought directly into the L2 classroom.

### Keywords

corpus-based teaching, corpus-oriented approach, register features, learner vocabulary, corpus classroom techniques

### Acknowledgements

The publication was prepared within the framework of the Academic Fund Program at the HSE University in 2020–2021 (grant №20-01-022)

### For citation

Gorina, Olga G., Tsarakova, Natalya S. Corpus Routes and Experiments in Language Teaching. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2021, vol. 19, no. 2, p. 36–53. DOI 10.25205/1818-7935-2021-19-2-36-53

## Введение

Постановка задач исследования, сопрягающего несколько областей знаний, таит большой потенциал, требующий раскрытия, поэтому сегодняшняя парадигма исследований нередко характеризуется как «время дефисных наук»<sup>1</sup>. Термин «корпусная лингводидактика», указывающий на пересечение корпусной лингвистики и лингводидактики еще не является устойчивой коллокацией, но это новое направление лингводидактики уверенно развивается. В нашем междисциплинарном исследовании мы обращаемся к компьютерным лингвистическим корпусам текстов, которые опираются на миллионные коллекции примеров естественного употребления языка, как для исследования изучаемого языка, так и для предъявления языко-

<sup>1</sup> См.: Голубкова Е. Е. Отзыв официального оппонента на диссертацию О. Г. Гориной / МГУ. 2014. С. 1. URL: [http://www.philol.msu.ru/~ref/001\\_79\\_14.htm](http://www.philol.msu.ru/~ref/001_79_14.htm) (дата обращения 19.09.2020).

вых данных при обучении. Примеры для лингвометодических материалов отбирались также из составленного нами корпуса, состоящего из около 3 млн словоупотреблений.

Актуальность исследования связана в первую очередь с тем, что корпусная лингвистика легко выявляет аутентичные употребления (usage) языкового сообщества, которые можно также использовать для разработки лингвометодических материалов и количественных оценок содержания обучения.

### 1. От первого конкорданса к корпусной лингводидактике

Бурное развитие корпусной лингвистики в последние десятилетия привело к коренным изменениям в области лексикографии, за которыми последовали изменения и в области преподавания иностранного языка, а также «намечаются контуры новой модели языка, которая в ряде существенных отношений отличается от привычных моделей, сложившихся в последней четверти XX века» [Плунгян, 2008. С. 7]. Эти изменения стали настолько радикальными, что их часто описывают как революционные.

Следует вспомнить, однако, что к идее составления корпусов с целью последующего обучения языкам с опорой на коллекции других аутентичных текстов человечество обратилось уже давно. Полные тексты высокого статуса собирались, хранились в библиотеках и изучались как образцы стиля, риторики, грамматики. Первый конкорданс к латинскому переводу библии «Вульгата» был задуман и создан еще в XIII в. Этот известный предшественник корпусных конкордансов был собран вручную усилиями 500 монахов-доминиканцев. Лишь отсутствие технологий и мощных компьютеров оставило многих исследователей в докорпусной эпохе.

Развитие технологий создало возможности для возникновения компьютерных корпусов, основанных на больших массивах отработанных языковыми сообществами данных. В результате те же самые задачи по сбору и извлечению аутентичных образцов и моделей стали решаться быстрее и эффективнее. Со временем существенно изменилась и технология составления толковых словарей, которая сегодня стала менее книжно-литературной и учитывает также разговорные особенности функционирования языков как системный, а не только как субъективный экспрессивно-стилистический фактор. Но распространение в преподавательской среде корпусных технологий исследования узуса языка (usage) стало доступным и относительно простым лишь после создания языковых корпусов и компьютерных средств их автоматической обработки с целью извлечения релевантной информации о функционировании единиц национального словаря в аутентичных текстах с учетом всей их жанрово-стилистической палитры, обеспечившей недостижимую ранее репрезентативность выборки.

В последние десятилетия также стали активно разрабатываться новые корпусные грамматики английского языка, такие как Longman Grammar of Spoken and Written English (LGSWE), которая стала результатом 6-летнего исследования ведущих лингвистов и основана на материалах сорокамиллионного корпуса (Longman Spoken & Written English Corpus, LSWEC). Как отмечал Р. Квирк, для всякого автора, планирующего исследование грамматики любого языка на земле с опорой на корпус, LGSWE станет незаменимым руководством<sup>2</sup> (цит по: [Biber et al., 1999]). В 2002 г. вышла преемница LGSWE, The Cambridge Grammar of the English Language (CGEL) [Huddleston, Pullum, 2002]. Опубликована грамматика устного и письменного английского, The Cambridge Grammar of English, CGE [Carter, McCarthy, 2011], основанная на материалах Кембриджского международного корпуса (Cambridge International Corpus, 700 млн словоупотреблений), который включает образцы британского английского, американского английского и корпус ошибок обучаемых, созданный на базе 60 000 экзаменационных работ Cambridge ESOL. Материалы Кембриджского корпуса ошибок обучаемых (the

<sup>2</sup> “For anyone planning corpus-based grammatical study, of any language on earth, the LGSWE will rapidly establish itself as indispensable” [p. iv].

Cambridge Learner Corpus) исследовались с целью разработки более современных и эффективных лингводидактических подходов.

## 2. Векторы развития корпусной лингводидактики

Говоря о становлении корпусной лингводидактики, можно выделить два вектора ее развития. Первое направление связано с информированием потенциальных составителей учебников и словарей различных языков о релевантном потенциале наличных корпусных баз речевых данных. Ни один издатель учебной продукции для изучения самых распространенных языков уже не может обойтись без корпусных данных, в то время как потребители учебных материалов фактически пользуются результатами корпусных исследований независимо от того, знают они об этом или нет. Это уже привело к формированию нового стандарта для учебников по языку, который обозначается пометой *corpus-informed* на обложках таких изданий. Вопрос заключается уже во введении определенных критериев, которым они должны отвечать, чтобы исключить манипулятивное использование индикатора *corpus-informed* на обложке.

В рамках второго направления ведется разработка аудиторных приемов работы с опорой на корпус, в котором пионеры подхода остановились, так сказать, «на пороге аудитории», (“at the classroom door” [O’Keeffe et al., 2007. P. xi]), чтобы дать преподавателям право самим определять, какой объем корпусных заданий нужно предъявлять обучаемым с учетом развития их промежуточной языковой системы и особенностями их когнитивного стиля усвоения речевого материала.

Вспомним о том, что в когнитивных исследованиях и лексическом подходе к обучению существует представление о том, что как *ментальный лексикон* [КСКТ, 1997] обучающегося, так и промежуточная грамматическая система (*intergrammar*), представляют собой смесь (*muddle*) единиц, усвоенных в разной степени, которая развивается не линейно и последовательно, а холистически [Lewis, 2008. P. 47]. Эта, так сказать, «система» с достаточно большим количеством «переменных» не имеет точного аналитического решения, а определяется лишь приблизительными, вероятностными оценками. При этом индикаторами нахождения формируемой языковой системы в состоянии промежуточности служат допускаемые обучаемыми ошибки. В связи с этим корпусно-ориентированное обучение может быть полезным и прогрессивным, так как способствует тому, чтобы предложить обучающимся лишь материал, отобранный в соответствии с их предметно-коммуникативными потребностями, интересами и стратегиями усвоения предмета, называемыми в англоязычной литературе когнитивными стилями [Bartlett, 1932; Bernadini, 2004].

### 2.1. Влияние корпусов на отбор содержания обучения

Влияние корпусных изысканий значительно: расширяется функциональный набор курс-маркеров, детализируется перечень видов коммуникативных актов и средств для реализации функциональных моделей устной деятельности (выражения согласия / несогласия, отказа и др.). Важно и то, что на самых ранних уровнях вводятся прежде неакцентируемые в преподавании элементы, такие как средства хеджирования или средства реагирования (*response tokens*) в диалоге, обеспечивающие продуктивный диалог даже при минимальном наборе активной лексики у говорящего, поскольку они свидетельствуют об активном внимании собеседников к репликам друг друга. Также на основе анализа аутентичных источников уточняются условия нормативного употребления средств когезии в различных видах письменной продукции.

### 2.2. Особенности корпусного формирования словарного состава

Корпусные ориентиры позволяют уточнить как количественные, так и качественные показатели в структуре запаса слов, их частеречной принадлежности и др., оптимальные для

формирования эффективного дискурса. Так, помимо традиционного деления на пассивный и активный словари, лингвостатистические данные о регистровых особенностях заставили обратить более пристальное внимание на различия между письменной и устной речью. Характеризуя эти различия, лингвисты обратились к понятию лексической плотности (lexical density), определив ее как процентное соотношение знаменательных слов в тексте или на отрезке речи. Этот показатель существенно различается в устной и письменной речи, в связи с чем составители LGSWE предлагают различать три категории слов: (i) содержательные<sup>3</sup> (lexical words), такие как: *opposition*, *minister*, которые в среднем по корпусу LSWЕ составляют 50 % от всех его слов; (ii) служебные (называемые ими *functional*): *over*, *the*, *of*; (iii) включения (inserts): *erm*, *please* [Biber et al., 1999], определенные типы которых, такие как стимулирующие ответ включения (response elicitors), оформления пауз хезитации (hesitators) и привлекающие внимание (attention signals), встречаются только в устной, как правило, спонтанной, речи.

Отмечается ряд других зависимостей. В частности, дистрибуция функциональных слов тесно связана с распределением других лексических единиц: низкая частотность существительных в спонтанной разговорной речи компенсируется высокой плотностью местоимений, и, наоборот, высокая частотность существительных в новостном жанре и научно-учебной прозе соответствует низкой плотности местоимений. А служебные слова, связанные с глагольными формами, варьируют свою частотность в зависимости от частотности самих смысловых глаголов. Самая высокая частотность вспомогательных глаголов и наречий наблюдается в разговорной речи и в литературной прозе. Подобная зависимость роста грамматических определителей при существительных наблюдается между определяющими функциональными словами и существительными.

В научно-учебной прозе и новостных жанрах (репортаж), фиксируется самая высокая плотность не только существительных, но также предлогов и определителей, поскольку они нужны для того, чтобы специфицировать эти существительные и дополнять их значения [Biber, 2016].

Следует отметить, что детализации различий между письменно-книжной и спонтанно-разговорной речью в русском и английском языках посвящены работы Г.А. Орлова, написанные в 60–90-х гг. XX в.; следует также упомянуть работы Ю. М. Скребнева по коллоквиалистике (см., например, [Орлов, 1991; Скребнев, 1985]). Вместе с тем полномасштабное корпусное изучение неоднородности академического устного дискурса, который выстраивается по нормам разговорной, а не письменно-книжной речи, стал возможен лишь после создания крупных корпусов академической устной речи: The MICASE (Michigan Corpus of Academic Spoken English) и его британского аналога корпуса лекций The BASE (British Academic Spoken English) [Flowerdew, 2012]. Эти исследования посвящены метадискурсивным элементам (*so let me just elaborate a little bit and then we ...*) и их сопоставлению на материале лекций и семинаров (*that's a very interesting remark / good observation / a really important point*); существенное внимание уделяется элементам хеджинга (см. о хеджинге далее). Анализ лекционного материала говорит о том, что использование хеджей в диалоге «лектор – студент» не столько манифестирует стратегию вежливости, сколько служит средством создания благоприятной, равноправной (friendly) среды для дискуссии [Mauranen, 2003a; 2003b; Hyland, 2009].

Отметим, что статистические различия по стилистическим регистрам (и, следовательно, по функциональным стилям и жанрам речи) стали одним из основных выводов в исследованиях лонгмановского корпуса. Самые значительные различия наблюдаются в таких полярных видах речевой деятельности, как диалогическая разговорная речь и новостной дискурс: наибольшая плотность знаменательных слов, несущих информацию, приходится на новостные тексты, а самый низкий – на обиходно-бытовые диалоги (т. е. спонтанную неподготов-

<sup>3</sup> В советско-российской лингвистике они часто называются знаменательными.

ленную разговорную речь), в которых фатическая контактоустанавливающая функция часто доминирует над информативной.

В целом по корпусу LSWECS типичная дистрибуция знаменательных («содержательных») слов по упомянутым полярным жанрово-стилистическим формам коммуникации составляет 63 и 41 % соответственно; число служебных единиц в письменном дискурсе превышает 44 %, а в устном 37 %, тогда как определенные типы «включений» вроде *erm, right, yeah, please*, заполняющих паузы (ср. fillers (G. Brown)) и свойственных лишь устной речи, составляют целых 15 %. Корпусная информация по другим жанрово-стилистическим формам устной и письменной речи дает дополнительное понимание о должном составе словарного запаса *специалиста* или о моделях ведения профессионального дискурса как в академической среде, так и в менее формальных ситуациях социального взаимодействия (например, офисной обстановке).

Сравнительный анализ данных по пяти современным корпусам английского языка, репрезентирующих три вида дискурса (устный бытовой, устный академический и письменный академический), показал, что самыми частотными в устных бытовых диалогах являются маркеры интерактивности, такие как: *I, You, yeah*, а также слова *like, please, thanks*. Эти единицы выполняют функции вербального реагирования на высказывание (response tokens) или действия собеседника. Однако смена прагматико-коммуникативной установки дискурса резко меняет эти показатели частотности. Так, в устном академическом корпусе на первый план по частотности выходят лексические единицы, специфичные для письменного академического дискурса, и это значит, что его содержание превалирует над устной формой. Самыми частотными становятся артикли *a, the*, свидетельствующие о высокой встречаемости сочетаний существительных (noun phrases); предлог *of* в составе существительных с зависимым словом (post-modified noun phrases); конструкции с формальным подлежащим *it*. Кроме того, статистическую значимость приобретают глаголы сообщения (reporting verbs): *report, believe, consider, assume*; характерны придаточные предложения с союзом *that*, или *that*-клаузы (that-clause). Также частотны предлоги *to, for, in*, оформляющие фразовые единства с предлогами (prepositional phrases) [Biber et al., 1999; O’Keeffe et al., 2007; Farr et al., 2004]<sup>4</sup>. Отметим, что свойственные устной речи интерактивные местоимения (*I, you*) в академическом дискурсе отсутствуют, равно как и слова *year, like, now*.

Лингводидактическая проблема состоит в том, что факт регистровых различий практически не осознается студентами первого и второго годов обучения, которые в подавляющем большинстве ситуаций общения прибегают к элементам неформального разговорного регистра, считая, что такие «модные» выражения, как *kind of, sort of, lots of, gonna, wanna* и сокращения, а также американизированное произношение *neither, either* и др., допустимы в любых ситуациях, так как являются для студентов формальными маркерами высокого уровня владения языком, беглости речи (= fluency). Вместе с тем о том, что, например, сочетания *should’ve, would’ve* в устной речи произносятся сокращенно, обучаемые не знают. Таким образом, наблюдается то самое смешение (muddling) в разной степени усвоенной информации и заблуждений о реальных жанрово-стилистических различиях, которые требуют внимания лингводидактики для формирования осознанности.

Необходимо отметить, что такие общие характеристики английский научного узуса, как логичность и обоснованность подачи фактического материала, отсутствие эмоционально окрашенных единиц лексики и грамматических конструкций, пунктуации, абсолютная обяза-

<sup>4</sup> Справедливости ради нужно отметить, что в работе А. Л. Пумпянского [1981. С. 11–18, 310–320] эти функциональные классы описаны гораздо подробнее, хотя и нужно признать, что корпус обеспечивает большую достоверность информации. Списки слов-организаторов, составленные Пумпянским на статистической основе, должны быть уточнены, конечно, с помощью корпусных баз данных, особенно, в той части, которая касается их распределения по жанрово-стилистическим разновидностям академического и научного дискурса. Показательно, что уже тогда у автора упоминались такие черты академического узуса, как «осторожность», что, вне всякого сомнения, созвучно нынешней идее хеджирования.

тельность поддержки логических связей с помощью средств когезии [Пумпянский, 1981], а также связность и преемственность на уровне абзацев, четкие жанровые структуры, правила цитирования, соблюдение конвенций, таких как отсутствие новых идей в заключении и т. д. [Рябцева, 1999], – всегда востребованное знание, так как навыки академического письма не сформированы у студентов вузов ни в английском языке, ни в родном. В диахроническом плане часто высказывается (в том числе и молодыми преподавателями-предметниками по результатам участия в международных конференциях) предположение о том, что употребление пассивных конструкций меняется в сторону уменьшения в современном письменном научном узусе. В таких случаях помощь корпуса, например, научных статей по профессии неоценима, так как корпус позволит немедленно верифицировать или опровергнуть кажущуюся тенденцию на аутентичном материале.

Корпусные исследования дополняют лексическую картину элементами хеджинга, который может наиболее эффективно в плане репрезентативности вводиться на основе корпусных исследований [Горина, Храброва, 2017], но еще почти не отражен в учебниках. Хедж-маркеры несут существенную социолингвистическую нагрузку, оформляя такие коммуникативные стратегии, как снижение категоричности и безапелляционности, повышение неопределенности и, как следствие, вежливости высказывания. Они также выступают в роли речевого средства «подстраховки», поскольку, ограничивая силу и степень допущений, помогают избежать абсолютной трактовки предложенных утверждений и, следовательно, обвинений в необоснованных обобщениях [Марюхин, 2010; O’Keeffe et al., 2007]. Как указывает Д. Кристалл, не всегда заметный на первый взгляд хедж, например *there are perhaps 1 500 such cases a year...*, обеспечит говорящему определенную линию защиты. В случае критики ученый может обратить внимание аудитории на приблизительность оценки числа случаев в год, выраженного наречием *perhaps*. Интересно также отметить, что наречие *invariably* может употребляться как в буквальном значении *always* ‘всегда’, так и в значении *usually / nearly always* ‘обычно / почти всегда’, являясь одним из самых распространенных хеджей в научном дискурсе [Crystal, 1988]. По этой причине роль хеджирования заметна и в академическом дискурсе, где оно помогает оформить стратегию защиты от потенциальной критики.

Приведенные выше доводы, безусловно, расставляют с точки зрения лингводидактики новые акценты в формировании структуры словарного запаса и инвентаря средств общения. Сосредоточенность корпусных исследований на регистровых различиях научно-академического дискурса в этой связи не случайна. Именно они часто ускользают от внимания обучающихся; кроме того, регистровые нарушения вообще не осознаются ими как ошибки, подобные орфографическим, грамматическим или пунктуационным. Между тем практика академических обменов и программ обучения в англоязычных университетах показывает, что студент может получить отказ, если мотивационное письмо (*motivation letter*) или характеристика (*letter of reference*) не соблюдают нормы формального общения и написаны с использованием разговорно-обиходной лексики и грамматики. Поэтому трудно переоценить важность осознания обучающимися жанрово-стилистических различий между письменной и устной речью, и, в частности, лингвокультурное значение средств хеджинга, так как именно эти умения зачастую дифференцируют успешного кандидата (на экзамене по английскому языку, по программам академической мобильности, при составлении / защите проекта научного исследования) и неуспешного.

Корпусная статистика имеет большое значение и для решения задач по отбору лексических минимумов вузовского уровня. Проведенные ранее исследования по проблеме составления минимума на основе двух корпусных коллекций текстов [Горина, 2014] доказали свою жизнеспособность.

Кроме того, количественные исследования ранговых распределений на лингвистическом материале могут дать количественные ориентиры для формирования вокабуляра обучаемого [Baayen, 2001; Gorina et al., 2020]. Корпусная статистика показывает резкое убывание частот

ности «контент-слов» (т. е. знаменательных общенаучных слов и специальных терминов, несущих основную информацию), а 50 % слов любого корпуса составляют группу гапаксов (от *hapax legomena* – «только раз названное», т. е. слов встретившихся в данном корпусе лишь единожды). Ранговый характер распределения слов по частотности прямо связан с показателем покрываемости текста минимальным списком либо с вероятностью столкнуться с заученным словом в аналогичном письменном тексте или в устном общении на профессиональную тему. Лексические минимумы продвинутых уровней часто страдают от «эффекта перевернутой пирамиды», связанного с недостаточной покрываемостью, когда обучаемый извлекает из больших объемов материала и заучивает «на всякий случай» массу лексических единиц, но затем понимает, что они принадлежат к категории низкочастотных (к гапаксам).

По количественным оценкам, расширение словарного запаса уровня B2, который составляет ок. 6 тыс. слов, до словаря объемом 10 тыс. слов увеличивает шансы встретить новые 4 тыс. слов в тексте всего на 3 %, а при увеличении вокабуляра с 10 тыс. до 16 тыс. слов – лишь на 2 %. Рецептивный (т. е. пассивный) словарный запас обучающегося объемом в 5–6 тыс. слов дает ему возможность понимать 90 % неспециального, случайным образом отобранного письменного или устного текста. Однако это означает, что каждое 10-е слово студенту незнакомо либо вызывает затруднение и может быть неправильно истолковано в контексте, если полагаться на догадку. Следовательно, текст объемом в 500 слов потребует 50 обращений к словарю, что сильно замедляет понимание текста и превращает его в декодирование. В профессиональной среде для полного погружения в текст специалисту требуется рецептивный словарь на уровне 97–98 % слов определенного подъязыка [O’Keeffe et al., 2007; Горина, 2014].

Именно по этой причине с точки зрения разумного использования усилий студента и преподавателя начиная с объема словаря в 5–6 тыс. общеупотребительных единиц дальнейшее инвестирование усилий должно быть направлено на специализированную тематику при тщательном отборе лексики, опирающемся на частотные словари, корпуса текстов по специальности, мнения экспертов. В этом случае усвоение лексики даст существенное приращение в понимании текстов на иностранном языке в избранном предметном направлении. Сегодня многие университеты обладают ресурсами для создания специализированных корпусов, корпусно-ориентированного отбора лексики и составления достоверных лексических минимумов, обеспечивающих необходимую покрываемость условного профессионального текста [Власова и др., 2019; West, 2013].

### 3. Корпусные ресурсы в лингвометодическом аспекте

Продолжая развивать тему жанрово-стилистических отличий, отметим, что корпусные исследования дают небезынтересные уточнения в отношении использования средств когезии в письменном тексте, в том числе национально-специфичных, несмотря на то что они хорошо представлены в классической учебной литературе.

#### 3.1. Средства когезии *on the one hand / on the other hand*

Приведем небольшой пример, демонстрирующий связь между частотностью и естественностью речи. Так, частотность употребления средств когезии *on the one hand / on the other hand* в речи носителей языка говорит о том, что употребление первого не предполагает обязательного употребления второго во второй части аргумента [Scott, Tribble, 2006; Salazar, 2014]. По данным BNC<sup>5</sup>, *on the other hand* является кластером, или лексическим пучком (*lexical bundle*<sup>6</sup>) с частотностью более 20 вхождений на 1 млн словоупотреблений совокупно по всем письменным и устным текстам корпуса. При этом *on the one hand* таковым не является,

<sup>5</sup> British National Corpus (BNC). URL: <https://www.english-corpora.org/bnc/> (accessed: 19.09.2020).

<sup>6</sup> В терминах [Biber et al., 1999].



поскольку частотность кластера составляет 14 вхождений на 1 млн словоупотреблений. По академическим текстам оба выражения по частотности причисляются к категории кластеров, но при этом *on the one hand* употребляется значительно реже и не входит даже в первые 40 самых употребительных выражений, в то время как *on the other hand* занимает 12-е место. Отметим, что процесс выделения кластеров имеет свои условности, связанные с принятым пороговым значением частотности на 1 млн словоупотреблений, а также объемом корпуса [Biber et al., 1999; Flowerdew, 2012; Bestgen, 2020; Scott, Tribble 2006].

По данным нашего собственного корпуса (около 3 млн словоупотреблений), наблюдается сходная с BNC картина частотности. Мы сопоставили частотность на профессионально значимых текстах коллекции письменных отчетов организаций, таких как исследовательский институт The Transatlantic Academy и др., – характер частотности сохраняется: *on the other hand* встречается почти в 3 раза чаще. Несмотря на эту статистику, в письменном аутентичном дискурсе *on the one hand* употребляется в несколько раз чаще, чем *on the other hand*, и это должно было бы найти отражение в письменных работах наших студентов, но фактически оба выражения употребляются у них строго симметрично, т. е. с одинаковой частотностью, вопреки норме аутентичной письменной речи.

Выборка нашего корпуса показывает<sup>7</sup>, что употребление кластеров тяготеет в предложении к инициальной позиции: *On the other hand, however, it has been difficult* (иногда в паре с еще одним противопоставлением), однако есть случаи парного употребления *On the other* в конце предложения: *in oil consumption on the one hand versus the region's paltry oil reserves on the other*. Исследовательский интерес представляют также случаи размещения выражений между подлежащим и сказуемым: *The EU, on the other hand, follows / The US, on the other hand, faces*.

Систематическое изучение коллокационной нагрузки текста, включая средства когезии, в профессиональных источниках и у обучающегося может помочь студентам сформировать письменный стиль, соответствующий ожиданиям их профессиональных сообществ.

### 3.2. Контрастирующие особенности слова *that* в различных функциях

Приведем пример исследовательского задания, в котором предлагается обратить внимание на контрастирующие особенности слова *that*. Как явствует из частотности, *that* в письменной речи функционирует в качестве союза, определителя, наречия, и, чаще всего, *that* вводит придаточное предложение, т. е. участвует в развитии аргументации. В устных высказываниях *that* чаще отсылает к общим знаниям. Это значит, что по контекстуальной семантике *that* можно судить о принадлежности высказывания к устной речи (первое предложение), либо к письменной (второе предложение):

1) *I wish you wouldn't say things like that.*

2) *UK Government has been right to support the demands that this issue be confronted and dealt with.*

К этим выводам могут прийти и сами обучаемые, проведя краткий анализ корпусных примеров. Началом исследования, которое может проводиться прямо в аудитории, станет задание на распределение 10–15 корпусных предложений по письменным и устным жанрово-стилевым формам текста. Кроме того, анализ академических текстов обязательно выявит группу глаголов, которые особенно частотны в сочетании с *that*. К ним относятся: *argue, assume, believe, claim, conclude, ensure, establish, estimate, find, mean, note, notice, reveal, say, see, show, suggest, think*. Обучаемым можно предложить подобрать примеры предложений с этими глаголами из своих академических статей, из корпусов академического дискурса.

<sup>7</sup> Еще одним вариантом развития исследования является обращение к конкордансам с целью выявления, например, позиции средств когезии в предложениях и особенностей пунктуации.

### 3.3. Выявление ошибок в употреблении глагола *suggest*

Полезным было бы также развить тему грамматических особенностей некоторых глаголов, часто встречающихся в научной литературе, и обратить внимание на частотность и фоссилизацию ошибок в употреблении глагола *suggest* в связи с его высокой встречаемостью и влиянием межъязыковой интерференции. Во-первых, на примере нескольких десятков аутентичных предложений необходимо привлечь внимание к отсутствию употреблений *suggest* с инфинитивом в корпусных примерах. Во-вторых, когда обучаемые такого употребления не найдут, нужно подчеркнуть, что употребление по аналогии с русским языком ошибочно.

Еще одна плохо усваиваемая особенность *suggest* связана с употреблением через *that*-клаузу и модальный глагол *should*. Студенты часто ставят под сомнение целесообразность изучения употребления через модальный глагол *should*, оценивая данный грамматически возможный вариант как слишком сложный и малочастотный. В этой связи нужно заметить, что преподаватели в своей практике нередко сталкиваются с подобными заблуждениями, поэтому доказуемость верных утверждений о нормах узуса с помощью контраргументов из корпуса является неоспоримым показанием к его применению.

Итак, сравнение числа вхождений *suggest* по всем текстам корпуса (BNC\_ALL) и по академическому узусу (BNC\_ACAD) свидетельствует о том, что на сочетание с *-ing* из 8 689 обнаруженных вхождений приходится лишь около 100; при этом с *-ing* глагол *suggest* употребляется по всему корпусу, но в большей степени характерен для разговорных жанров. Вместе с тем в академическом регистре предпочтительнее употребление с придаточным через союз *that*. Подчеркнем, что в целом, по данным BNC, *suggest + -ing*, во-первых, употребляется во много раз реже, во-вторых, к академическим контекстам относится менее 1 % его употреблений, в то время как почти 50 % употреблений *suggest that* приходится на академический узус. Распределение варианта употребления *suggest that* (4 575 случаев) представлены в табл. 1<sup>8</sup>.

Тем не менее, на 4 575 вхождениях *suggest that* корпусная статистика предоставляет весомую доказательную базу необходимости изучения грамматического паттерна *suggest + that + [should]*, особенно в курсах академического письма. Возможность немедленной верификации, в том числе и повторной верификации результатов, избавляет преподавателя от неподтвержденных суждений.

Данные о частотности вариантов употребления глагола *suggest + -ing* и *suggest + that* по всем жанрам BNC и лишь по академическому регистру BNC приведены в табл. 2. Жанрово-стилистические различия можно актуализировать, изучив вопрос о том, какой способ употребления предпочтителен в научном отчете, курсовой работе, неформальном электронном письме или устной беседе.

Для корпусной лингводидактики большую ценность представляют предложения с четко просматриваемыми жанрово-стилистическими особенностями. Так, в комбинации *suggest + -ing* на первое место по частотности выходят *suggest coming / leaving / using / going* в таких жанровых формах, как *W\_fict\_prose*; *W\_newsp\_brdsh\_t\_nat\_misc*; *W\_non\_ac\_polit\_law\_edu*<sup>9</sup>:

i) *This is our vacation! I didn't suggest coming on this heap of a boat to work like a house-servant;*

ii) *She hoped that he wouldn't suggest coming up for a coffee; she didn't think she could rise to a social occasion at this particular moment...*

<sup>8</sup> Отметим, что в силу технических особенностей автоматического поиска в таблицу не вошли такие употребления, как *suggest + om 1 do 5 close + that*, например: *SUGGEST TWO TYPES OF LETTERING AIDS THAT; SUGGEST, HOWEVER INDIRECTLY, THAT; SUGGEST THE RADICAL INTELLECTUAL APPROACH THAT; SUGGEST VERY CLEARLY THAT; SUGGEST OR RECOMMEND THAT; SUGGEST TO READERS THAT; SUGGEST HERE IS THAT; SUGGEST, HOWEVER, THAT.*

<sup>9</sup> *W\_fict\_prose*: BNC Written, novels; *W\_newsp\_brdsh\_t\_nat*: BNC Written, broadsheet national newspapers: miscellaneous texts; *W\_non\_ac\_polit\_law\_edu*: BNC Written, non-academic prose: politics, laws, education.

Таблица 1

Число вхождений глагола *suggest* по всему корпусу BNC ALL

Table 1

Number of *suggest* entries in the full BNC


SECTION	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	NON-ACAD	ACADEMIC	MISC
FREQ	4575	305	173	208	246	869	<b>1883</b>	891
WORDS [M]	100	10.0	15.9	7.3	10.5	16.5	<b>15.3</b>	20.8
PER MIL	45.75	30.61	10.87	28.64	23.50	52.68	<b>122.82</b>	42.76
SEE ALL SUB-SECTIONS AT ONCE								

Таблица 2

Сопоставление частотности *suggest* + *-ing* и *suggest* + *that* по BNC \*

Table 2

Comparison of *suggest* + *-ing* and *suggest* + *that* frequencies in the BNC

Поисковый запрос	Частотность	
	по всему корпусу ALL freq.	по академическому компоненту ACAD freq.
<i>suggest</i> [vvg]	79	12 (15,19 %)
<i>suggest that</i>	4575	1883 (41,16 %)

\* Приведем расшифровку некоторых условных обозначений, используемых в BNC: ALL freq. – частотность по всем текстам / компонентам BNC; ACAD freq – частотность по академическому компоненту; *suggest* [vvg] – употребление глагола *suggest* с *-ing*-формой; *suggest that* – употребление глагола *suggest* через *that*-клаузу.

Несмотря на то что в некоторых случаях эти жанровые формы относятся к письменным текстам, в примерах из корпуса могут просматриваться стилистические признаки разговорного регистра речи, поскольку предложения могут представлять собой прямую или несобственно-прямую речь.

Вместе с тем отметим грамматическое явление, характерное для формата *spoken meetings* («планёрки»), когда союз *that* используется для развития предложений с явно наблюдаемой лексикализацией чанков: [i] *I / we / it [noun] would suggest that* (314 случаев), а также достаточно частотными [ii] *what I would suggest is...* (14 случаев) и заслуживающими внимания [iii] *I was / am / going [gonna] to suggest* (10 случаев).

Прагматика их функционирования такова:

i) **I would suggest that** while... / **findings would suggest that there was** / since **logic would suggest that** those who 'consume' / **We would suggest that** there is no comparable description of teacher development as such;

ii) So **what I would suggest is** that we reply to this for you;

iii) **what I was going to suggest is** that we should seek to negotiate with them a turnover rent / But **I was going to suggest** that we order a copy of the book.

Зарегистрированное явление составляет, на наш взгляд, важный для обучаемых элемент профессиональной компетенции по обеспечению взаимодействия в рамках совещаний, в неформальной обстановке «совещаний на бегу».

### 3.4. Паттерн с модальным глаголом *should*

Корпус также предлагает богатый иллюстративный материал для изучения паттерна с модальным глаголом *should*: *It is important / vital / essential / crucial that ... should*. Для продвинутых уровней можно рекомендовать отбор показательных примеров, демонстрирующих явления *subjunctive* в вышеупомянутых конструкциях, т. е. опущения *should* и употребления через инфинитив без частицы *to* [A], а также возможность перфектных инфинитивов (*should have done, should have been done*) только при употреблении ряда прилагательных, таких как *odd / strange / typical / natural / interesting / surprised / surprising / funny* и др., выражающих человеческие реакции [B].

#### A

i) It is **essential** that we **should** maintain both our sensitivity and our effectiveness, if we are to protect and watch over the interests of the people we work for and provide the kind of services they need;

ii) ... it is **essential** that the bureaucratic mind **be confined** to its proper function;

iii) It is **essential** that all these issues **be addressed** openly, free from ideology and rhetoric and from...;

iv) It is **vital** that commercial activities **be legally regulated**.

#### B

i) It seemed **odd** that he **should** feel ill at ease among the people whose vernacular he has espoused so wholeheartedly;

ii) ... it is **odd** that these others **should not have been tried out** until within our own time;

iii) The staff nurse said that she had thought it **strange** that this statement **should have been volunteered** 'out of the blue' moments after her mother had arrived;

iv) Besterman claimed that he was largely self-educated at the British Museum library and it is **typical** that his first separate publication **should have been** a bibliography of the well-known theosophist Annie Besant q.v. [1924].

Помимо глагола *suggest*, ряд частотных в академическом дискурсе глаголов можно ранжировать по степени уверенности: [i] глаголы, выражающие более высокую степень уверенности: *conclude, find\*, report, demonstrate, note, show\*, describe, present, state, explain, prove*;

и [ii] глаголы, оформляющие менее уверенные высказывания: *argue, imply, postulate, claim, indicate, propose, content, maintain, suggest\**, *hypothesize* [Conrad, Biber 2009]<sup>10</sup>.

Основными обучающими приемами работы с корпусными образцами являются сортировка, «вертикальное чтение» конкорданса для выявления грамматических особенностей и составление профиля слова, поиск коллокаций, трансформация информации – из конкорданса в таблицу или диаграмму – по определенным условиям. Для фиксированных выражений типа *bear ... resemblance* можно рекомендовать заполнение пропуска на основе конкорданса, поиск по которому в данном случае даст ограниченный список слов: *striking, strong, uncanny, little, no, any, some*.

Рекомендуется также проводить сравнение найденных вариантов из корпуса со словарной дефиницией.

### Выводы и заключение

В данной статье мы обратились к некоторым аспектам влияния зарубежной «корпусной революции» на российскую лингводидактику в плане существенного пересмотра содержания обучения языку.

Предложенные и проиллюстрированные здесь методы обучающих корпусных микроисследований и экспериментов, имеют неоспоримые достоинства как способ индивидуализации лингводидактических материалов и повышения информированности самого преподавателя.

Кроме того, необходимость таких экспериментов продиктована временем, поскольку высветились противоречия между эксплицитными, отраженными в правилах, и имплицитными, проявляющимися в реальном функционировании, языковыми системами. В то время как традиционные грамматики и искусственно составленные предложения отличаются «объясняемостью» и логичностью, реальные, сложившиеся в языковом сообществе употребления, хаотичны, не систематизированы и требуют изучения и осознания принципов их функционирования [Гвишиани, 2008].

Хорошую иллюстрацию этой мысли дает в своем корпусном исследовании И. В. Соловьева, изучившая реальный социокультурный статус извинения *pardon* в английском языке и пришедшая к выводу, что его интуитивный теоретический анализ от известного социолога К. Фокс страдает неточностью [Соловьева, 2019. С. 128–129].

**Вокабуляр.** Корпусные ориентиры успешно имплементированы в процессы уточнения количественных и качественных показателей в структуре запаса слов студента вуза, что поможет преодолеть рассогласованность в существующих у студентов и требуемых специалистом объемах словарного запаса академического, профессионального дискурса. Установлено, что язык относится к ранговым распределениям, и уже аналитически доказано, что 50 % любого корпуса текстов имеет единичное вхождение в этот корпус [Gorina et al., 2020]; следовательно, покрытие и понимание текстов растет медленно по отношению к количеству новых выученных слов. Поэтому начиная с запаса объемом 5–6 тыс. слов целесообразно инвестировать усилия в наращивание тщательно отобранной лексики по релевантным для обучаемого тематикам, в том числе профессиональной направленности. В связи с этим оправданными будут усилия по компиляции небольших профессионально-ориентированных баз данных,

<sup>10</sup> Звездочкой обозначены самые употребительные. Безусловно, здесь делается акцент не на предъявлении наиболее полного списка академических глаголов, а на том, что студент выступает в роли исследователя, учится отбирать «живые» языковые данные и систематизировать их самостоятельно. Т. Джонс, стоявший у истоков обучения с помощью лингвистических баз данных, отмечал, что задачей обучаемого является «совершать открытия» (discovery learning – эвристический метод) в иностранном языке. При этом преподаватель становится помощником, информатором, куратором исследования, обеспечивая контекст, в котором эти открытия происходят, и, развивая стратегии, с помощью которых обучаемый сможет «научиться учиться» (learn how to learn) [Johns, King, 1991].

которые могут составляться силами университетов для выявления наиболее частотной лексики, обеспечивающей достаточную покрываемость условного профессионального текста. На сегодня существуют и готовые корпус-менеджеры, работающие даже без разметки (WordSmith), которые дают набор корпусных инструментов для статистической оценки лингвистических баз данных. Однако этот лингводидактический потенциал корпусных инструментов зачастую остается нереализованным, в том числе и по причине инерции преподавательской среды.

*Регистр, когезия и коммуникативные стратегии.* Как показывает наш опыт, в студенческой среде регистровые различия часто не осознаются как ошибки, подобные орфографическим, грамматическим или пунктуационным. Более того, средства хеджирования как элементы вежливости и снижения безапелляционности высказываний автоматически в речь на изучаемом языке не включаются. Следовательно, необходимо акцентировать внимание на явлении хеджинга, которое, относясь к разряду активно используемых в стандартном английском языке, тем не менее, почти не используется обучаемыми. Решение задачи повышения осознанности упомянутых явлений может быть найдено в контрастивном анализе письменной и устной речи, корпусном поиске хеджей, проводимых аудиторно, на основе больших диверсифицированных корпусов. Сопоставление частотности и конкордансов – нарезок текстов корпуса по искомой единице – также эффективный прием для выявления особенностей употребления средств когезии, например, в академическом узусе, в том числе статистических.

*Расширение содержания обучения.* Обновление содержания обучения может состояться за счет расширения номенклатуры языковых средств и включения периферийных явлений, которые в диахроническом плане становятся более явственно выраженными и интересны обучаемым с точки зрения будущей профессии. Хорошим примером такой практики стал корпусный анализ двойной модальности М. А. Сальковой и О. А. Мачиной [2018], по результатам которого авторы рекомендовали знакомить будущих профессионалов с подобными явлениями.

Следует сказать, что индивидуализированные учебные материалы, основанные на аутентичных данных, вызывают интерес и доверие со стороны обучаемых, поскольку частотность и употребимость теперь доказуема и может быть верифицирована прямо в аудитории. Таким образом, умение использовать корпус как эффективную справочную систему снимает зависимость от языковой интуиции, в том числе преподавателя, а диалог «преподаватель – студент» приобретает новое направление, в котором актуализация знания происходит во время дискуссии, обсуждения, постановки задач, проверки гипотез. Таким образом, в действие вступает принцип, радикально отличающийся от подачи готового материала, не подлежащего обсуждению. Эффективность корпусных приемов заключается в сдвиге акцента с исследований, проведенных за студентов, на изыскания, которые ведут сами студенты и которые должны составлять неотъемлемую часть современной аудиторной практики.

### Список литературы

- 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
- Гвишиани Н. Б. Практикум по корпусной лингвистике: Учеб. пособие по английскому языку. М.: Высш. шк., 2008. 191 с. ISBN 978-5-06-005478-1.
- Горина О. Г. Использование технологий корпусной лингвистики для развития лексических навыков студентов-регионоведов: Дис. ... канд. пед. наук / МГУ. М., 2014. 325 с.

- Горина О. Г., Храброва В. Е.** Лингвистический хеджинг как коммуникативная стратегия (в русле корпусных исследований) // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2017. Т. 15, № 3. С. 44–53. DOI 10.25205/1818-7935-2017-15-3-44-53
- КСКТ – Краткий словарь когнитивных терминов / Под общ. ред. Е. С. Кубряковой. М.: Филол. ф-т МГУ им. М. В. Ломоносова, 1997. 245 с. ISBN 5-89042-018-6.
- Марюхин А. П.** Непрямая коммуникация в научном дискурсе (на материале русского, английского, немецкого языков): Автореф. дис. ... канд. филол. наук / Ин-т языкознания РАН. М., 2010. 24 с.
- Орлов Г. А.** Современная английская речь: Учеб. пособие для вузов по специальности «Английский язык и литература». М.: Высш. шк., 1991. 240 с.
- 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.
- Пумпянский А. Л.** Введение в практику перевода научно-технической литературы на английский язык. М.: Наука, 1981. 344 с.
- Рябцева Н. К.** Научная речь на английском языке: Руководство по научному изложению; словарь оборотов и сочетаемости общенаучной лексики; новый словарь-справочник активного типа. М.: Флинта: Наука, 1999. 601 с.
- Салькова М. А., Мачина О. А.** Двойная модальность и корпусный анализ: Актуализация обучения грамматическому аспекту современного английского языка // Вестник МГЛУ. Гуманитарные науки. 2018. Вып. 15 (810). С. 19–34.
- Скребнев Ю. М.** Введение в коллоквиалистику / Под ред. О. Б. Сиротининой. Саратов: Изд-во Сарат. ун-та, 1985. 210 с.
- Соловьева И. В.** Социокультурный статус формулы извинения *pardon* в английском языке по данным корпусных баз // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2019. Т. 17, № 1. С. 125–133. DOI 10.25205/1818-7935-2019-17-1-125-133
- Vaayen, R. H.** Word frequency distributions. Dordrecht, The Netherlands, Kluwer Academic Publ., 2002, 356 p. ISBN 1402009275.
- Bartlett, F. C.** Remembering: A Study in Experimental Social Psychology. Cambridge, Cambridge Uni. Press, 1932, 317 p. URL: <http://www.bartlett.psychol.cam.ac.uk/RememberingBook.htm> (accessed: 12.03.2020).
- Bernadini, S.** Corpora in the classroom: An overview and some reflections on future developments. In: Sinclair, J. McH. (ed.). How to use corpora in language teaching. Amsterdam [u.a.], Benjamins, 2004, 308 p. (Studies in Corpus Linguistics. Vol. 12). ISBN 978-90-272-2283-1.
- Bestgen, Y.** Comparing Lexical Bundles across Corpora of Different Sizes: The Zipfian Problem. *Journal of Quantitative Linguistics*, 2020, vol. 27, no. 3, p. 272–290 DOI 10.1080/09296174.2019.1566975
- Biber, D.** Using multi-dimensional analysis to explore cross-linguistic universals of register variation. In: Lefer, M.-A., Vogeleer, S. (eds.). Genre- and Register-related Discourse Features in Contrast. Amsterdam [Phil.], John Benjamins Publishing Company, 2016, p. 7. (Benjamins Current Topics. Vol. 87).
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E.** Longman Grammar of Spoken and Written English. S. 1.: Pearson Education Limited, 1999. 1204 p. ISBN 0-582-23725-4.
- Carter, R., McCarthy, M.** Cambridge Grammar of English: A comprehensive guide: Spoken and Written English Grammar and Usage. 6<sup>th</sup> ed. S. 1.: Cambridge Uni. Press, 2011. 973 p. ISBN 978-0-521-85767-3.
- Conrad, S., Biber, D.** Real Grammar: A Corpus-Based Approach to English. Pearson, Longman, 2009, 150 p.
- Crystal, D.** On Keeping One's Hedges in Order. *English Today*, 1988, vol. 4. no. 3 (15), p. 46–47.
- Farr, F., Murphy, B., O'Keeffe, A.** The Limerick corpus of Irish English: design, description and application. In: Farr, F., O'Keeffe, A. (eds.). Corpora, Varieties and the Language Classroom. Special Edition of Teanga 21. Dublin, IRAAL, 2004, p. 5–29.

- Flowerdew, L.** Corpora and Language Education. S. l., Palgrave Macmillan, 2012, 347 p.
- Gorina, O. G., Tsarakova, N. S., Tsarakov, S. K.** Study of Optimal Text Size Phenomenon in Zipf-Mandelbrot's Distribution on the Basis of Full and Distorted Texts. Author's Frequency Characteristics and derivation of Hapax Legomena. *Journal of Quantitative Linguistics*, 2020, vol. 27, no. 2, p. 134–158, DOI 10.1080/09296174.2018.1559460
- Huddleston, R. D., Pullum, G. K.** The Cambridge Grammar of the English Language. Cambridge Uni. Press, 2002. ISBN 0-521-43146-8.
- Hyland, K.** Academic Discourse. London, Continuum, 2009.
- Johns T., King, Ph.** Should you be persuaded – two samples of data-driven Learning materials. *Classroom Concordancing: ELR Journal*, 1991, vol. 4, p. 1–16. URL: <http://www.lexically.net> (accessed: 17.03.2021).
- Lewis, M.** Implementing the lexical approach: Putting theory into practice. Heinleceange, 2008, 223 p.
- Mauranen, A.** 'But here's a flawed argument': socialization into and through metadiscourse. In: Leistyna, P. and Meyer, C. (eds.). *Corpus Analysis: Language Structure and Language Use*. Amsterdam, Rodopi, 2003a, p. 19–34.
- Mauranen, A.** The corpus of English as lingua franca in academic settings. *TESOL Quarterly*, 2003b, no. 37 (3), p. 513–527.
- O'Keeffe, A., McCarthy, M., & Carter, R.** From corpus to classroom: Language use and language teaching. Cambridge, Cambridge Uni. Press, 2007.
- Salazar, D.** Lexical Bundles in Native and Non-native Scientific Writing: Applying a corpus based study to language teaching. Amsterdam, Phil., John Benjamins Publishing Company, 2014, ix, 212 p. (Studies in Corpus Linguistics. Vol. 65)
- Scott, M., Tribble, C.** Textual Patterns: key words and corpus analysis in language education: Studies in Corpus Linguistics. Amsterdam, Philadelphia, John Benjamins, 2006, 200 p.
- West, R. et al.** The Word Family Framework [WFF]. 2013. URL: <http://www.teachingenglish.org.uk/article/word-family> (accessed 15.02.2019).

### References

- Baayen, R. H.** Word frequency distributions. Dordrecht, The Netherlands, Kluwer Academic Publ., 2002, 356 p. ISBN 1402009275.
- Bartlett, F. C.** Remembering: A Study in Experimental Social Psychology. Cambridge, Cambridge Uni. Press, 1932, 317 p. URL: <http://www.bartlett.psychol.cam.ac.uk/RememberingBook.htm> (accessed: 12.03.2020).
- Bernadini, S.** Corpora in the classroom: An overview and some reflections on future developments. In: Sinclair, J. McH. (ed.). *How to use corpora in language teaching*. Amsterdam [u.a.], Benjamins, 2004, 308 p. (Studies in Corpus Linguistics. Vol. 12). ISBN 978-90-272-22831.
- Bestgen, Y.** Comparing Lexical Bundles across Corpora of Different Sizes: The Zipfian Problem. *Journal of Quantitative Linguistics*, 2020, vol. 27, no. 3, p. 272–290 DOI 10.1080/09296174.2019.1566975
- Biber, D.** Using multi-dimensional analysis to explore cross-linguistic universals of register variation. In: Lefer, M.-A., Vogeleeer, S. (eds.). *Genre- and Register-related Discourse Features in Contrast*. Amsterdam [Phil.], John Benjamins Publishing Company, 2016, p. 7. (Benjamins Current Topics. Vol. 87).
- 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. ISBN 0-582-23725-4.
- Brief dictionary of cognitive science terms. Ed. by E. S. Kubryakova. Moscow, MSU Press, 1997, 245 p. (in Russ.)



- Carter, R., McCarthy, M.** Cambridge Grammar of English: A comprehensive guide: Spoken and Written English Grammar and Usage. 6<sup>th</sup> ed. S. l.: Cambridge Uni. Press, 2011. 973 p. ISBN 978-0-521-85767-3.
- Conrad, S., Biber, D.** Real Grammar: A Corpus-Based Approach to English. Pearson, Longman, 2009, 150 p.
- Crystal, D.** On Keeping One's Hedges in Order. *English Today*, 1988, vol. 4. no. 3 (15), p. 46–47.
- Farr, F., Murphy, B., O'Keeffe, A.** The Limerick corpus of Irish English: design, description and application. In: Farr, F., O'Keeffe, A. (eds.). *Corpora, Varieties and the Language Classroom*. Special Edition of Teanga 21. Dublin, IRAAL, 2004, p. 5–29.
- Flowerdew, L.** Corpora and Language Education. S. l., Palgrave Macmillan, 2012, 347 p.
- Gvishiani, N. B.** Workshop on Corpus Linguistics: a Textbook. Moscow: Higher school, 2008, 191 p. ISBN 978-5-06-005478-1.
- Gorina O. G.** Using corpus linguistics tools in teaching L2 lexical skills to Area Studies students. Cand. Ped. Sci. Dis. Moscow, 2014, 325 p. (in Russ.)
- Gorina, O. G., Khrabrova, V. E.** Linguistic Hedging as a Communicative Strategy (within the Scope of Corpus Studies). *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2017, vol. 15, no. 3, p. 44–53. (in Russ.) DOI 10.25205/1818-7935-2017-15-3-44-53
- Gorina, O. G., Tsarakova, N. S., Tsarakov, S. K.** Study of Optimal Text Size Phenomenon in Zipf-Mandelbrot's Distribution on the Basis of Full and Distorted Texts. Author's Frequency Characteristics and derivation of Hapax Legomena. *Journal of Quantitative Linguistics*, 2020, vol. 27, no. 2, p. 134–158, DOI 10.1080/09296174.2018.1559460
- Huddleston, R. D., Pullum, G. K.** The Cambridge Grammar of the English Language. Cambridge University Press, 2002. ISBN 0-521-43146-8.
- Hyland, K.** Academic Discourse. London, Continuum, 2009.
- Johns T., King, Ph.** Should you be persuaded – two samples of data-driven Learning materials. *Classroom Concordancing: ELR Journal*, 1991, vol. 4, p. 1–16. URL: <http://www.lexically.net> (accessed: 17.03.2021).
- Lewis, M.** Implementing the lexical approach: Putting theory into practice. Heinleceengage, 2008, 223 p.
- Maryukhin, A. P.** Indirect communication in academic discourse (based on Russian, English and German languages). Cand. Philol. Sci. Thesis. Institute of Linguistics RAS. Moscow, 2010, 24 p. (in Russ.)
- Mauranen, A.** 'But here's a flawed argument': socialization into and through metadiscourse. In: Leistyna, P. and Meyer, C. (eds.). *Corpus Analysis: Language Structure and Language Use*. Amsterdam, Rodopi, 2003a, p. 19–34.
- Mauranen, A.** The corpus of English as lingua franca in academic settings. *TESOL Quarterly*, 2003b, no. 37 (3), p. 513–527.
- O'Keeffe, A., McCarthy, M., Carter, R.** From corpus to classroom: Language use and language teaching. Cambridge, Cambridge Uni. Press, 2007.
- Orlov, G. A.** Modern English: a textbook for university students of English and Literature Studies Moscow, Higher School Publ., 1991, 240 p. (in Russ.)
- Plungyan, V. A.** Corpus as a Tool and as an Ideology: On Some Lessons of Modern Corpus Linguistics. *Russian language in scientific coverage*, 2008, no. 2 (16), p. 7–20. (in Russ.)
- Pumpyansky, A. L.** Introduction to the practice of translating scientific and technical literature into English. Moscow, Nauka, 1981, 344 p. (in Russ.)
- Ryabtseva, N. K.** English for Scientific Purposes: A Guide to Scientific Writing; Dictionary of Academic Expressions and Collocations; a New Dictionary-Reference Activator. Moscow, Flinta: Nauka, 1999, 601 p. (in Russ.)
- Salazar, D.** Lexical Bundles in Native and Non-native Scientific Writing: Applying a corpus based study to language teaching. Amsterdam, Phil., John Benjamins Publishing Company, 2014, ix, 212 p. (Studies in Corpus Linguistics. Vol. 65)

- Salkova, M. A., Machina, O. A.** Double Modality and Corpus-Based Analysis: Actualization of Teaching Modern English Grammar. *Vestnik of Moscow State Linguistic University. Humanities Sciences*, 2018, iss. 15 (810), p. 19–34. (in Russ.)
- Scott, M., Tribble, C.** Textual Patterns: key words and corpus analysis in language education: Studies in Corpus Linguistics. Amsterdam, Philadelphia, John Benjamins, 2006, 200 p.
- Skrebnev, Yu. M.** Introduction to Colloquialistics. Ed. by O. B. Sirotinina. Saratov, Saratov Uni. Press, 1985, 210 p. (in Russ.)
- Soloveva, Inna V.** Language Corpora Data as the Source of Sociocultural Information about *Pardon* Apology. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 1, p. 125–133. (in Russ.) DOI 10.25205/1818-7935-2019-17-1-125-133
- Vlasova, E. A., Karpova, E. L., Olshevskaya, M. Yu.** Vocabulary: How Many Words Are Enough? Principles of Minimizing Learners' Vocabulary. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2019, vol. 17, no. 4, p. 63–77. (in Russ.) DOI 10.25205/1818-7935-2019-17-4-63-77
- West, R. et al.** The Word Family Framework [WFF]. 2013. URL: <http://www.teachingenglish.org.uk/article/word-family> (accessed 15.02.2019).

Материал поступил в редколлегию  
Date of submission  
26.09.2020

### Сведения об авторах / Information about the Authors

**Горина Ольга Григорьевна**, кандидат педагогических наук, доцент департамента иностранных языков Санкт-Петербургской Школы гуманитарных наук и искусств, Национальный исследовательский университет «Высшая школа экономики» (Санкт-Петербург, Россия)

**Olga G. Gorina**, Candidate of Sciences (PhD) in Theory and Methods of Education, Associate Professor of the Department of Foreign Languages, St. Petersburg School of Arts and Humanities, National Research University “Higher School of Economics” (St. Petersburg, Russian Federation)

gorina@bk.ru

ORCID 0000-0002-9184-326X

**Царакова Наталья Сергеевна**, тьютор департамента иностранных языков Санкт-Петербургской Школы гуманитарных наук и искусств, Национальный исследовательский университет «Высшая школа экономики» (Санкт-Петербург, Россия)

**Natalya S. Tsarakova**, Tutor of the Department of Foreign Languages, St. Petersburg School of Arts and Humanities, National Research University “Higher School of Economics” (St. Petersburg, Russian Federation)

ntcarakova@hse.ru