

О возможности использования корпуса NOW в курсе английского для специальных целей для студентов специальности «Биотехнология»

И.В. Кузнецова¹, М.С. Коган²

¹ Университет ИТМО,

² Санкт-Петербургский политехнический университет Петра Великого

ingakuznetsova2014@mail.ru, kogan_ms@spbstu.ru

Аннотация

В статье исследуются возможности применения лингвистических корпусов для обучения английскому для специальных целей (АСЦ) студентов – биотехнологов. Доказывается целесообразность поиска релевантного корпуса, находящегося в свободном доступе, по сравнению с созданием собственного специального корпуса для данной предметной области. Обосновывается целесообразность исследования/проверки релевантности корпусов COCA и NOW как дополнительного ресурса в курсе АСЦ для направления подготовки «Биотехнология». В качестве критериев релевантности корпуса были выбраны следующие: 1) наличие в корпусе биотехнологических терминов, 2) их частотность и 3) доступность полных текстов, фрагменты которых представлены в конкордансах с терминами по биотехнологии. В ходе эксперимента подготовленные с помощью корпуса NOW задания были опробованы в группе бакалавров-биотехнологов 3-го курса, изучающих АСЦ. Рассматриваются примеры заданий, используемые в эксперименте. Это задания с раздаточными материалами, подготовленными преподавателем на основе материалов корпуса, а также задания на прямое использование корпуса студентами – биотехнологами для решения таких задач как поиск коллокатов, выбор правильного артикля, поиск и анализ нетипичных форм множественного числа биотехнологических терминов и др.

Показано, что студенты успешно применяли полученные в ходе работы с корпусом знания для развития продуктивных навыков (говорение, письмо) и своих аналитических способностей. Отзывы студентов об экспериментальном курсе в целом положительные. В заключение рассматриваются возможные направления дальнейших исследований в области применения лингвистических корпусов в курсе АСЦ.

Ключевые слова: английский для специальных целей, биотехнология, лингвистические корпуса, контекст, освоение специальной лексики

Библиографическая ссылка: Кузнецова И.В., Коган М.С. О возможности использования корпуса NOW в курсе английского для специальных целей для студентов специальности «Биотехнология» // Компьютерная лингвистика и вычислительные онтологии. Выпуск 4 (Труды XXIII Международной объединенной научной конференции «Интернет и современное общество», IMS-2020, Санкт-Петербург, 17 – 20 июня 2020 г. Сборник научных статей). — СПб.: Университет ИТМО, 2020. С. 29-43. DOI: 10.17586/0000-0000-2020-4-29-43

Введение

Одной из наиболее существенных проблем, с которой сталкиваются преподаватели при разработке курса обучения английскому для специальных целей в любой профессиональной области, является отсутствие учебников со специальной лексикой.

В немногих учебниках по овладению специальной лексикой в области биотехнологий, представленных сегодня на рынке образовательной литературы, часто используются тексты устаревшей тематики, рассматривающие технологические процессы давно усовершенствованные и измененные в реальной жизни, а поэтому не отражающие современную лексику для описания биотехнологий и использующие устаревшие термины. Более того, зачастую упражнения в таких учебниках разработаны с применением грамматико-переводного метода, что не отвечает современным требованиям к образовательному процессу, основанному на коммуникативном методе обучения иностранном языкам. Коммуникативный метод предполагает повышение мотивации студентов и использование подходов, ставящих в центр самого обучающегося, а не преподавателя, который принимает на себя роль фасилитатора в процессе обучения. Федеральный государственный образовательный стандарт высшего профессионального образования (ФГОС ВПО) по направлению 240700 «Биотехнология» [1] предполагает умение выпускника использовать иностранный язык в межличностном общении и профессиональной деятельности, а также владеть навыками выражения своих мыслей и мнения в межличностном и деловом общении, что предполагает необходимость поиска новых методов обучения, повышающих уровень автономии студентов.

Лингвистические корпусы являют собой собрание огромного количества текстов и предоставляют обучающимся достаточное количество примеров употребления носителями языка тех или иных терминов по специальности и коллокаций. Это является существенным преимуществом использования корпусов для обучения специальному английскому, так как студенты часто испытывают трудности с правильным подбором слов в устном и письменном общении из-за отсутствия достаточного опыта применения специальной лексики, а также интерференции терминов родного языка в иностранную речь.

Ряд исследований показали, что использование корпусов в целях обучения иностранному языку способствует развитию автономии и мотивации студентов [2; 3]. Профессиональная специфика лексики дисциплины Биотехнология обнаруживает большое количество слов, связанных с химией, биологией и материаловедением, которые будучи легко узнаваемы в письменной форме, имеют совершенно другое значение и/или иначе произносятся в английском языке. Поэтому умение студента самостоятельно обратиться к корпусу для исследования значений слова и его коллокаций очень важно для развития лексических навыков.

Однако при использовании корпусов для модернизации и усовершенствования курса обучения, возникает проблема выбора подходящего лингвистического корпуса, в котором содержались бы необходимые термины для обучения специальной лексике. Отсутствие специальных корпусов по многим научным и техническим направлениям делает логичным создание собственных небольших корпусов известных как Do It Yourself (DIY) corpora по определенной тематике. По этому пути идут некоторые исследователи, обучающие студентов применять методы корпусного анализа к имеющемуся у них собранию научных статей по их специальности для развития навыков написания научных статей [4; 5; 6], освоения незнакомой лексики и т.п. [7]. Однако, трудоемкость создания специальных корпусов преподавателем для студентов определенного направления подготовки, о которой также говорят исследователи, (см., например, [8; 9, p. 163]) вынуждает обратиться к поиску других решений этой проблемы, а именно, к поиску подходящего корпуса доступного в интернете.

В нашей работе мы сравнили преимущества и возможности двух корпусов: COCA [10] и NOW [11], созданных исследователями университета Бригама Янга под руководством М. Дэвиса. Корпус современного американского варианта английского языка (Corpus of Contemporary American [English]) является большим, структурированным, постоянно пополняемым корпусом общего назначения [12]. Он, охватывает тексты разных типов и жанров; содержит около одного миллиарда словоупотреблений, встречаемых в более чем двухстах тысячах текстов устной и письменной речи разных стилей и жанров. Ежегодно с 1990 г. корпус пополнялся двадцатью миллионами лексем.

Корпус NOW (News on the Web) был создан в 2016 году [13]. На сегодняшний день корпус NOW содержит 10 млрд слов, начиная с 2010г, включая новые примеры использования профессиональных терминов. За 4 года его объем увеличился в 3 раза. До появления корпуса NOW, корпус COCA применялся для обучения специальному английскому как студентов языковых вузов, так и специалистов неязыковых профилей. Некоторые исследователями подтверждают пригодность использования корпуса COCA для разработки упражнений, способствующих формированию «информационно-технологической составляющей иноязычной компетенции» при подготовке специалистов неязыкового вуза [14]. Л.П. Тарнаева и Е.С. Осипова высоко оценили возможности COCA для развития идиоматической компетенции при обучении переводчиков, расширения переводческого кругозора, развития лингвистического мышления, формирования навыков самостоятельной работы с информационно-коммуникационными ресурсами [15]. Ряд зарубежных исследований посвящен оценке возможностей COCA для обучения профессиональному английскому [16; 17]. Примеры использования корпуса NOW в качестве дополнительного ресурса к учебникам для студентов в сфере IT и переводчиков в сфере профессиональной коммуникации даны в работах [18; 19; 20].

Очевидно, что у каждого корпуса есть свои особенности, которые необходимо исследовать, чтобы решить, какой из корпусов больше подходит для использования в курсе английского для специальных целей для биотехнологов. Для этого необходимо ответить на следующие вопросы:

- 1) Какова частотность выбранных терминов в каждом корпусе.
- 2) Какова доступность исходных текстов, являющихся источниками минимальных контекстов, содержащих заданный термин, которые выдаются по запросу.

После этого нужно разработать задания по использованию корпуса/корпусов в учебном процессе и провести их апробацию.

1. Метод

Для сравнения возможностей применения корпусов COCA и NOW был проведен анализ частотности терминов специфических для области биотехнология в обоих корпусах. Затем мы проверили доступность полных текстов, источников минимальных контекстов с помощью имеющихся в корпусе мета-данных.

После этого были разработаны упражнения с использованием выбранного корпуса.

Упражнения были апробированы в течение семестра в рамках курса Английский язык для специальных целей в группе студентов третьего курса ИТМО по специальности «Биотехнология». В экспериментальном обучении участвовало 10 человек. В контрольной группе, осваивавшей стандартную программу без элементов корпусной лингвистики, было 12 студентов. Исходный уровень владения английским языком в обеих группах был B2-C2 по шкале CEFR. Возраст участников эксперимента колебался в пределах 20 – 23 года. В конце семестра для обеих групп было проведено тестирование по изученному материалу и опрос студентов для выяснения их мнения об использовании корпусных подходов в курсе.

1.1. Анализ частотности терминов предметной области Биотехнология в корпусах COCA и NOW

Для ответа на 1-й вопрос был проведен поиск текста, максимально обеспечивающего репрезентативность вокабуляра необходимого для освоения лексики по дисциплине английский для специальных целей для биотехнологов. Для этого в поисковик Google были введены термины «biotechnology», «chemical industry». В числе первых ссылок оказалась ссылка на текст Biotechnology in the chemical industry на сайте The essential chemical industry online [21]. Текст состоит из 1270 слов. Было решено провести его анализ с помощью инструмента Vocabulary Profile на сайте Т. Кобба Compleat Lexical Tutor. Инструмент позволяет получать как графическое распределение слов по рангам /категориям частотности (рис.1), так и табличное представление слов с указанием частотного диапазона, включая слова, не найденные в объединенном частотном распределении слов из корпусов COCA + BNC, выбранном для анализа.

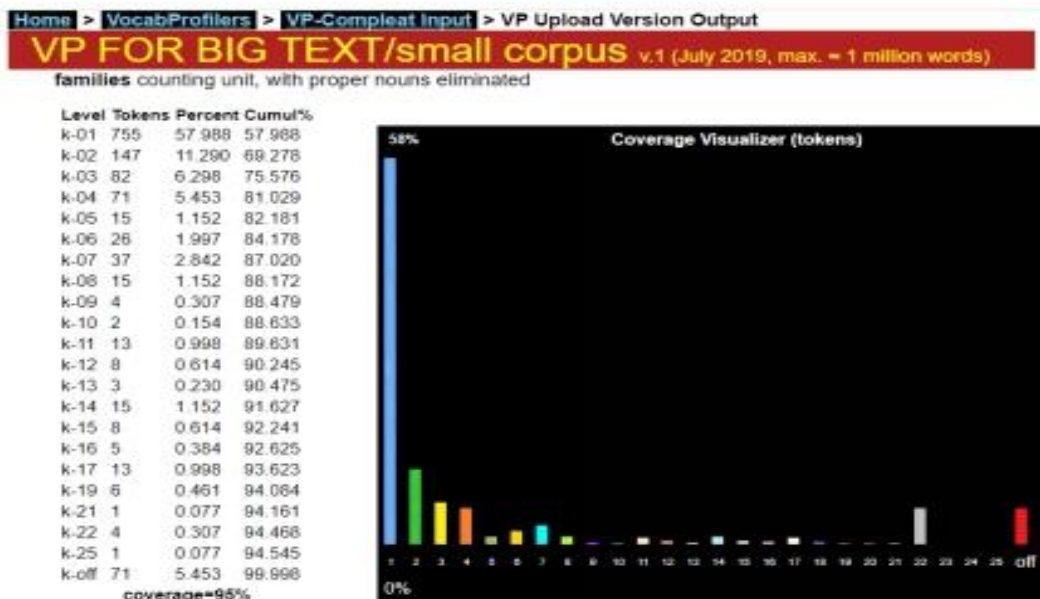


Рис. 1. Графическое представление результатов анализа текста с помощью инструмента VocabProfiler на сайте CompleatLexicalTutor [22]

Наше внимание привлекли слова, не обнаруженные в корпусе. Их оказалось сравнительно немного: 5,5% (71 слово). Имен собственных среди них не было. Явных опечаток, возможно, случившихся при изменении форматов (html □ word.doc □ txt □ VocabProfilers) мы обнаружили 6. Например, система распознала как одно слово выражение thedirectproduction*.

Из оставшихся слов были отобраны 10. К ним были добавлены 5 словосочетаний (4, состоящие из 2-х слов и одно, состоящее из 3-х слов): citric acid, lactic acid, calcium hydroxide, calcium citrate, fungus Aspergillus niger. Инструмент VocabProfilers работает по принципу подсчета частотности отдельных слов (Word-by-Word Count Index), обоснованный создателями ресурса в статье 2013 г. [23]. Это приводит к тому, что устойчивые словосочетания или компаунды разбиваются на отдельные слова или выносятся в список off-list. Проверка в больших корпусах типичных словосочетаний из конкретной предметной области важна, т.к. как мы показали, корпус может содержать большое количество примеров употребления каждого слова из словосочетания по отдельности (turbine head) и не содержать примеров употребления словосочетания,

которое означает напор турбины [7] и является очень распространенным термином в гидроэнергетике¹.

Для проверки частотности из off-list списка были отобраны десять терминов, состоящих из одного слова, 4 выражения, содержащих два слова, и один термин из трех слов. Полученные в ходе анализа частотности этих слов и выражений в корпусах COCA и NOW представлены в табл. 1.

Таблица 1. Сравнение частотности специальной лексики по теме «Биотехнология» в корпусах COCA и NOW

№	Слово/выражение	Частотность в корпусе	
		COCA	NOW
1	acidulant	1	6
2	lactobacillus	186	1594
3	alkoxylates	0	18
4	diol	17	103
5	equimolar	19	151
6	dicarboxylic	15	65
7	glutamicum	3	27
8	octanoate	2	9
9	polytrimethylene	0	6
10	bacterium clostridium	12	86
11	lactic acid	366	2810
12	calcium hydroxide	31	116
13	calcium citrate	21	55
14	citric acid	196	1257
15	fungus Aspergillus	1	5

В данной выборке 6 слов и выражений имели очень низкую встречаемость в корпусе COCA (меньше 4), при этом термины *alkoxylate* и *polytrimethylene* не был найден в корпусе ни в какой форме. Все термины были найдены в корпусе NOW, хотя наименее частотные/узкоспециальные встречались 5-6 раз. Для сравнения двух корпусов нами был выбран фактор частотности, так как он особенно важен для преподавателя английского языка, обращающегося к корпусу. Чем больше частотность употребления слова, тем больше примеров можно извлечь из конкорданса. Как видно из таблицы 1, корпус NOW содержит гораздо больше примеров употребления профессиональной лексики чем корпус COCA.

При выполнении учащимися самостоятельных поисковых заданий у них появляется больший по сравнению с корпусом COCA выбор для обнаружения контекста наиболее близкого к интересующей их области. Это в свою очередь должно способствовать лучшему усвоению ими лексических единиц, так как согласно исследованиям Т. Кобба с соавторами для запоминания нового слова с ним необходимо встретиться в различных контекстах от 6 до 10 раз [24].

1.2. Анализ источников, содержащих заданный термин

Во второй части эксперимента мы проанализировали результаты сравнения источников текстов, полученных с помощью корпуса NOW и COCA по запросу одного и того же термина. В качестве иллюстрации разницы полученных источников текстов между корпусами NOW и COCA, было выбрано существительное *plasmid* – слово, являющееся одним из ключевых слов и отражающее тему одного из самых важных разделов курса

¹ см. подробнее на сайте <https://mash-xxl.info/info/528972/> (дата обращения: 10.03.2020)

обучению специальному английскому в области биотехнологии. В глоссарии биотехнологических терминов [25] оно употребляется более 20 раз в объяснении разных терминов.

Результаты запроса слова *plasmid* в корпусе NOW приводятся на рис. 2.

SEARCH	FREQUENCY	CONTEXT	ACCOUNT
1 19-12-28 US advances.sciencemag.org	A B C	actin (CBA) promotor was performed as described previously (13). The plasmid DNA of both eGFP-2a-huTau constructs and GFP were then tested for in vivo	
2 19-12-28 US advances.sciencemag.org	A B C	. Additional data related to this paper may be requested from the authors. The plasmid encoding reporter 2A tau constructs can be provided by B.T.H. pe	
3 19-12-28 US nature.com	A B C	achieve equal amount of Beclin-1 protein level at time point 0, amounts of Flag-Beclin-1 plasmid transfected between empty vector and LAT1 were 3:1. P	
4 19-12-27 US Nature	A B C	The HDR-HygB repair template, containing the hygromycin B resistance cassette (HygB) from plasmid pTK079, was constructed by fusing the HygB casse	
5 19-12-27 US Nature	A B C	, according to the manufacturer's instructions. Positive clones were sequenced and the resulting plasmid designated pHygR-FUM1. The HDR-HygB repair	
6 19-12-27 US Nature	A B C	plasmid designated pHygR-FUM1. The HDR-HygB repair template for microhomology recombination was PCR amplified from plasmid pHygR-FUM1 with p	
7 19-12-27 US Nature	A B C	of helper Vero cells with in vitro RNA transcripts synthesized using T7 RNA polymerase or plasmid DNAs engineered to contain the eukaryotic cytomegalo	
8 19-12-27 US Nature	A B C	in?~7.9). consistent with literature reports(17,57,58,59). The human Eg5-513 plasmid , a gift from Susan Gilbert (Rensselaer Polytechnic Institute) 33 was tran	
9 19-12-27 US Nature	A B C	thanic Susan Gilbert (Rensselaer Polytechnic Institute) for the gift of the Eg5-513 expression plasmid . Timothy Mitchison (Harvard University) for the stable	
10 19-12-27 US PLOS	A B C	synthesized (JDT, Iowa, USA). This fragment was cloned into a plasmid vector (PGEM-T Easy Vector System – Promega, Madison, USA) according to	
11 19-12-27 US PLOS	A B C	region23. # For standard curve construction we used in vitro transcribed RNA instead of plasmid DNA as DNA amplification efficiency do not always corre	
12 19-12-27 US PLOS	A B C	(ATCC, Manassas, VA, USA). E. coli DH5a carrying GFP-expressing plasmid (pBBR122 encoding gfp) was also used for this study 27. E. coli	
13 19-12-27 US PLOS	A B C	on LB agar with chloramphenicol (10 ⁷ g per ml) to maintain the plasmid . Pseudomonas aeruginosa ATCC29213 and Enterococcus faecalis ATCC29212 wer	
14 19-12-27 US PLOS	A B C	primers and probes sequences used in this study are listed in Table 1. Plasmid cloning and RNA in vitro transcription # A double-stranded DNA fragment	
15 19-12-27 US PLOS	A B C	was used as the E. coli host for construction and propagation of cDNA clones. Plasmid extraction was performed using Pure Link Quick Plasmid Miniprep	

Рис. 2. Результаты поиска слова “*plasmid*” в корпусе NOW

При анализе результатов запроса обращает на себя внимание тот факт, что большинство текстов – недавно появившиеся публикации, самые свежие из которых появились не более двух месяцев тому назад. (По другим запросам были примеры, пополнившие корпус NOW буквально накануне). Тексты же, найденные с помощью корпуса COCA по данному запросу, были опубликованы достаточно давно: нам не удалось найти примеры позже 2012 года.

Второй важный аспект – легкий переход к полному тексту, из которого была взята строка конкорданса, по гиперссылке названия журнала. Например, первая строка конкорданса на скриншоте, приведенного выше, ведет к статье в журнале Science Advances [26]. Как видно из приведенных на рис. 2 результатов запроса по поиску слова “*plasmid*”, подавляющее число источников текстов корпуса NOW являются текстами научных статей из научных журналов. Первые 100 строк конкорданса представляют собой выдержки из текстов научных журналов Science Advances, Nature Communications, Scientific Reports, и Public Library of Science (PLOS). При проведении аналогичного поиска в корпусе COCA, мы столкнулись с невозможностью перейти к полному тексту из расширенного контекста, даже при наличии гиперссылки на источник информации.

Как и в случае с терминами, приведенными в табл.1, количество примеров слова *plasmid* в корпусе NOW намного больше чем в COCA (2967 против 476). На основе полученных данных можно сделать вывод, что корпус NOW 1) содержит терминологию, обязательную для усвоения студентами, обучающимися по специальности «Биотехнология»; 2) предоставляет доступ к гораздо большему количеству полных текстов статей, недавно опубликованных в научных журналах. Это говорит в пользу выбора корпуса NOW.

2. Разработка заданий на прямое и не прямое использование корпуса NOW в курсе АСЦ для обучающихся по направлению «Биотехнология»

Далее мы приводим разработанные задания на прямое и не прямое использование NOW в курсе обучения специальному английскому студентам неязыкового вуза по направлению подготовки «Биотехнология».

Поскольку использование корпусов предполагает определенную техническую подготовку, на начальном этапе целесообразно рассматривать работу с раздаточными материалами, содержащими задания, составленные преподавателем (подход, получивший название в специальной литературе *hands-off*). Авторы немногочисленных работ, посвященных изучению английского для специальных целей с использованием подходов корпусной лингвистики, показывают, что прежде чем студенты научатся самостоятельно эффективно работать с корпусными ресурсами при овладении профессиональным английским, им необходимо поработать с упражнениями, разработанными преподавателем [26, 27].

2.1. Примеры заданий, разработанных преподавателем на основе корпусных данных

Типичными заданиями, которые можно разработать, используя корпус, являются упражнения на сопоставление частей коллокаций (*Match collocations*), заполнение пробелов в тексте словом из предложенного списка (*Multiple choice*), либо без предлагаемых вариантов (*Open cloze*). Упражнение на словообразование (*Word formation*) является наиболее сложным для выполнения учащимися, когда для заполнения пробела требуется видоизменить одну из частей коллокации, предложенной преподавателем. Однако, его можно достаточно просто подготовить с помощью корпуса, если использовать функцию поиска по лемме (*lemma*) и «по маске» (*wild card*).

Задачей более продвинутого этапа работы с корпусами является обучение учащихся самостоятельному обращению к корпусам для анализа правил на употребление лексики, предлогов, и коллокаций профессионального английского, используемого в их деятельности.

2.2. Примеры заданий на прямое использование корпуса обучающимися

2.2.1. Первое обращение к корпусу

При первом прямом обращении к корпусу важно создать у обучающихся стабильную позитивную реакцию и положительные эмоции при работе с этим сложным инструментом. Важно показать студентам, что работа с корпусом имеет практическое применение для решения вопросов, возникающих при изучении иностранного языка.

К таким вопросам относится, например, выбор предлога в разных предложных конструкциях. Студентам было предложено проверить, какие предлоги употребляются в выражениях типичных для резюме или сопроводительного письма при поиске работы по специальности (инженер-биотехнолог): *apply for the position; working under the pressure; was in charge for; creating chemicals from renewable biomass, etc.* Поиск в корпусе целесообразно проводить с использованием поиска по лемме глагола, используя поле *Word/phrase*, и не использовать поле *Collocates*, так как в этом случае результат будет содержать больше шума и не даст возможность сформировать запрос для поиска выражения, в состав которого входят 2 предлога, например, «*was in charge for*». Отсутствие вариантов для выражения типа *put biofuels on an equal footing with fossil fuels* (запрос: *PUT biofuels _i* an equal footing _i* fossil fuels*) делает логичным формулирование запроса в более общих терминах с использованием символа *** для любого слова или *_nn** для любого существительного вместо *biofuels*. Анализ полученных результатов позволяет

сделать однозначный вывод о предлогах в выражении *on equal footing with* и обратить внимание на малочисленность примеров, содержащих такую длинную фразу, даже в очень большом корпусе. Для запроса `PUT_nn*_i* an equal footing_i*` мы получили всего 20 примеров, большинство из которых встретились в корпусе NOW один раз.

2.2.2. Изучение основных функций корпуса для прямого обращения к корпусам. Развитие лексических навыков

Следующая серия заданий направлена на уточнение значений терминов путем обращения к расширенному контексту или полному тексту статей, в которых они встретились, поиск и анализ синонимов (запрос в поле `Word/phrase` выглядит так: `[=ingestion]`) и коллокатов.

Для начинающих пользователей, какими являются студенты-биотехнологи, мы считаем принципиально важным в формулировке задания указывать по крайней мере один способ его выполнения. Например, «Enter one of the nouns for example, “Preservative” and click COLLOCATES, then click PoS and choose “adj. ALL” from the dropdown menu. Click 2 on Left and then “Find matching strings” to see examples of adjectives used with this word. Which 5 adjectives are used most frequently with this noun? Click on each of them to see the context. Note down your findings». Студенты должны были найти типичные сочетания терминологических единиц с прилагательными и глаголами.

Для глаголов, описывающих процесс ферментации (*aerate, oxidize, slacken, hasten, halt, esterify, dehydrate, etc*), было предложено найти существительные, наречия и предлоги, с которыми они образуют коллокаты, и представить результаты поиска в виде таблицы. Если внимательные студенты предложат альтернативные варианты формулировки запроса – это может стать предметом обсуждения на занятии, как и сравнение полученных вариантов при разных способах выполнения задания [28].

2.2.3. Задания на прямое обращение к корпусу NOW для решения грамматических задач

Корпуса позволяют привлечь внимание учащихся к некоторым грамматическим трудностям, которым уделяется недостаточное внимание в курсах грамматики, например, к особым случаям образования множественного числа существительных. Например, С.Ю. Стрелкова показала целесообразность использования BNC для анализа категории числа существительных греческого происхождения (*politics, statistics, mathematics, etc.*), обнаружив достаточно много примеров их употребления во множественном числе в значениях отличных от названия дисциплин, и согласующихся с глаголами во множественном числе – то, что как правило, не рассматривается в учебниках [29]. Большое количество существительных в области Биотехнология, образующих множественное число не по правилам, (*bacteria, data, criterion, analysis, и т.п.*) делает актуальным изучение этого грамматического аспекта с помощью корпуса. Типичные запросы могут представлять поиск по лемме форм указанных существительных, например, `DATUM` (поле `Word/phrase`), поиск всех слов, содержащих элемент `dat*` (поле `Word/phrase`) с последующим анализом полученных результатов и выделением слов, относящихся понятию «данные».

2.2.4. Примеры заданий на формирование продуктивных навыков (говорения и письма)

Задания на формирование продуктивных навыков включали обсуждение лексико-грамматических особенностей изученных терминов, краткий пересказ /формулирование основных идей текстов из корпуса NOW, содержащих изучаемую терминологию, и составление мини-гlossариев. Мини-гlossарии содержали следующую информацию: определение значения термина, перевод на русский язык, типичные коллокации с предлогами, существительными, прилагательными и глаголами. Описание термина

Preservative представлено в таблице 2. Первую часть работы студенты делали самостоятельно дома, а на занятии сравнивали полученные результаты по одинаковым терминам и дополняли свое описание, если у одногруппников были интересные находки.

Таблица 2. Описание термина Preservative

Определение	Перевод	Коллокации			
		Существительные	Прилагательные	Глаголы	Предлоги
something that preserves or tends to preserve, esp a chemical added to foods to inhibit decomposition	консервант	wood, food, products, market, chemicals, coatings	natural, chemical, artificial, cosmetic, mercury-based	BE, called, used, has, contain, added	in, for, to, of

К развитию коммуникативных навыков мы также относили задания на конспектирование и последующее комментирование презентаций одногруппников, сравнение вариантов перефразирования фрагмента текста с количественным анализом количества использованных слов и предложений и качественным анализом использованных приемов перефразирования (изменение порядка слов, частей речи, использование синонимов).

3. Анализ результатов экспериментального обучения с использованием подходов корпусной лингвистики в курсе АСЦ

Апробация описанных выше заданий проводилась в рамках курса Английский язык для специальных целей в течение осеннего семестра 2019-2020 учебного года в ИТМО. В экспериментальной группе было 10 студентов-бакалавров третьего курса по направлению подготовки «Биотехнология». В конце курса АСЦ, как в экспериментальной, так и контрольной группе, была проведено итоговое тестирование. Итоговый тест был одинаковым для обеих групп. Тест состоял из трех разделов: лексико-грамматического, проверявшего усвоение профессиональной лексики (4 задания – 50 баллов), аудирования, проверявшего восприятие профессиональной речи на слух (2 задания – 30 баллов), письменного, в котором испытуемые должны были письменно изложить основные биотехнологические аспекты работы с биоматериалом (1 задание – 20 баллов). Максимальное количество баллов за тест – 10 (фактический результат, разделенный на 10, по шкале BARS). Навык устного общения на профессиональные темы оценивался по представленной презентации по специальности.

Средний результат теста в экспериментальной группе составил 9,8 (SD=0,42) балла, в контрольной – 7,9 (SD=0,9). В контрольной группе максимальный балл за тест не получил никто, в экспериментальной это удалось сделать 80% учащихся. При уровне значимости 0,05 $T_{эмп}=6,07$, что больше критического значения $T_{0,05}=2,086$. Таким образом, можно сделать вывод, что различия характеристик контрольной и экспериментальной групп после окончания эксперимента статистически достоверны с вероятностью 95%.

Презентации, подготовленные студентами экспериментальной группы, характеризовались более уверенным владением профессиональной лексикой, меньшим количеством грамматических ошибок, использованием синтаксически более сложных предложений, имели более четкую структуру по сравнению презентациями, с которыми выступали студенты контрольной группы.

Не менее важным для нас были результаты анкетирования, проведенного в экспериментальной группе в конце обучения. Объем статьи не позволяет провести анализ всех ответов студентов. Его результаты показали, что студенты не считают дополнительную нагрузку, связанную с освоением и использованием корпусов,

как чрезмерную. Студенты высоко оценили возможности, которые дает прямое обращение к корпусу при выполнении учебных заданий. Студенты отметили, что работа с корпусом помогла им улучшить навыки устного и письменного профессионально-ориентированного общения на английском языке. Например, они стали увереннее использовать профессиональную лексику и лексические обороты при составлении сопроводительного письма для приема на работу, поняли, что корпус может помочь им при подготовке доклада или презентации по профессиональной тематике.

В течение семестра также проводился мониторинг посещаемости студентами занятий и выполнения домашних заданий. В экспериментальной группе оба этих показателя были выше, чем в контрольной, что позволяет говорить о более высокой мотивации студентов экспериментальной группы, которую мы целиком относим за счет использования подходов корпусной лингвистики в курсе АСЦ.

К основным достоинствам использования корпусов на занятии по АСЦ можно отнести также развитие аналитических способностей студентов, их навыков критического мышления, навыков самостоятельно находить ответы на вопросы, возникающие при изучении и использовании профессиональной лексики. Интересным в этом отношении является следующий пример: они догадались, что показателем исчисляемости существительного в корпусе являются примеры употребления заданного существительного с неопределенным артиклем.

Практика работы с корпусами в классе и при выполнении домашних заданий дает студентам новые возможности дальнейшего совершенствования знаний и навыков, связанных с профессиональным общением на английском языке, открывает перспективы самостоятельной подготовки к выступлениям на международных семинарах и конференциях, способствуя их профессиональному росту, развитию кругозора и успешной карьере в выбранной сфере деятельности.

К трудностям работы с корпусами можно отнести следующие:

- ограниченное количество выполняемых в течение дня запросов (не более 50 в день для студентов нелингвистического профиля);
- прерывание работы после нескольких запросов настойчивым предложением приобрести лицензию на использование корпуса (заметим, что работу можно возобновлять быстрее, чем рассчитывают создатели ресурса, если скопировать заданный запрос, нажать клавишу Reset, вставить запрос в очистившееся поле и повторить поиск, нажав на клавишу Find matching strings);
- невозможность проведения такой работы при возникновении проблем технического характера (отсутствие интернет связи, нехватка зарядки в устройстве);
- неудобство работы с устройствами с маленьким экраном.

К ограничениям данного эксперимента можно отнести небольшое количество участников. Поэтому полученные результаты носят предварительный характер и требуют подтверждения на более представительной выборке.

Выводы

В ходе работы была произведена оценка возможностей применения корпусов COCA и NOW как ресурсов для овладения специальной лексикой. Было показано, что корпус NOW, позволяет получить быстрый доступ к полным текстам первоисточников, устраняя необходимость составления узкоспециального корпуса, так как большинство терминов могут быть найдены в корпусе NOW в количестве достаточном для усвоения профессиональной лексики студентами по специальности «Биотехнология». В статье описаны упражнения, разработанные на основе подходов корпусной лингвистики для применения в курсе английского для специальных целей для студентов-биотехнологов, рассчитанные как на прямое обращение к корпусу, так и на работу с раздаточными материалами, подготовленными преподавателем.

Обращаясь для изучения специальных терминов к современным аутентичным текстам, студенты начинают понимать всю важность умения прочесть и понять текст на иностранном языке, сформулировать его основную идею и использовать информацию для выражения собственных мыслей на иностранном языке. Помимо информационной и профессионально-образовательной пользы такого современного аутентичного текста, студенты также осознают практическую пользу взаимодействия с корпусом, позволяющим находить ответы на возникающие вопросы. Корпусы позволяют студентам самостоятельно изучать, а впоследствии и применять в письменной и устной коммуникации современные обороты научного языка в их профессиональной области биотехнологий.

Успешное выполнение студентами заданий свидетельствует о целесообразности как прямого, так и непрямого использования корпуса NOW при обучении специальному английскому. Таким образом, использование корпуса NOW в курсе английского для специальных целей раскрывает потенциал как учащихся, получающих возможность совершенствовать свой профессиональный английский, изучая аутентичные тексты, так и преподавателя, приобретающего огромный ресурс и возможность для разработки интересных современных заданий для курса АСЦ. В дальнейшем необходимо продолжить исследования, направленные на изучение возможных способов оптимизации применения лингвистических корпусов в курсах английского для специальных целей, и апробировать их на большем количестве учащихся.

Литература

- [1] Текст стандарта 240700 «Биотехнология» ФГОС ВО. URL: <http://fgosvo.ru/fgosvpo/7/6/1/23> (дата обращения: 10.03.2020).
- [2] Oveshkova A.N. Work with English corpora as a means of promoting learner autonomy // Образование и наука. 2018. Т. 20. № 8. С. 66–87. DOI: 10.17853/1994-5639-2018-8-66-87.
- [3] Филошина И.О. Использование корпусного подхода в автономизации изучения и преподавания английского языка на фармацевтическом факультете // Фармацевтическое образование, наука и практика: горизонты развития. Материалы Всероссийской научно-практической конференции с международным участием, посвященной 50-летию фармацевтического факультета КГМУ (Под редакцией В.А. Лазаренко, И.Л. Дроздовой, И.В. Зубковой, О.О. Куриловой). Курск. 2016. С. 742–747.
- [4] Chang Ji-Y. The use of general and specialized corpora as reference sources for academic English writing: A case study // ReCALL. 2014. Vol. 26, № 2. P. 243–259. doi:10.1017/S0958344014000056.
- [5] Charles M. Getting the corpus habit: EAP students' long-term use of personal corpora // English for Specific Purposes. 2014. Vol. 35. P. 30–40. DOI:10.1016/j.esp.2013.11.004.
- [6] Charles M. Using Do-It-Yourself corpora in EAP: a tailor-made resource for teachers and students // The Journal of Teaching English for Specific and Academic Purposes. 2018. Vol. 6, № 2, Special Issue. P. 217–224. DOI: 10.22190/JTESAP1802217C.
- [7] Almazova N., Kogan M. Organizing polytechnic post-graduate students individual work on required reading corpora (within ESP course) // Университетский журнал. 2013. № 6. С. 13-25.
- [8] Rozanova Ya., Kudryashova A., Zamyatina O. Computer-aided research of ESP class materials: vocabulary potential and learning opportunities // Язык и культура. 2018. № 42. С. 152–162. DOI: 10.17223/19996195/42/9.
- [9] Cobb T. Necessary or nice? Computers in second language reading // In Z. Han & N. Anderson (Eds) L2 Reading Research and Instruction: Crossing the Boundaries. Ann Arbor: University of Michigan Press. 2009. P. 144 – 174.
- [10] Davies Mark. The Corpus of Contemporary American English (COCA): One billion words, 1990-present. URL: <https://www.english-corpora.org/coca/> (дата обращения: 10.03.2020).

- [11] Davies Mark. Corpus of News on the Web (NOW): 10 billion words from 20 countries, updated every day. URL: <https://www.english-corpora.org/now> (дата обращения: 10.03.2020).
- [12] Davies M. The 385+ million word Corpus of Contemporary American English (1990-2008+): Design, architecture, and linguistic insights // *International Journal of Corpus Linguistics*. 2009. Vol. 14, № 2. P. 159-188.
- [13] Davies M. The new 4.3 billion word NOW corpus, with 4–5 million words of data added every day // *Corpus linguistics. International conference*. 25-28 July 2017, University of Birmingham. Conference Abstracts (S. Hunston and M. Nasseri, eds.). 2017. P. 507 – 508 URL: <http://paulslals.org.uk/ccr/CL2017ExtendedAbstracts.pdf> (дата обращения: 29.02.2020).
- [14] Генкин Ю.Ю., Степко М.Л., Пересыпкин А.П. Информационно-технологическая составляющая формирования коммуникативных компетенций в процессе обучения иноязычной речевой деятельности с использованием инструментальных средств корпусной лингвистики // *Научные ведомости БелГУ. Серия: Гуманитарные науки*. 2017. № 14 (263). С. 51–57.
- [15] Тарнаева Л.П., Осипова Е.С. Использование ресурсов корпусной лингвистики при подготовке переводчиков в сфере профессиональной коммуникации // *Филологические науки. Вопросы теории и практики*. 2016. № 9-1 (63). С. 205–209.
- [16] Lessard-Clouston M., Chang T. Corpora and English Language Teaching: Pedagogy and Practical Applications for Data-Driven Learning. *TESL Reporter*. 2014. Vol. 47. P.1-20.
- [17] Boulton A. Integrating corpus tools and techniques in ESP courses // *ASp - La revue du GERAS, revue.org (en ligne) / Bordeaux: GERAS (imprimé)*. 2016. Vol. 69, № 69. P.111-135. DOI: 10.4000/asp.4826.hal-01250346.
- [18] Коган М.С., Куликова Е.В. Использование подходов корпусной лингвистики при обучении переводчиков в сфере профессиональной коммуникации // *Вопросы методики преподавания в вузе*. 2018. Т. 7, № 24. С. 65–78. DOI: 10.18720/HUM/ISSN 2227-8591.24.7. URL <https://hum.spbstu.ru/userfiles/files/pdf/VMP24.pdf> (дата обращения: 25.05.2020)
- [19] Коган М.С., Куликова Е.В. Корпус как ресурс для изучения лексико-грамматических аспектов английского языка для специальных целей // IX международная научно-практическая конференция «ПЕРЕВОД. ЯЗЫК. КУЛЬТУРА» (отв. редактор Л.В. Коцюбинская) 25 мая 2018г. СПб.: ЛГУ им. А.С. Пушкина, 2018. С.97 – 103. URL: https://www.elibrary.ru/download/elibrary_35415781_81651132.pdf (дата обращения: 25.05.2020)
- [20] Коган М.С., Куликова Е.В. Дидактические возможности лингвистических корпусов для обновления учебников по иностранному языку // *Магия ИННО: интегративные тенденции в лингвистике и лингводидактике: сборник научных трудов*. В 2 томах. Том 2 / под редакцией Е. Б. Морозовой; Моск. гос. ин-т междунар. отношений (ун-т) М-ва иностр. дел Рос. Федерации. Москва: МГИМО – Университет, 2019. С.215 – 223.
- [21] Домашняя страница сайта The essential chemical industry online. URL: <https://www.essentialchemicalindustry.org/materials-and-applications/biotechnology-in-the-chemical-industry.html> (дата обращения: 10.03.2020).
- [22] Инструмент VocabProfiler доступен по адресу URL: <https://www.lexutor.ca/vp/> (дата обращения: 10.03.2020).
- [23] Crossley S., Cobb T., McNamara D. Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*. 2013. Vol. 41, № 4. P. 965-981.
- [24] Zahar R., Cobb T., Spada N. Acquiring vocabulary through reading: Effects of frequency and contextual richness // *Canadian Modern Language Review*. 2001. Vol. 57, № 4. P. 541-572.

- [25] Glossary of terms commonly used in biotechnology. URL: <https://link.springer.com/content/pdf/bbm%3A978-1-4020-5149-4%2F1.pdf> (дата обращения: 10.03.2020).
- [26] Wegmann S. Experimental evidence for the age dependence of tau protein spread in the brain // *Science Advances*. 2019. Vol. 5, № 6. URL: <https://advances.sciencemag.org/content/5/6/eaaw6404.full> (дата обращения: 10.03.2020).
- [27] Boulton A. Hands-on/hands-off: Alternative approaches to data-driven learning // Thomas J. and Boulton A. (eds.) *Input, process and product: Developments in teaching and language corpora*. Brno: Masaryk University Press. 2012. P. 152–168.
- [28] Kogan M., Yaroshevich A., Ni O. Corpus-based teaching of German compound nouns and lexical bundles for improving academic writing skills // *Lidil*. 2018. № 58. [Electronic edition]. URL: <http://journals.openedition.org/lidil/5438> (дата обращения: 29.02.2020). DOI: 10.4000/lidil.5438.
- [29] Popova N., Kogan M. Alternative approaches to fulfilling call tasks as a way of stimulating communication among foreign language learners // *Proceedings of 2014 CALL conference "Research Challenges in CALL"* Antwerp, 7 -9 July 2014. Antwerp: University of Antwerp. 2014. P. 201–207.
- [30] Стрелкова С.Ю. Politics is или politics are: о некоторых случаях формального и понятийного согласования в британском варианте английского языка // *Иностранные языки в школе*. 2011. № 10. С. 76-80.

On Prospects of Using the NOW Corpus in ESP Course for Undergraduates Majoring in Biotechnology

I.V. Kuznetsova¹, M.S. Kogan²

¹ ITMO University, ² Peter the Great St. Petersburg Polytechnic University

The article considers the possibilities of using the COCA and NOW corpora as supplementary sources to English for Special Purposes (ESP) course for students majoring in biotechnology. It describes some of the characteristics of these two corpora and compares them, performing analysis as to which corpus is best suited for the purpose of teaching biotechnology terms to students. The compared and analyzed characteristics include the frequency of biotechnology terms in each corpora and the possibility of gaining access to full texts of the extracts presented in concordances with biotechnology terms. Based on this information, the authors make their conclusion about the possibility of using the corpora for teaching biotechnology terms and development of professional vocabulary exercises. Next, the authors describe corpus-based exercises developed for teaching lexico-grammatical properties of professional vocabulary, and also show some examples of tasks aimed at the improvement of students' productive and analytical skills. The article discusses difficulties related to the practical issues of integrating developed corpus-based exercises in the ESP classroom mentioned in students' responses to the ESP course review questionnaire. In conclusion, the authors reflect on study limitations and outline further research in the field of using corpora for teaching ESP.

Keywords: English for specific purposes, linguistic corpora, biotechnology, context, professional vocabulary acquisition

Reference for citation: Kuznetsova I.V., Kogan M.S. On prospects of using the NOW corpus in ESP course for undergraduates majoring in biotechnology // *Computer Linguistics and Computing Ontologies*. Vol. 4 (Proceedings of the XXIII International Joint Scientific Conference «Internet and Modern Society», IMS-2020, St. Petersburg, June 17-20, 2020). - St. Petersburg: ITMO University, 2020. P. 29 – 43. DOI: 10.17586/0000-0000-2020-4-29-43

References

- [1] Text of the standard 240700 "Biotechnology" FSES HE. URL: <http://fgosvo.ru/fgosvpo/7/6/1/23> (date accessed: 03/10/2020). [In Russian].
- [2] Oveshkova A.N. Work with English corpora as a means of promoting learner autonomy // *Education and Science*. 2018. Vol. 20, № 8. P. 66–87. DOI: 10.17853/1994-5639-2018-8-66-87.
- [3] Filoshina I.O. The use of the corpus approach in the autonomization of the study and teaching of the English language at the Faculty of Pharmacy // *Pharmaceutical Education, Science and Practice: Development Horizons. Materials of the All-Russian scientific and practical conference with international participation, dedicated to the 50th anniversary of the pharmaceutical faculty of KSMU* (Edited by V.A.Lazarenko, I.L.Drozдова, I.V. Zubkova, O.O. Kurilova). Kursk. 2016. P. 742–747. [In Russian].
- [4] Chang Ji-Y. The use of general and specialized corpora as reference sources for academic English writing: A case study // *ReCALL*. 2014. Vol. 26, № 2. P. 243–259. doi: 10.1017/S0958344014000056.
- [5] Charles M. Getting the corpus habit: EAP students' long-term use of personal corpora // *English for Specific Purposes*. 2014. Vol. 35. P. 30-40. DOI.org/10.1016/j.esp.2013.11.004.
- [6] Charles M. Using Do-It-Yourself corpora in EAP: a tailor-made resource for teachers and students // *The Journal of Teaching English for Specific and Academic Purposes*. 2018. Vol. 6, № 2, Special Issue. P. 217-224. DOI: 10.22190/JTESAP1802217C.
- [7] Almazova N., Kogan M. Organizing polytechnic post-graduate students individual work on required reading corpora (within ESP course) // *University Journal*. 2013. № 6. P. 13-25. [In Russian].
- [8] Rozanova Ya., Kudryashova A., Zamyatina O. Computer-aided research of ESP class materials: vocabulary potential and learning opportunities // *Language and culture*. 2018. № 42, P. 152–162. DOI: 10.17223 / 19996195/42/9.
- [9] Cobb T. Necessary or nice? Computers in second language reading // In Z. Han & N. Anderson (Eds) *L2 Reading Research and Instruction: Crossing the Boundaries*. Ann Arbor: University of Michigan Press. 2009. P. 144 - 174.
- [10] Davies Mark. The Corpus of Contemporary American English (COCA): One billion words, 1990-present. URL: <https://www.english-corpora.org/coca/> (date accessed: 03/10/2020).
- [11] Davies Mark. Corpus of News on the Web (NOW): 10 billion words from 20 countries, updated every day. URL: <https://www.english-corpora.org/now> (date accessed: 03/10/2020).
- [12] Davies M. The 385+ million word Corpus of Contemporary American English (1990-2008 +): Design, architecture, and linguistic insights // *International Journal of Corpus Linguistics*. 2009. Vol. 14, № 2. P. 159-188.
- [13] Davies M. The new 4.3 billion word NOW corpus, with 4–5 million words of data added every day // *Corpus linguistics. International conference. 25-28 July 2017, University of Birmingham. Conference Abstracts* (S. Hunston and M. Nasseri, eds.). 2017. P. 507 - 508 URL: <http://paulslals.org.uk/ccr/CL2017ExtendedAbstracts.pdf> (date accessed: 29.02.2020).
- [14] Genkin Yu.Yu., Stepko M.L., Peresypkin A.P. Information and technological component of the formation of communicative competencies in the process of teaching foreign language speech activity using the tools of corpus linguistics // *Scientific Bulletin of BelSU. Series: Humanities*. 2017. № 14 (263). P. 51-57.
- [15] Tarnaeva L.P., Osipova E.S. Using the resources of corpus linguistics in the preparation of translators in the field of professional communication // *Philological Sciences. Questions of theory and practice*. 2016. № 9-1 (63). P. 205–209. [In Russian].
- [16] Lessard-Clouston M., Chang T. Corpora and English Language Teaching: Pedagogy and Practical Applications for Data-Driven Learning. *TESL Reporter*. 2014. Vol. 47. P.1-20.

- [17] Boulton A. Integrating corpus tools and techniques in ESP courses // *ASp - La revue du GERAS*, revue.org (en ligne) / Bordeaux: GERAS (imprimé). 2016. Vol. 69, № 69. P. 111-135. 10.4000 / asp.4826.hal-01250346.
- [18] Kogan M.S., Kulikova E.V. Using the approaches of corpus linguistics in teaching translators in the field of professional communication // *Questions of teaching methods at the university*. 2018. Vol. 7, № 24, P. 65–78. DOI: 10.18720 / HUM / ISSN 2227-8591.24.7. URL <https://hum.spbstu.ru/userfiles/files/pdf/VMP24.pdf> (date accessed: 05/25/2020)
- [19] Kogan M.S., Kulikova E.V. Corpus as a resource for studying lexical and grammatical aspects of the English language for special purposes // IX International Scientific and Practical Conference “TRANSLATION. LANGUAGE. CULTURE” (editor-in-chief L. V. Kotsyubinskaya) May 25, 2018 SPb.: Leningrad State University im. A.S. Pushkin, 2018. P. 97 - 103. URL: https://www.elibrary.ru/download/elibrary_35415781_81651132.pdf (date accessed: 05/25/2020). [In Russian].
- [20] Kogan M.S., Kulikova E.V. Didactic possibilities of linguistic corpora for updating textbooks on a foreign language // *Magic INNO: integrative tendencies in linguistics and linguodidactics: collection of scientific papers*. In 2 volumes. Volume 2 / edited by EB Morozova; Moscow state in-t international. relations (un-t) M-va foreign. cases Ros. Federation. Moscow: MGIMO - University, 2019. P. 215 - 223. [In Russian].
- [21] The home page of The essential chemical industry online. URL: <https://www.essentialchemicalindustry.org/materials-and-applications/biotechnology-in-the-chemical-industry.html> (date accessed: 03/10/2020).
- [22] The VocabProfiler tool is available at the URL: <https://www.lex tutor.ca/vp/> (accessed 03/10/2020).
- [23] Crossley S., Cobb T., McNamara D. Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System*. 2013. Vol. 41, № 4. P. 965-981.
- [24] Zahar R., Cobb T., Spada N. Acquiring vocabulary through reading: Effects of frequency and contextual richness // *Canadian Modern Language Review*. 2001. Vol. 57, № 4. P. 541-572.
- [25] Glossary of terms commonly used in biotechnology. URL: <https://link.springer.com/content/pdf/bbm%3A978-1-4020-5149-4%2F1.pdf> (date accessed: 03/10/2020).
- [26] Wegmann S. Experimental evidence for the age dependence of tau protein spread in the brain // *Science Advances*. 2019. Vol. 5, № 6. URL: <https://advances.sciencemag.org/content/5/6/eaaw6404.full> (date accessed: 03/10/2020)
- [27] Boulton A. Hands-on / hands-off: Alternative approaches to data-driven learning // Thomas J. and Boulton A. (eds.) *Input, process and product: Developments in teaching and language corpora*. Brno: Masaryk University Press. 2012. P. 152–168.
- [28] Kogan M., Yaroshevich A., Ni O. Corpus-based teaching of German compound nouns and lexical bundles for improving academic writing skills // *Lidil*. 2018. No. 58. [Electronic edition]. URL: <http://journals.openedition.org/lidil/5438> (date accessed: 02/29/2020). DOI: 10.4000 / lidil.5438.
- [29] Popova N., Kogan M. Alternative approaches to fulfilling call tasks as a way of stimulating communication among foreign language learners // *Proceedings of 2014 CALL conference “Research Challenges in CALL”* Antwerp, 7 -9 July 2014. Antwerp: University of Antwerp. 2014. P. 201–207.
- [30] Strelkova S.Yu. Politics is or politics are: on some cases of formal and conceptual agreement in the British version of the English language // *Foreign languages at school*. 2011. № 10. P. 76-80. [In Russian].