

РАЗРАБОТКА ПРОФЕССИОНАЛЬНО ОРИЕНТИРОВАННОГО АНГЛОЯЗЫЧНОГО УЧЕБНОГО КОРПУСА ТЕКСТОВ ДЛЯ ПОДГОТОВКИ АВИАЦИОННЫХ СПЕЦИАЛИСТОВ

Киргинцева Н.С.¹, Жехрова М.В.²

¹Федеральное государственное казенное военное образовательное учреждение высшего профессионального образования «Военный учебно-научный центр Военно-воздушных сил «Военно-воздушная академия имени профессора Н.Е. Жуковского и Ю.А. Гагарина», Воронеж, e-mail: natalysn@yandex.ru;

²Санкт-Петербургский государственный университет гражданской авиации, Санкт-Петербург, e-mail: marizh7@bk.ru

В статье продемонстрирована перспективность разработки профессионально ориентированных англоязычных учебных корпусов текстов, предназначенных для профессиональной подготовки авиационных специалистов. Перечислен ряд преимуществ «корпусного подхода» к обучению иностранным языкам будущих специалистов. Показано, что применение учебных корпусов развивает не только рецептивные и продуктивные, но и исследовательские умения обучающихся, а также их коллокационную компетенцию. Описана технология компьютерной поддержки разработки лингвистических корпусов с использованием программ комплекса AntLab. Она включает последовательное прохождение семи шагов, часть из которых может быть пропущена при составлении учебных корпусов. Выявлено, что в качестве менеджера корпуса можно использовать любую программу, работающую с текстовыми файлами: например, программы-аутлайнеры Zim, Outwiker и программы, реализующие wiki-технологии (например, TiddlyWiki). Смоделирован процесс работы преподавателей по созданию учебных корпусов с использованием аутлайнера Zim и программы AntConc, входящей в состав комплекса AntLab. Сделан вывод о том, что использование инструментов программы AntConc позволяет получать информацию, которая может быть использована для построения терминологических словарей, разработки упражнений, а также развития коллокационной компетенции будущих авиационных специалистов.

Ключевые слова: учебный корпус, корпусный подход, коллокационная компетенция, авиационный специалист, аутлайнер.

DEVELOPING A PROFESSIONALLY-ORIENTED ENGLISH LANGUAGE LEARNER CORPUS FOR PREPARING AVIATION SPECIALISTS

Kirgintseva N.S.¹, Zhekhrova M.V.

¹ Military Educational and Scientific Center of the Air Force «N.E. Zhukovsky and Y.A. Gagarin Air Force Academy», Voronezh, e-mail: natalysn@yandex.ru;

² St. Petersburg State University of Civil Aviation, Saint Petersburg, e-mail: marizh7@bk.ru

The article demonstrates the viability of developing professionally-oriented learner text corpus intended for preparing aviation specialists. A variety of advantages of the “corpus approach” to teaching foreign languages to prospective specialists is enumerated. It is demonstrated that applying learner corpus develops not only learners’ receptive and productive skills, but also their researching faculties, as well as their collocations competence. The computer sustainment technology of developing linguistic corpus with the help of AntLab software programs is described. It includes continuous undertaking of seven steps, some of which can be skipped when composing learner corpus. It is illicited that any software application working with text files can be used as the corpus manager: for example, the Zim and Outwiker outliner software, and wiki-based software such as TiddlyWiki. The process of teacher’s work for creating learners corpus with the help of Zim outliner and AntConc, a component of AntLab software has been modelled. A conclusion was made that using the AntConc tools allows to get information, which can be applied for composing terminology dictionaries, designing exercises, and developing prospective aviation specialists collocations competence.

Keywords: learner corpus, corpus approach, collocations competence, aviation specialist, outliner.

Для специалистов, проходящих профессиональную подготовку в области авиации, способность осуществлять коммуникацию на английском языке является почти настолько же важной, как и умение работать с авиационной техникой. Объясняется это, прежде всего, тем,

что именно английский язык является принятым языком международного общения и авиации и именно он используется для радиообмена. На необходимости развития способности применять современные коммуникативные технологии, в том числе на иностранном языке, для академического и профессионального взаимодействия фокусируют внимание Федеральные государственные образовательные стандарты высшего образования, по которым ведется подготовка авиационных специалистов в вузах нашей страны. В них речь также идет о том, что в результате освоения программ профессиональной подготовки специалисты должны быть способны представлять информацию в требуемом формате с использованием информационных, компьютерных и сетевых технологий. Составленные на основе Федеральных стандартов рабочие программы дисциплины «Иностранный язык» часто актуализируют необходимость осуществлять поиск, анализ и систематизацию информационных ресурсов на иностранном языке с помощью информационно-коммуникационных технологий, что еще раз подтверждает актуальность выбранной темы исследования.

Одним из инструментов, позволяющих проводить анализ и систематизацию профессионально-релевантной информации на иностранных языках, является электронный языковой (лингвистический) корпус. Навыки работы с корпусными данными крайне необходимы как преподавателю иностранного языка, так и авиационному специалисту наряду с навыками использования электронных словарей, машинных переводчиков, интернет-ресурсов и других технических средств [1].

Актуальным является также создание учебных корпусов текстов, релевантных конкретной предметной области, что предполагает выбор речевых образцов профессионального общения, используемых носителями языка с высокой долей вероятности. В свою очередь, это «оптимизирует формирование и совершенствование лексических навыков и умений и создает условия для внедрения рациональной методики обучения иностранному языку» [2, с. 216].

Преимущество использования учебных корпусов состоит в том, что они позволяют преподавателю иностранного языка актуализировать собственную лингвистическую компетенцию (всем известно, что любой язык развивается, со временем меняются его лексика и грамматика, а узус начинает отличаться от принятой «нормы»). Также плюсом данного подхода является возможность применения учебных корпусов для разработки заданий на основе аутентичных материалов (indirect use), а также непосредственно на занятиях по иностранному языку для анализа обучающимися языкового материала с точки зрения используемых в «живой» речи коллокаций (direct use).

Еще одно преимущество «корпусного подхода» к обучению иностранным языкам состоит в том, что самостоятельная работа обучающихся с языковым корпусом способствует развитию не только рецептивных, но и продуктивных умений устной и письменной коммуникации в соответствии с условиями и требованиями конкретной ситуации профессионального общения [1]. Всё это, а также тот факт, что работа с учебным корпусом активизирует и исследовательскую составляющую учебного процесса и повышает мотивацию обучающихся к изучению иностранного языка [1], говорит о несомненной актуальности разработки и использования учебных корпусов в процессе обучения иностранному языку будущих авиационных специалистов. Цель данного исследования – изучить возможности компьютерной поддержки разработки профессионально ориентированного англоязычного учебного корпуса текстов для подготовки авиационных специалистов.

Результаты исследования и их обсуждение

В научной литературе представлено большое количество определений языкового корпуса. На основе их анализа автор работы [1] выводит следующее: «языковой корпус - это репрезентативная выборка текстов или их фрагментов, представляющая собой определенным образом унифицированный, структурированный и размеченный текстовый массив конечного объема, доступный в электронном формате и предназначенный для решения конкретных исследовательских задач». Следует подчеркнуть, что в свете сказанного выше корпус может быть использован не только в исследовательских целях, но и в учебных.

Одной из главных характеристик корпуса является его репрезентативность (сбалансированность), под которой понимается пропорциональное представление в корпусе текстов требуемой тематики и жанров, что обеспечивает необходимую полноту и типичность представленных в нем языковых данных. Хотя, говоря об учебных корпусах, предназначенных для обучения авиационных специалистов английскому языку, следует подчеркнуть их временную гомогенность (они относятся к современному периоду) и жанровую ограниченность (большинство текстов носят научно-популярный характер).

В работе О.Ю. Павловой [1] отмечается большая роль языковых корпусов при составлении словарей, грамматик, справочных и учебных материалов. При этом отмечается проблема недостатка специализированных тематических корпусов. Кроме того, в данной работе даются рекомендации по использованию корпусов текстов, которые могут помочь при создании и использовании учебных корпусов (например, поиск коллокаций - проверка лексической сочетаемости). Отмечается, что владение коллокациями считается важным компонентом языковой компетенции, и что языковой корпус представляется незаменимым средством формирования коллокационной компетенции.

Составление тематических списков ключевых (наиболее частотных) слов, которые обучающиеся могут получить с помощью программных средств, а также поиск иллюстративного материала - другие варианты применения учебных корпусов. Их достоинством в этом смысле является то, что «... языковые данные представлены там в своем естественном контексте, что дает возможность изучения как непосредственного лексического окружения языковых единиц, так и их функционирования в более широком контексте на уровне фрагмента текста» [1].

Важным моментом является компьютерная поддержка (поддержка корпуса текстов комплексом программ по обработке данных, обеспечивающих выявление контекстов слова, статистическую инвентаризацию, автоматическую словарную обработку и т.д.).

Теоретически эту проблему можно решить с использованием программ комплекса AntLab. В статье [3] дано описание разработки объединенной группы исследователей из Японии и Новой Зеландии по созданию бесплатных и надежных сервисов, под названием AntLab. Данные программы активно используются в корпусных исследованиях как за рубежом, так и у нас в стране. На наш взгляд, часть из этих сервисов полезна для организации квантитативного отбора содержания обучения для CALL-программы. Работа с программой AntConc описана в работах [4; 5].

На первом этапе проводится анализ известных в мире библиотек и наукометрических баз данных с помощью программы AntCorGen с целью отбора исследовательского материала. Затем программами AntFileConverter, EndCodeAnt и VariAnt проводится перевод текстов в нужный формат, проверка правописания и удаление незначимых для анализа символов. С использованием программы AntConc создается частотный словарь, выделяются ключевые слова и выявляются устойчивые словосочетания с ключевыми словами. Затем с применением программы ProtAnt наращивается объем исследуемых текстов по ключевым словам, выявленным на предварительном этапе или по эталонному тексту. Данная программа осуществляет отбор текстов, близких по лексическому составу к поставленной учебной задаче. После этого следует вернуться к работе по лингвостатистическому исследованию расширенного состава учебных текстов в программе AntConc, а также к профилированию полученного словаря и разметке текстов. С помощью программы AntMover производится фрагментация и структурно-семантическая разметка текстов по предложениям с сохранением каждого предложения в отдельный файл. Далее программой ProtAnt обрабатываются файлы отдельных предложений и выявляются те из них, которые по лексическому или грамматическому составу подходят для кодирования упражнений в CALL-программу.

На наш взгляд, данная технология предназначена для создания лингвистических корпусов текста и поэтому содержит ряд операций, избыточных при создании учебного

корпуса текстов. Кроме того, программа AntCorGen жестко привязана (подключается) только к базе PLOS ONE, при этом количество текстовых ресурсов в этой базе ограничено (на момент написания статьи 47 текстов). Однако несомненным достоинством данного комплекса программ является их бесплатность при некоммерческом использовании и возможность работы отдельных программ вне комплекса, то есть их автономность.

В качестве менеджера корпуса можно использовать любую программу, работающую с текстовыми файлами. Она позволяет создавать базу данных (корпус текстов) и обеспечивает совместимость с лингвистическим программным обеспечением. Для этой роли хорошо подходят программы-аутлайнеры Zim, Outwiker и программы, реализующие wiki-технологии (например, TiddlyWiki).

В данной работе мы рассмотрим использование программы Zim. Она является свободным и кросс-платформенным продуктом, обладающим широким функционалом. В обычном режиме она работает с текстовыми ресурсами, создавая базу документов в виде страниц в формате *.txt. Программа позволяет форматировать информацию, создавать гиперссылки на внутренние и внешние файлы, вставлять графические изображения. Следует подчеркнуть, что возможность подключения внешних программ вообще является важным свойством программы. Данная особенность позволяет подключать программы из коллекции AntLab, в частности программу AntConc. Кроме того, нами был подключен текстовый редактор для редактирования исходного кода текстовых ресурсов. Схема корпуса показана на рисунке 1.

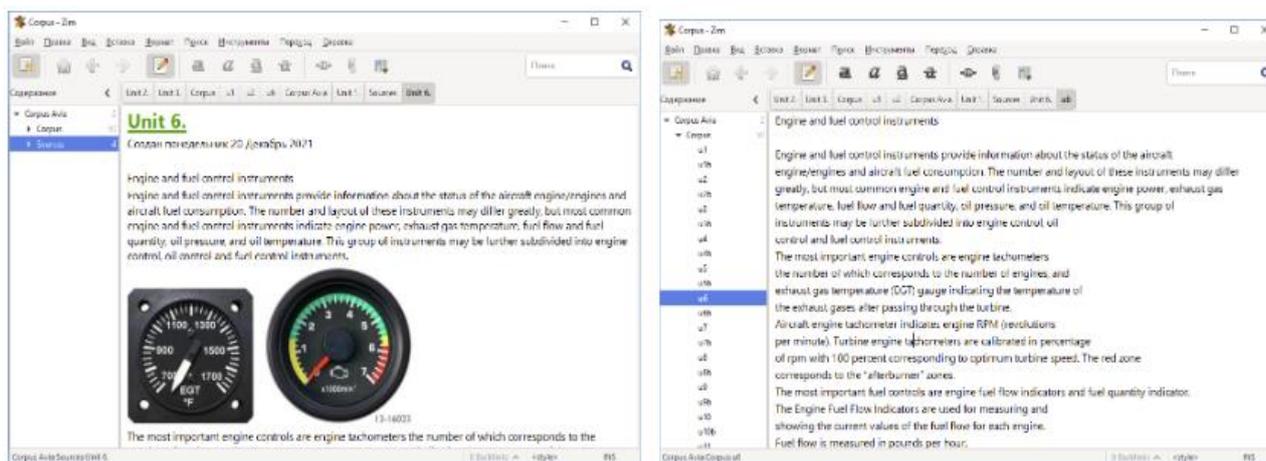


Рис. 1. Схема корпуса

Процесс создания моделирует процесс работы преподавателей по созданию учебных корпусов. Корпуса текстов создаются на основе книг, монографий, статей и т.д. В качестве основы (ядра) корпуса в нашем исследовании использовался материал авторского учебника «Английский язык. Авиационное и радиоэлектронное оборудование» [6].

Он включает материал, размещенный в 25 разделах (units). Тематически материал можно разделить на следующие блоки: авиационное оборудование; радиоэлектронное оборудование; беспилотные летательные аппараты.

В процессе создания учебного корпуса материал учебника в формате документа Word (docx) был переведен в базу программы Zim посредством простого копирования материала. При этом была проведена разметка по разделам, тестовым материалам и т.д. Следует заметить, что поскольку учебный корпус не требует строгого соблюдения требований к разметке текста, с учетом возможности использования в нем программ, разработанных для корпусной лингвистики, необходимо было провести работу по предварительной подготовке материала. Все разделы были приведены к формату текстовых файлов с расширением *.txt в формате кодировки utf-8, причем были удалены переносы и цифры. В учебном корпусе системе были созданы два раздела: источники (Sources) и корпус (Corpus) (рис. 2).



Источники

Фрагмент корпуса

Рис. 2. Источники и фрагмент текста корпуса

Подготовленный таким образом материал использовался для дальнейшей обработки с помощью программы AntConc. После загрузки корпуса использовался инструмент «Word List». Это позволило получить упорядоченный по частоте использования список слов. Результат работы программы AntConc в режиме Word List отражен на рисунке 3.

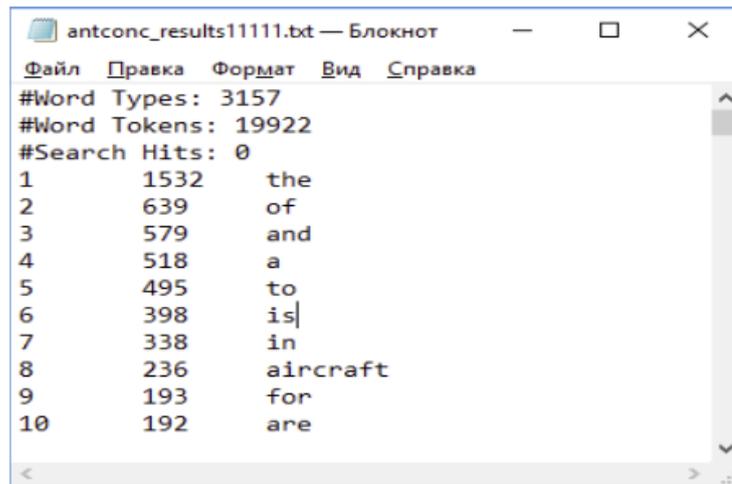


Рис. 3. Результат работы программы AntConc в режиме Word List

В данном корпусе было выявлено 19992 (19696) токена и 3157 (3270) словоформ. Среди словоформ можно выделить служебные части речи, которые не несут серьёзной смысловой нагрузки, и значимые части речи, которые имеют большое значение для передачи учебной информации. Для того чтобы убрать незначимые (служебные) элементы, в программе можно использовать так называемый stoplist. После подключения данного инструмента программа выдала новый список (рис. 4).

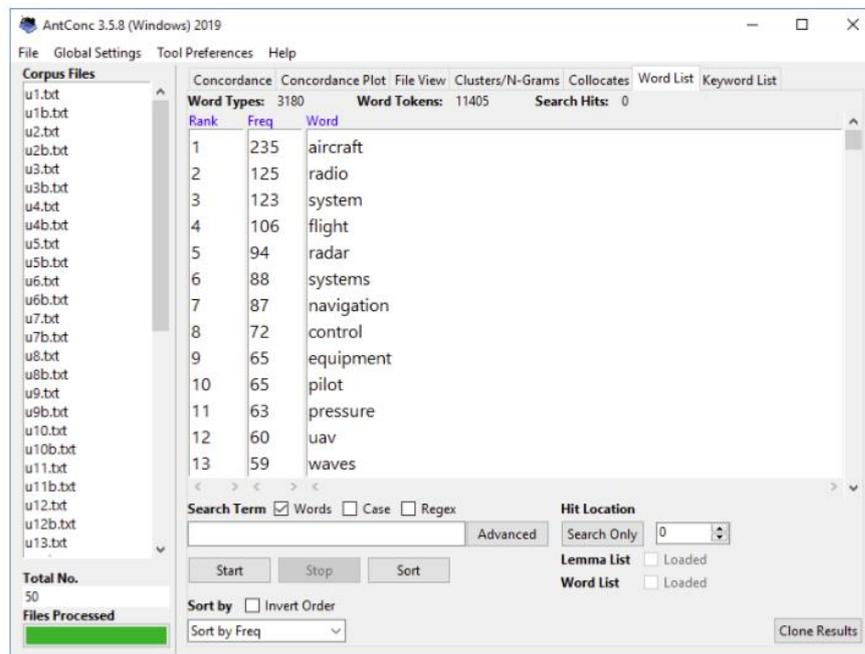


Рис. 4. Список словоформ значимой части речи

Так как пособие было рассчитано для использования в учебном процессе на различных факультетах, представляет интерес оценить распределение лексики по тематическим разделам специальностей «Техническая эксплуатация и восстановление электросистем и пилотажно-

навигационных комплексов боевых летательных аппаратов» и «Техническая эксплуатация транспортного радиоэлектронного оборудования». Для этого в полученном списке было выбрано по 10 терминов (словоформ), наиболее часто встречаемых в корпусе.

Полученные два списка были исследованы с помощью программы AntConc на определение их статистических и лингвистических характеристик. В таблице приведены по 10 наиболее частотных терминов из разделов «Радиооборудование» и «Авиационное оборудование».

Списки терминов радиооборудования и авиационного оборудования

#Word Types: 10			#Word Types: 10		
#Word Tokens: 315			#Word Tokens: 523		
#Search Hits: 0			#Search Hits: 0		
1	63	pressure	1	129	radio
2	58	fuel	2	94	radar
3	50	electrical	3	88	navigation
4	32	power	4	59	waves
5	29	engine	5	43	antenna
6	20	energy	6	30	communication
7	18	inertial	7	21	transmitter
8	17	alternator	8	21	frequency
9	14	battery	9	21	receiver
10	14	gyro	10	17	beam

С помощью инструмента «Concordance» были получены результаты поиска в формате «ключевое слово» в контексте, представляющем информацию об использовании словоформы в корпусе текстов.

Использование инструмента Concordance plot позволило получить информацию о размещении исследуемых слов в текстах корпуса в виде штрих-кода, что отражено на рисунке 5.

gyro



alternator



Рис. 5. Результаты работы инструмента Concordance plot с терминами *gyro* и *alternator*

Визуальная оценка показывает, что термин «gyro» используется в 4 модулях, а термин «alternator» только в одном, причем есть возможность определить, в каких конкретных разделах отдельного текста размещаются данные термины.

Далее мы можем, используя инструмент Words Clusters, осуществить отбор слов с заданным количеством элементов слева и справа от заданного слова. То есть оценить, с какими словами чаще всего используются данные термины в учебном корпусе.

Ниже приведены результаты работы инструмента Words Clusters с термином **gyro**.

#Total No. of Cluster Types: 7

#Total No. of Cluster Tokens: 14

1	4	2	the gyro
2	3	2	a gyro
3	3	3	directional gyro
4	1	1	and gyro
5	1	1	mechanical gyro
6	1	1	rotating gyro
7	1	1	stationary gyro

С термином **gyro** слева от него чаще всего размещаются следующие слова: *directional*, *mechanical*, *rotating*, *stationary*.

Инструмент Collocates дает возможность сделать статистический анализ по словам, стоящим слева или справа от исследуемого элемента «alternator». Ниже приведен результат поиска коллокаций словоформы *alternator* при сортировке результатов по силе коллокации.

#Total No. of Collocate Types: 90

#Total No. of Collocate Tokens: 170

1	3	0	3	1.98935	battery
2	2	1	1	1.12428	alternator
3	3	0	3	0.79671	power
4	4	2	2	0.56789	electrical
5	2	1	1	0.35376	engine
6	1	1	0	0	zero
7	1	0	1	0	work
8	1	1	0	0	will

Самая сильная степень коллокации у термина **alternator** наблюдается с термином **battery**. При щелчке по термину **battery** открывается окно **Concordance**, где можно увидеть примеры совместного использования словоформ в конкретном тексте.

1 basic components. These consist of an **alternator**, a **battery**, switches, circuit
u3b.txt

2 ammeter shows electrical performance of an **alternator** in relation to the **battery**, a
u3b.txt

3 will show a charge. If the **alternator** is not charging the **battery** or
u3b.txt

Таким образом, как видим, результаты работы программы **AntConc** могут использоваться как в исследовательских целях, так и в процессе обучения. Обучающиеся могут самостоятельно анализировать узус, тем самым совершенствуя свою коллокационную компетенцию.

Выводы

Результаты анализа статистических показателей данного корпуса показали, что, несмотря на то что учебник, на базе которого он был создан, ориентирован на несколько специальностей, в лексическом составе содержится материал, в котором преобладает тематика профиля подготовки специалистов по эксплуатации радиооборудования. Показано, что использование инструментов программы **AntConc** позволяет получать информацию, которая может быть использована для построения терминологических словарей, разработки упражнений, а также развития коллокационной компетенции будущих авиационных специалистов. Планируется, что данный корпус будет развиваться как преподавателями, так и обучающимися в рамках коллаборативных технологий.

Список литературы

1. Павлова О.Ю. Использование языковых корпусов в обучении иностранному языку // Язык и культура. 2021. № 54. С. 283-298.
2. Горина О.Г. Использование технологий корпусной лингвистики для развития лексических навыков студентов-регионоведов в профессионально-ориентированном общении на английском языке: дис. ... канд. пед. наук. Санкт-Петербург, 2014. 308 с.
3. Nymm V., Piotrovskaya X., Nomm S. Using Stochastic Learning Theory and Corpus Tools in CALL. R. Piotrowski Readings in Language Engineering and Applied Linguistics: Proceedings, Saint-Petersburg, November 27, 2017. St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, Herzen State Pedagogical University of Russia. Saint-Petersburg: Creative Commons ССО, 2018. P. 166-176.
4. Котюрова И.А. Корпусные исследования с помощью сервиса antconc в условиях работы в вузе // Язык и культура. 2020. № 52. С. 36-50.
5. Гришенкова Е.Г. Использование специального корпуса текстов в профессионально-ориентированном обучении иностранным языкам // Современные лингвокоммуникативные практики. Саратов, 2021.С. 17-25.
6. Киргинцева Н.С., Поталуй В.В., Ширшикова Е.А., Неровная Н.А., Варушкина А.В., Романчук В.О. Английский язык. Авиационное и радиоэлектронное оборудование. Воронеж: ВУНЦ ВВС «ВВА», 2021. 362 с.