

Исследование английской фразеологии с помощью подходов корпусной лингвистики

И.А. Комарова, М.С. Коган

Санкт-Петербургский политехнический университет Петра Великого

irinal5april@yandex.ru, kogan_ms@spbstu.ru

Аннотация

Статья посвящена перспективам использования подходов корпусной лингвистики в формировании лексико-фразеологических навыков обучающихся. Т.к. в рамках учебного процесса разработка дидактических материалов с использованием корпусных данных осуществляется с учетом рекомендованных для изучения дисциплин учебных пособий, была предпринята попытка верифицировать фразеологические единицы, включенные в пособия по практике речевого общения и перевода, в корпусе современного американского английского языка (СОСА). Приведено обоснование выбора корпуса СОСА для проведения подобного исследования.

При верификации 535 отобранных фразеологизмов в корпусе были обнаружены 262 единицы в своем нормативном виде и 71 в трансформированном. В работе представлен список самых частотных фразеологизмов, список фразеологизмов, не обнаруженных в корпусе, но рекомендуемых к изучению в пособии по практике перевода, а также и наиболее типичные примеры трансформаций фразеологических единиц, предлагаемых к изучению в пособии по практике речевого общения. Сравнение полученных результатов поиска трансформированных фразеологических единиц с имеющимися в литературе данными позволяет сделать вывод о целесообразности продолжения изучения вариантности фразеологизмов обязательных для освоения студентами лингвистических направлений.

Ключевые слова: корпусная лингвистика, английская фразеология, корпус СОСА, поисковый запрос, верификация фразеологизмов, частотность, нормативный вид, типичные трансформации, дидактические материалы

Библиографическая ссылка: Комарова И.А., Коган М.С. Исследование английской фразеологии с помощью подходов корпусной лингвистики // Компьютерная лингвистика и вычислительные онтологии. Выпуск 3 (Труды XXII Международной объединенной научной конференции «Интернет и современное общество», IMS-2019, Санкт-Петербург, 19 – 22 июня 2019 г. Сборник научных трудов). — СПб: Университет ИТМО, 2019. С. 40–49. DOI: 10.17586/2541-9781-2019-3-40-49

1. Введение

Одной из центральных проблем преподавания иностранного языка является овладение его словарным запасом как наиболее важным и существенным аспектом из тех, что должны быть усвоены студентами. Однако полноценное усвоение лексической системы языка невозможно без овладения его фразеологическим фондом, который является отражением целостной картины мира и лучшим доказательством реальности культуры конкретного языкового коллектива.

В широком смысле фразеология перестает быть периферийной областью языкознания, где хранятся только исключения и речевые ошибки, а осознается как фундаментальное

свойство языка. В 1991 году ведущий специалист корпусной лингвистики Джон Синклер сформулировал принцип идиоматичности. Он заключается в том, что «говорящий имеет в своем распоряжении большое число полуоформленных фраз, которые представляют собой уже готовые единицы, даже несмотря на то, что при анализе и можно разбить на сегменты» [1].

На сегодняшний день существует множество классификаций фразеологизмов, разработанных на основании различных принципов. Основопологающей для данного исследования стала структурно-семантическая классификация А.В. Кунина, который выделил четыре типа английских фразеологических единиц: 1) номинативные (именные, адъективные, адвербиальные и предложные); 2) номинативно-коммуникативные (глагольные); 3) междометные и модальные (фразеологизмы, выражающие эмоции, волеизъявления); 4) коммуникативные (со структурой простого и сложного предложения) [2]. Таким образом, вслед за А.В. Куниным, в данной работе мы берем за основу широкое понимание фразеологизмов, предполагающее включение в свой состав пословиц и поговорок как коммуникативных фразеологических единиц.

Некоторые зарубежные исследователи также включают в эту категорию самые разные выражения, пословный анализ которых не позволяет понять их смысл. Кроме глагольных выражений с дополнениями (*pass the buck*) и идиоматичных фразовых глаголов, к фразеологизмам относят предложные выражения с существительными (*after a fashion, off the wall*), дву- и трехчленные выражения *high and mighty, mix and match, lock, stock and barrel*, застывшие сравнения *as mad as a hatter, as black as your hat*, культурные аллюзии, цитаты, крылатые выражения, слоганы и т.д. [3, p. 81-82].

2. Корпусная лингвистика и проблема формирования лексико-фразеологических навыков

Особое значение формирование лексико-фразеологических навыков приобретает на основном и продвинутом этапах обучения, поскольку дальнейшее овладение языком как средством общения осуществляется в основном за счет расширения и обогащения словаря. Практически все учебные пособия по изучению английского языка содержат разделы, посвященные изучению, иногда тематическому, идиоматических выражений. Эти низкочастотные лексические единицы являются очень выразительными, яркими и потому более привлекательными для обучающихся продвинутого уровня чем высокочастотные нейтральные выражения повседневного общения.

В систему знаний и навыков по овладению фразеологической лексикой обычно включают следующие компоненты:

- понимание семантики единиц в определенном контексте или ситуации общения;
- комбинирование единиц с другими словами на основе правил лексической и грамматической сочетаемости;
- распознавание в речи границ каждой отдельно взятой фразеологической единицы;
- использование знакомых единиц в процессе осуществления речевой деятельности.

Последнее не всегда достижимо, потому что, даже зная идиоматическое выражение, иностранец не чувствует «себя в праве» его использовать, особенно в общении с носителями языка [3, p.97].

В целом усвоение фразеологии осуществляется на языковом и речевом уровнях, а завершается на культурологическом, на котором обучающиеся знакомятся с национально-культурной картиной мира. Полученные знания объединяются и способствуют развитию лингвокультурологической компетенции.

Непосредственный доступ к аутентичному материалу в области иноязычной фразеологии обучающийся может получить с помощью возможностей корпусной лингвистики. Корпусная лингвистика представляет собой целую методологию, которую можно применить ко многим языковедческим проблемам. В этом смысле, по мнению

В.П. Захарова и С.Ю. Богдановой, корпусную лингвистику можно назвать «пучком методов из разных областей лингвистических исследований» [4, с.9]. Основным понятием корпусной лингвистики является электронный языковой корпус. Выделяют два направления использования электронных корпусов в языковедческих исследованиях: базирующийся на корпусе (corpus-based) и направляемый корпусом (corpus-driven) [1, 5]. Вместе с тем следует отметить, что выделение этих двух подходов несколько условно. Представляя весь спектр современных корпусных исследований, Т. МакЭнери и А. Харди пишут о том, что различия между двумя основными подходами к изучению языковых явлений (corpus-driven и corpus-based) стираются [6].

При изучении фразеологии с помощью данных подходов фокус у лингвиста-исследователя и у преподавателя/обучающегося будет разным.

Для лингвистов-исследователей и лексикографов поиск идиом в корпусе представляет проблему, потому что «компьютер не знает, что такое идиома». Поэтому автоматические методы поиска идиоматических выражений в электронном корпусе с использованием существующих программных средств возможны лишь частично, подобные исследования предполагают большой объем ручной работы. Не рассматривая в этом контексте верификацию идиоматических выражений и фразеологизмов, зафиксированных в словарях, А. О'Кеефф с соавторами описывают методику поиска идиоматических выражений в корпусе на основе слов, «склонных» к образованию идиом, т.к. они составляют основу когнитивных метафор, связанных с частями тела, денег, цвета, света и т.п. Исследователи обнаружили, что словоформа *face* встречается 520 раз в корпусе CANCODE, из них не менее 15 – в разных идиоматических выражениях общим количеством около 60 [3, p.83].

Д.О. Добровольский считает, что работа с корпусом позволяет:

- определить частотность каждой из включаемых в словарь идиом;
- определить, можно ли считать данное словосочетание идиомой;
- определить стандартную с позиции современного узуса форму леммы, уточнив модели управления соответствующих идиом;
- определить набор наиболее значимых вариантов каждой идиомы;
- определить структуру полисемии каждой идиомы и уточнить описание ее конкретных значений;
- описать типичные модификации структуры каждой идиомы;
- определить типичное окружение описываемых идиом и типы контекстов, в которых их появление воспринимается как наиболее естественное [5, с. 23-24].

Он также отмечает, что относительно низкие показатели частоты встречаемости идиомы в корпусе не означают, что данная идиома малоупотребительна. Важно, что корпус является дополнительным ресурсом, позволяющим лексикографу определить степень употребительности каждого отдельного фразеологизма.

Для преподавателя английского языка, обращающегося к корпусу, фактор частотности очень важен.

Наличие в корпусе идиом, включенных в учебные пособия, в случае высокой частотности является весомым доказательством их употребительности в современном дискурсе, что служит дополнительной мотивацией к их изучению и использованию в речи. Низкая частотность, или ее отсутствие в большом корпусе заставляет задуматься о целесообразности изучения именно данной идиомы в практическом курсе английского языка или перевода.

При большом количестве разных контекстов есть возможность найти интересные контексты употребления идиомы в значении, в котором она предлагается для изучения. Это позволит увеличить вероятность ее усвоения обучающимися, т.к. для запоминания незнакомой лексической единицы, по мнению исследователей, с ней необходимо встретиться 6-10 раз в разных контекстах [7].

При высокой частотности «канонической» идиомы шансы обнаружить ее лексическую вариативность и/или выявить употребления, которые не соответствуют принятой языковой норме, возрастают.

Высокая частотность позволяет разработать поисковые задания для обучающихся, способствующие развитию у обучающихся лингво-компьютерной компетенции.

Поэтому целью данной работы была верификация идиоматических выражений и фразеологизмов из учебных пособий в корпусе СОСА и поиск их модификаций, представленных в этом корпусе.

3. Особенности структуры корпусов, созданных в университете Бригама Янга (BYU), и синтаксиса запросов в них

Достоинства корпуса СОСА или CoSA, созданного в 2008 г. в университете Бригама Янга, для пользователей, в том числе не очень опытных, перечислены А. Болтоном [8]:

- это большой, структурированный, постоянно пополняемый корпус общего назначения, охватывающий тексты разных типов и жанров; содержит более 560 миллионов словоупотреблений, встречаемых в более чем двухсот тысяч текстов устной и письменной речи разных стилей и жанров. Ежегодно с 1990 г. корпус пополняется двадцатью миллионами новых лексем. Последнее добавление текстов было сделано/произошло в декабре 2017 г.
- высокая вариативность синтаксиса запросов к корпусу;
- относительная простота в использовании;
- доступность онлайн при бесплатной регистрации;
- стабильность/устойчивость работы с минимальным риском того, что корпус будет удален из сети, «рухнет», будет перемещен на новый сайт или изменит интерфейс.

Относительно изменения адреса сайта и интерфейса прогноз оказался неверным: в 2016 г. произошло существенное изменение интерфейса корпуса по сравнению с предыдущей версией, и упрощение формулировки запросов (в первую очередь за счет удаления многочисленных квадратных скобок в случае частеречного поиска, поиска по лемме или поиска синонимов). В 2019 г. изменился адрес сайта. Но как в случае с другими известными сайтами это произошло с минимальными неудобствами для пользователей: их автоматически перенаправляют на новый сайт.

Архитектура корпуса, основанная на реляционных базах данных (БД) по мнению разработчиков системы имеет важные преимущества.

Во-первых, запросы легко масштабируются. Это означает, что время выполнения запроса практически не зависит (увеличивается очень незначительно) при выполнении однотипных запросов в 100 млн BNC, 560 млн СОСА или 7,47 млрд NOW.

Даже такие сложные запросы, как поиск синонимов глагола *clean* в разных формах в сочетании с любыми существительными с предшествующим определенным артиклем выполняются меньше чем за 2 сек [9, p.166]. Формула запроса, приведенная в указанной публикации, выглядела вот так: `[[=clean]].v* the n*` В новой версии программы тот же запрос формулируется следующим образом: `=CLEAN_v* the NOUN`. Всего находится 1971 пример, удовлетворяющий запросу. В первой сотне самых частотных выражений: *cleaning the house/kitchen/bathroom; scrubbing the floor; mopped/mopped the floor; wipe the dust*.

Во-вторых, архитектуру системы можно усложнять, добавляя, например, новые уровни разметки, что практически не повлияет на скорость выполнения запросов.

Данный корпус может использоваться для получения лингвистических и статистических данных о словоформах. В частности, он позволяет исследователю проанализировать информацию о частотности лексем и грамматических категорий, проследить изменение контекстов в разные периоды времени, получить данные о совместной встречаемости языковых единиц.

Запрос необходимо вводить в форму поиска на главной странице корпуса. Здесь же можно выбирать дополнительные функции (выдача списка лексем, сопоставление слов в разделах, построение виртуального корпуса и т.д.) и просматривать диапазон запросов, которые предлагает корпус. Для поиска в СОСА одной формы слова необходимо писать его строчными буквами, для поиска всех его форм – прописными. Запрос на нахождение фразы вводится с пробелами между словами, а символы «звездочка» (*), «вопросительный знак» (?) используются, если необходимо уточнить букву в написании слова или проанализировать сочетаемость лексемы с другими словами. Следует также обратить внимание на возможность поиска синонимов и антонимов слова, различных частей речи, а также сопоставления фраз и лексем по жанрам и по годам (табл.1).

Таблица 1. Виды поисковых запросов в СОСА

№	Вид запроса	Примеры
1	Single word / Одиночное слово	mysterious, skew
2	Phrase / Фраза	make up, on the other hand
3	Any word more / Добавление любого слова	* than, * bit
4	Wildcard / Запрос с подстановочными знаками	*icity, *break*, b?t?er
5	Lemma (forms) / Формы слова	DECIDE, <u>CURVE</u> n
6	Part of speech / Часть речи	rough NOUN, VERB money
7	Alternants / Антонимы	<u>fast</u> / <u>slow</u>

В СОСА можно не только выполнить поисковый запрос (например, найти формы фразеологизма «water under the bridge») и проанализировать список частотности употребления введенной лексемы, но и построить конкорданс – список всех употреблений одной лексемы в контексте со ссылками на источник, а также сохранить результаты в виде электронного списка. Пользователь может просматривать, изменять и удалять созданные списки в любое время [9].

СОСА может использоваться как непосредственно в учебной аудитории для анализа примеров употребления иноязычных фразеологизмов, так и для составления учебных материалов, упражнений, направленных на совершенствование фразеологического навыка. При таком подходе преподаватели студенты выступают не только как активные участники процесса обучения, но и как исследователи, которые работают с корпусом в поиске лингвистической информации для решения коммуникативной задачи на изучаемом языке.

4. Верификация фразеологизмов в корпусе СОСА

Источником отбора материала послужили словари и учебные пособия, используемые в качестве дидактических материалов при обучении студентов-лингвистов в Гуманитарном институте СПБПУ и на кафедре теоретической и прикладной лингвистики ЮЗГУ (г. Курск), а именно: Language Leader: Upper-Intermediate [10], «Учебные задания по переводу с русского языка на английский» Н.В. Поповой [11], «Английские и русские пословицы и поговорки в иллюстрациях» М.И. Дубровина [12], «3500 английских фразеологизмов и устойчивых словосочетаний» П.П. Литвинова, учебное пособие [13].

Проводимый компьютером анализ фразеологизмов позволяет раскрыть особенности сочетаемости идиом, уточнить их возможные грамматические формы. Кроме того, изучение материалов корпусов помогает не только определить более и менее частотные варианты, но и выявить единицы, которые ещё не зафиксированы в словарях.

Полученные результаты позволили сделать следующие выводы:

Фразеологические единицы достаточно широко употребляются в современном английском языке, поскольку из 535 отобранных лексем в корпусе было обнаружено 262 единицы в своем нормативном виде и 71 – в трансформированном. Наиболее частотные единицы представлены в табл. 2.

Таблица 2. Список наиболее частотных фразеологических единиц, найденных в СОСА

№	Фразеологическая единица	Вид запроса	Кол-во контекстов	Пример употребления
1	Not to give a damn	GIVE * damn	985	I don't give a damn about what anybody thinks about me
2	Now and again	Now and again	758	I like a good plate of rice and vegetables every now and again
3	Empty-handed	Empty-handed	621	Every afternoon, she came home empty-handed and avoided
4	To be scared to death	BE scared to *	585	I was scared to death, also really excited
5	To try desperately	TRY desperately	492	As the river rises, the people of Fargo, North Dakota try desperately to head off disaster
6	Easier said than done.	Easier said than *	452	It's a feat that's easier said than done and one that's been accomplished by ...
7	To be in full swing	BE in full swing	437	Come on, the party's in full swing
8	Turn a blind eye to	TURN * blind EYE to	393	You know, if you turn a blind eye to it, you're a fool
9	To have the upper hand	HAVE * upper *	355	I think Romney has an upper hand because it will be a referendum
10	Ill at ease	Ill at ease	263	Sitting in the stall, he was still ill at ease even after the urgency had passed

Вместе с тем при заданном режиме поиска в корпусе не были обнаружены примеры употребления 202 фразеологизмов. Примеры фразеологизмов, приведенных в пособии Н.В. Поповой [11], не обнаруженные в корпусе СОСА, приведены в табл. 3.

Таблица 3. Список фразеологических единиц из пособия [11], не обнаруженных в СОСА

№	Фразеологическая единица	Формула запроса	Примечания
1	Care for as for the apple of one's eye.	care for as for the apple of * eye; care for as for	--
2	Sit around and wait indefinitely	SIT around and wait indefinitely	Есть примеры при запросе <i>SIT around and wait</i>
3	Actions can boomerang	Actions can boomerang	Есть примеры при запросе <i>can boomerang</i>
4	The game isn't worth the candle	The game * worth the candle	--
5	Things don't work the first time	Things * work the first time	--
6	The pot shouldn't call the kettle black	The pot * call the kettle black	Есть 1 пример с антонимическим значением <i>The pot does call the kettle black</i>
7	Many commanders sink the ship	Many commanders sink the ship; commanders sink the ship	Есть примеры по запросу <i>many cooks spoil the broth</i>

Наличие вариативности идиом признается, в той или иной степени, большинством отечественных и зарубежных исследователей. Например, Т.Н. Федуленкова, исследовавшая вариативность английских глагольных фразеологических единиц (ФЕ) с компонентом *keep*, обнаружила, что константно-вариантная зависимость компонентов характерна почти для половины из 250 исследованных фразеологизмов.

Проведенный нами анализ фразеологизмов позволил обнаружить следующие типичные трансформации:

- расширение состава фразеологизма за счет введения добавочного лексического компонента;
- редукция, или сокращение состава, фразеологизма;
- изменение грамматической формы;
- лексическое варьирование компонентов фразеологизма.

В табл. 4 представлены примеры фразеологизмов из УМК Language Leader (уровень Upper-Intermediate) и формулы запросов, по которым проводился поиск их трансформаций в корпусе COCA. Таблица 5 содержит примеры указанных трансформаций.

Таблица 4. Примеры фразеологизмов из УМК Language Leader

№	Фразеологизм	Формула запроса (количество примеров)
1	actions speak louder than words	ACTION SPEAK LOUD (102)
2	get straight to the point	GET STRAIGHT TO * POINT (28); STRAIGHT TO * POINT (68)
3	burn one's bridges	BURN * BRIDGE (98)
4	keep an open mind	KEEP * * MIND (305)
5	be in two minds	in * minds (11); BE in * minds (8)

Таблица 5. Примеры трансформаций фразеологизмов из таблицы 4

№	Редукция	Расширение	Лексические изменения	Грамматические изменения
1	The three firefighters' action spoke loud and clear	Luciana's speech evaporates into silence, and her last action speaks louder than words...	I intend to let my actions speak louder than anything more that can be said about this truly regrettable moment.	I could have done that. Action speaks louder than words.
2	They were strict and straight to the point and you loved them for that. You respected them.	To get right straight to the point , Don Pancho was a real parrandero...	I suggest you come straight to the point , " he said	I can see it in her eyes. Getting straight to the point , she said
3	--	Both moves burned important bridges between the provinces and federal government, Mr. Gardiner says.	She wasn't ready to burn any bridges yet	And now, in a way, he has burned his boats in terms of any expectation that there would be support for him
4	--	--	He inspires us to keep going on and move forward and keep a positive mind as we go through our struggles in life.	Again, keeping an open mind , Nikki refused to conclude there even was a killer in the classic sense
5	--	That fella seems to be caught in two minds	COM SHERWIN is in several minds.	--

Интересно отметить, что поговорка Actions speak louder than words уже становилась объектом подобного исследования на материале корпуса COCA [15]. Таксономия трансформаций, представленная авторами, приведена в таблице 6.

Таблица 6. Таксономия трансформаций пословицы *Actions speak louder than words*¹

№	Вид трансформации	Количество трансформаций
1	Canonical / Канонический (Нормативный)	34 (22,1%)
2	Minor: Syntactic / Незначительные: Синтаксические	32 (20,8%)
3	Minor: Modal and polarity / Незначительные: Модальность и полярность	22 (14,3%)
4	All substitutions / Все подстановки	42 (27,2%)
5	Substitutions with simple adaptation to context / Подстановки с простой адаптацией к контексту	21 (13,6%)
6	Deletions / Сокращения (Редукция)	7 (4,7%)
7	Additions / Добавления (Расширение состава ФЕ)	53 (34,4)
8	Metaplasm (anti-proverbs, e.g. <i>Words speak louder than action</i>) / «Перевертыши»	3 (1,9%)
9	Author's 'signalling' about a proverbial occurrence (<i>As we all know, actions speak louder than words</i>) / Авторское «указание» на пословицу, которая следует	1

¹ Примечание: Адаптировано из [15, p.139, 141, 142, 143, 145].

Авторы отметили, что она была самой частотной из исследованных ими 6 пословиц (154 примера, из них 34 в канонической форме [15, p.139]. За прошедшее время количество примеров употребления этой пословицы в канонической форме увеличилось больше чем в 2 раза: мы обнаружили 73 примера. Однако, общее количество примеров при нашем способе поиска (табл. 4) оказалось меньше. Он не позволил найти «перевертыши» (*Words speak louder than action*) и комплексные трансформации типа *These electronic transactions speak much louder than words because they reflect actual behavior* [15, p.147].

Полученные нами результаты являются первым шагом к более полному изучению варьирования данных фразеологических единиц, к дальнейшей систематизации знаний о соотношении между фиксированностью лексем и степенью их трансформации.

5. Заключение

Находящийся в открытом доступе корпус английского языка может быть использован в качестве ценного ресурса при изучении фразеологических единиц, а также в качестве источника аутентичных примеров для разработки различного рода упражнений и корпусных заданий нового типа для студентов-лингвистов.

Задания могут касаться как простой проверки наличия определенного фразеологизма в корпусе и, следовательно, его употребляемости в современном английском языке, так и более тонкого анализа, направленного на выявление частотности использования фразеологизмов, контекста их употребления, изменение частотности в течение времени, наличие зафиксированных трансформаций фразеологизма, что должно привести к повышению не только лингвокультурологической, но и лингвокомпьютерной компетенции студентов.

Сравнение найденных нами трансформаций пословицы *Actions speak louder than words* с данными, имеющимися в литературе [14], показывает, что формулировка запроса и методика поиска приводят к разным результатам. Следовательно, целесообразно продолжить такое исследование для других фразеологизмов, включенных в УМК *Language Leader* для того, чтобы получить больше материала для разработки упражнений с опорой на корпусные данные. Методы корпусной лингвистики оказывают значительную помощь в процессе преподавания и должны быть использованы для создания эффективных пособий по обучению иноязычной фразеологии, поскольку лингвистический корпус представляет собой не только базу для проверки различных представлений о языке,

но и ориентирован прежде всего на прикладное изучение фразеологических единиц, анализ их функционирования в аутентичных текстах и реальных ситуациях общения.

Литература

- [1] Sinclair J. *Corpus, concordance, collocation*. Oxford: Oxford University Press, 1991. 170 p.
- [2] Кунин А.В. *Курс фразеологии современного английского языка: Учеб. для ин-тов и фак. иностр. яз. М.: Высш. шк., Дубна: Изд. центр «Феникс», 1996. 381 с.*
- [3] O’Keeffe A., McCarthy M., Carter R. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press, 2007. 315 p.
- [4] Захаров В.П., Богданова С.Ю. *Корпусная лингвистика: Учебник для студентов направления «Лингвистика»*. 2-е изд., перераб. и дополн. СПб.: СПбГУ. РИО. Филологический факультет, 2013. 148 с.
- [5] Добровольский Д.О. *Корпусы текстов и двуязычная фразеогрфия // Вестник Новосибирского государственного педагогического университета*. 2015. № 5 (27). С. 23 – 37.
- [6] McEnery T., Hardie A. *Corpus Linguistics: Method, theory and practice*. Cambridge: Cambridge University Press, 2012. 312p.
- [7] 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.
- [8] Boulton A. *Beyond concordancing: Multiple affordances of corpora in university language degrees // Procedia - Social and Behavioral Sciences*. 2012. № 34. P. 33–38.
- [9] 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.
- [10] Cotton D, Falvey D., Kent S. *Language Leader: Upper-Intermediate*. UK: Person Longman, 2012. – 190 p.
- [11] Попова Н.В. *Учебные задания по переводу с русского языка на английский: учебное пособие по английскому языку*. Санкт-Петербург: Нестор, 2003. 130 с.
- [12] Дубровин М.И. *Английские и русские пословицы и поговорки в иллюстрациях*. Москва: Просвещение, 1993. 349с.
- [13] Литвинов П.П. *3500 английских фразеологизмов и устойчивых словосочетаний*. Москва: Астрель: АСТ, 2007. 285 с.
- [14] Федуленикова Т.Н. *Современная английская фразеология со структурным компонентом keep // Язык и культура*. 2018. № 43. С. 114–125.
- [15] Arnaud P.J.L., Maniez F., Renner V. *Non-Canonical Proverbial Occurrences and Wordplay: A Corpus Investigation and an Enquiry Into Readers’ Perception of Humour and Cleverness. Wordplay and Metalinguistic // Metadiscursive Reflection: Authors, Contexts, Techniques, and Meta-Reflection*. 2015. P.135–160. URL: <https://doi.org/10.1515/9783110406719-007> (дата обращения: 15.03.2019).

Exploring English Phraseology with Corpus Linguistics Methods

I. Komarova, M. Kogan

Peter the Great Saint Petersburg Polytechnic University

The paper discusses prospects of applying corpus linguistics approaches to formation of lexicophraseological skills in linguistics students. Didactic materials based on corpus data should supplement English textbooks which contain lists of phraseological units compulsory for study. These phraseological units were verified in the Corpus of Contemporary American English

(COCA). The paper explains the relevance of COCA for this kind of research. Of total 535 selected for verification phraseologisms 262 were found in their canonical form, and 71 - in modified forms. The paper contains the list of the most frequent phraseologisms, the list of the phraseologisms from the Manual for Translation Practice not found in the COCA, and the list of typical modifications/transformations of phraseologisms from a General English textbook – Language Leader. The comparison of the revealed modified phraseologisms with the data from special literature proves that it is reasonable to continue studying transformations of the selected phraseologisms.

Keywords: corpus linguistics, English phraseology, corpus COCA, search query, verification of phraseologisms, frequency, canonical form, typical transformations, didactic materials

Reference for citation: Komarova I.A., Kogan M.S. Exploring English Phraseology with Corpus Linguistics Methods // Computer Linguistics and Computing Ontologies. Vol. 3 (Proceedings of the XXII International Joint Scientific Conference «Internet and Modern Society», IMS-2019, St. Petersburg, June 19-22, 2019). - St. Petersburg: ITMO University, 2019. P. 40 – 49. DOI: 10.17586/2541-9781-2019-3-40-49

Reference

- [1] Sinclair J. Corpus, concordance, collocation. Oxford: Oxford University Press, 1991. 170 p.
- [2] Kunin A.V. Kurs frazeologii sovremennogo anglijskogo yazyka: Ucheb. dlya in-tov i fak. inostr. yaz. M.: Vyssh. shk., Dubna: Izd. centr «Feniks», 1996. (In Russian).
- [3] O’Keeffe A., McCarthy M., Carter R. From Corpus to Classroom: Language Use and Language Teaching. Cambridge: Cambridge University Press, 2007.
- [4] Zaharov V.P., Bogdanova S.YU. Korpusnaya lingvistika: Uchebnik dlya studentov napravleniya «Lingvistika». 2-e izd., pererab. i dopoln. SPb.: SPbGU. RIO. Filologicheskij fakul'tet, 2013. (In Russian).
- [5] Dobrovol'skij D. O. Text corpora and bilingual phraseography // Novosibirsk State Pedagogical University Bulletin. 2015. Vol.5, No 5. P.23–37.
- [6] McEnery T., Hardie A. Corpus Linguistics: Method, theory and practice. Cambridge: Cambridge University Press, 2012.
- [7] Zahar R., Cobb T., Spada N. Acquiring vocabulary through reading: Effects of frequency and contextual richness // Canadian Modern Language Review. 2001. Vol.57, No 4. P. 541–572.
- [8] Boulton A. Beyond concordancing: Multiple affordances of corpora in university language degrees // Procedia - Social and Behavioral Sciences. 2012. No 34. P. 33–38.
- [9] 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, No 2. P. 159-188.
- [10] Cotton D, Falvey D., Kent S. Language Leader: Upper-Intermediate. UK: Person Longman, 2012.
- [11] Popova N.V. Uchebnye zadaniya po perevodu s russkogo yazyka na anglijskij: uchebnoe posobie po anglijskomu yazyku. Sankt-Peterburg: Nestor, 2003. (In Russian).
- [12] Dubrovin M.I. Anglijskie i russkie posloviцы i pogovorki v illyustracijah. Moskva: Prosveshchenie, 1993. (In Russian).
- [13] Litvinov P.P. 3500 anglijskih frazeologizmov i ustojchivyh slovosochetaniј. Moskva: Astrel': ACT, 2007. (In Russian).
- [14] Fedulenkova T.N. Modern English phraseology with the structural component ‘keep’: Types of variability // Language and Culture. 2018. No 43. P.114–125.
- [15] Arnaud P.J.L., Maniez F., Renner V. Non-Canonical Proverbial Occurrences and Wordplay: A Corpus Investigation and an Enquiry Into Readers’ Perception of Humour and Cleverness. Wordplay and Metalinguistic // Metadiscursive Reflection: Authors, Contexts, Techniques, and Meta-Reflection. 2015. P.135–160. URL: <https://doi.org/10.1515/9783110406719-007> (access date: 15.03.2019).