

ЯЗЫКОЗНАНИЕ
РАЗДЕЛ I. ТЕОРИЯ ЯЗЫКА

УДК 811

DOI: 10.18413/2313-8912-2018-4-2-0-1

Гэ Цзиншэнь¹
Куприева И.А.²

ЛИНГВИСТИЧЕСКАЯ ОСНОВА ОРГАНИЗАЦИИ
БАЗЫ ДАННЫХ

¹) Китайский Океанский Университет
ул. Юйшань, No.5, район Шинан, город Циндао, провинция Шаньдун
Китай
E-mail: jingshen@bk.ru

²) ФГАОУ ВО «Белгородский государственный национальный исследовательский
университет»
ул. Победы, 85, Белгород, 308015
Россия
E-mail: kuprieva@yandex.ru

Печатается при поддержке Гранта Президента для государственной поддержки молодых российских ученых докторов наук РФ МД-2687.2017.6 «Вариативность английского языка Великобритании в условиях влияния политического, социального и экономического факторов (на материале Интернет-дискурса)».

Аннотация. В настоящей статье рассматриваются лингвистические параметры организации базы данных на примере фразеологических единиц современного английского языка. Поскольку количество устойчивых выражений различной степени срощенности в развитых языках исчисляется десятками тысяч, изучение столь обширного фактического материала представляется достаточно сложным и трудоемким процессом. Соответственно, для удобства интерпретации и максимального рассмотрения таких сложных объектов исследования целесообразно оптимизировать процесс исследования, что сводится не просто к делению фразеологизмов на группы, которые определяются с учетом общих понятийных основ классификации и принципа антропоцентризма, но и к выбору универсальных критериев, которые впоследствии позволят выявить общие и частные основания для последующих сопоставительных исследований. Таким образом, данная статья ставит своей целью описание некоторых лингвистических стратегий организации баз данных, которые актуальны для использования в качестве контента параллельного мультязыкового корпуса.

Ключевые слова: фразеологизмы, корпус, база данных, лингвистическая составляющая, фразообразующий компонент.

Ge Jingshen¹
Irina A. Kuprieva²

LINGUISTIC BASIS FOR THE ORGANIZATION
OF A DATABASE

¹⁾ Ocean University of China
No.5 Yushan Rd., Shinan District, Qingdao City, Shandong Province
China
E-mail: jingshen@bk.ru

²⁾ Belgorod State National Research University
85 Pobeda St., Belgorod, 308015
Russia
E-mail: kuprieva@yandex.ru

Supported by the President's Grant for the state support of young Russian Doctors of Sciences MD-2687.2017.6 "The variability of the English language in Great Britain influenced by political, social and economic factors (based on Internet discourse)".

Abstract. The article covers some linguistic parameters of the organization of a database of phraseological units in the modern English language. Since the number of stable expressions of varying degrees of coalescence in certain languages is extremely large, the study of such a vast factual material seems to be a rather complex and time-consuming process. Accordingly, for the convenience of interpretation and better consideration of such complex objects of research, it is expedient to optimize the research process, which is reduced not only to the division of phraseological units into groups that are defined taking into account the general conceptual bases of classification and the principle of anthropocentrism, but also to the selection of universal criteria that will subsequently reveal general and private grounds for subsequent comparative studies. Thus, this article aims to describe some linguistic strategies for organizing databases that are relevant for use as content of a parallel multilingual corpus.

Key words: phraseological units; corpus; database; linguistic ingredient; phrase-forming component.

Основная часть:

Исследования в области фразеологии независимо от давления той или иной научной догмы, инновационного инструментария и интегративного подхода тем не менее вскрывают все больше и больше лакун для детального описания. Речь идет преимущественно о той скрытой от лингвистического взора стороне вопроса, которая ассоциирована с восприятием, переработкой, категоризацией и вербальным отражением синкретизма языка и культуры. Это позволяет говорить о вербальном отражении этнокультурного

кода в коннотации, ассоциатах, символикe, коррелирующих в сознании носителей языка с теми или иными объектами.

Как справедливо отмечает Н.Ф. Алефиренко, язык и культура выступают в качестве взаимопредполагающих сущностей, где язык являет собой основной инструмент познания культуры, выступая в качестве вербального коррелята национальной культуры, а культура, обретая в языке семиотическое воплощение, оживает в нем и динамично функционирует [1].

Идентичного мнения придерживаются сторонники лингвокультурологического и антропоцентрического направлений, отмечая, что в лексическом и фразеологическом фонде языка «находят отражение национальный характер, психический склад народа, его история и культура» [8, с. 85], таким образом сам по себе язык несет в себе особенности национальной специфики носителей языка [2, с. 416].

С.Г. Тер-Минасова связывает вербализацию народной мудрости и опыта непосредственно с существованием фразеологизмов как аккумуляторов национальной культуры и мировоззрения этноса [6, с. 80]. Однако, как показывают многочисленные исследования в области фразеологического фонда нации, определение фразеологизма не может быть однозначным, что позволяет не разграничивать устойчивые выражения различного уровня сращенности и представлять объект исследования как целый довольно массивный по объему пласт культурно-исторического фонда языка. Это обстоятельство также дает возможность развития перспективных корпусных сопоставительных исследований на материале нескольких дистантных языков. Для таких исследований требуется грамотный выбор критериев, на основании которых возможен компаративный анализ.

В предпринимаемом нами исследовании в качестве основного принимается определение фразеологических единиц, предложенное Н.М. Шанским, который полагает, что фразеология ассоциирована со всеми устойчивыми выражениями и включает в качестве основного объекта исследования крылатые выражения, пословицы, поговорки и проч. Нам близка точка зрения лингвиста о том, фразеологизм представляет собой воспроизводимую в готовом виде языковую единицу, включающую два и более ударных компонента словного характера, фиксированную по семантике, композиции и структурной организации [8, с. 156].

Таким образом, принимая в качестве основного рабочего определения ФЕ формулировку, предложенную Н.М. Шанским, мы, опираясь на мнения авторитетных фразеологов, определяем устойчивые словосочетания как единицы, отвечающие следующим критериям:

- критерию расчлененности, который определяет способность конstituентов ФЕ функционировать самостоятельно;

- критерию количественности, который определяет ФЕ как комбинацию из двух и более ударных слов;

- критерию композиционной идентичности, определяющий структурное сходство любого фразеологизма и словосочетания;

- критерию идиоматичности как показатель семантического единства компонентов, предопределяющего возникновение нового значения, нерелевантного для каждого из элементов в отдельности;

- критерию устойчивости, который определяет некоторую долю инвариативности композиционного плана;

- критерию воспроизводимости в готовом виде, который определяет ФЕ в качестве готовых для употребления в речи элементов;

- критерию коммуникативной направленности, определяющему прагматический потенциал ФЕ;

- критерию номинативности, определяющему способность фразеологизма с учетом семантической спаянности его конstituентов выступать в качестве номинанта объекта, явления, процесса и называть предмет или явление, давать им характеристики;

- критерию экспрессивности как способности фразеологизма передавать дополнительные оценочные смыслы.

Таким образом, в качестве фразеологических единиц в настоящем исследовании выступают единицы вторичной номинации в комплексе составных элементов, особенно если речь идет об устойчивых выражениях, не допускающих трансформации качествен-

ного и/или количественного состава. Такие устойчивые словосочетания высокой степени сращенности выполняют номинативно-характеризующую функцию, в синтаксическом аспекте выступают как единое целое, переосмыслены, а для актуализации такого значения требуют коллокации лексем определенных тематических групп. При этом важным и интересным обстоятельством употребления фразеологизмов является их некоторая прецедентность, что предполагает отсутствие необходимости объяснения контекста при употреблении, его внедрения ФЕ в коммуникативную ситуацию без ущерба для восприятия транслируемого значения. Здесь также необходимо отметить, что большинство фразеологизмов известны носителям языка. В случае их незнания, коммуникант, как правило, требует уточнения или разъяснения семантики такого словосочетания, то есть запрашивает контекст, из которого ясно, что это особая единица, а не сочетание слов, и который дает возможность ее правильно интерпретировать. При этом не теряются не только денотативное содержание, но и коннотативные свойства ФЕ – экспрессивность, стилистическая и эмоциональная окраска. Таким образом, важной характеристикой фразеологизмов является их взаимосвязь и обусловленность культурной традицией и историей народа-создателя.

Учет представленных критериев является обязательным для настоящего исследования, которое ставит своей целью многогранное параллельное исследование групп фразеологизмов английского, китайского, русского, украинского языков.

Итак, представленные критерии являются общими для всех исследуемых нами фразеологических единиц независимо от степени сращенности и языковой группы. Само по себе рабочее определение термина “фразеологизм” или “фразеологическая единица” в сочетании с указанными критериями позволяет

накопить обширный пласт эмпирического материала в любом языке мира. Однако эта задача в настоящей лингвистической разработке является исключительно первостепенной. Более того, получаемый посредством сплошной выборки поток фактических данных слишком обширен для их обработки даже в одном отдельно взятом языке, не говоря уже о сопоставительных исследованиях. Последние требуют особой, детальной проработки дополнительных критериев для последующей типологии и поиска точек соприкосновения.

Представляется, что поиск таких оснований для мультиязыкового компаративного анализа возможен с учетом данных лингвокультурологии, лингвоаксиологии, когнитивного направления и принципа антропоцентризма. Учитывая тот факт, что перспективное корпусное исследование нацелено на обработку языкового материала сразу четырех языков, опишем общие параметры организации и тэгирования лингвистической информации, которые впоследствии и в перспективе разработки будут детализироваться согласно языковой группе.

Итак, полученные при помощи сплошной выборки из аутентичных источников фактических данных фразеологические единицы составляют довольно внушительный по тематической дифференциации, структурной организации, компонентному составу языковой пласт, крайне неудобный для непосредственного лингвистического анализа. Такой эмпирический массив целесообразно делить по принципу основного компонента, являющегося структурным элементом такого языкового единства. Подобная тематическая ранжировка по критерию “тематическая соотнесенность фразеобразующего элемента” имеет под собой антропоцентрические и когнитивные основания, подтверждающие факт словесного выражения метафоры и

символизма, организованного именно в человеческом сознании благодаря когнитивным механизмам восприятия и отражения мира.

Предварительно представив группы фразеологизмов по тематической соотнесенности фразообразующих элементов, следует произвести анализ лексикографического толкования каждой из единиц при учете существования последнего. Для этой работы следует опираться преимущественно на толковые и фразеологические словари для выявления данных о мотивированных и немотивированных номинациях. В случае с немотивированными номинациями целесообразно принимать в расчет этимологические данные и предысторию вопроса. На этом же этапе, если позволяют лексикографические сведения, следует уточнить коннотацию (или положительное/отрицательное базовое оттеночное значение фразеологизма). Это значение определяется с учетом ценностной шкалы лингвокультуры, которая предопределяет ценностные ориентиры и оценку того или иного явления, описанию которого способствует тот или иной фразеологизм. В таком случае на первый план выходят лингвокультурные и лингвоаксиологические принципы, которые позволяют обозначить положительную или отрицательную базовую полярность фразеологической единицы в ее исходном оригинальном значении.

Для составления последующего конкорданса в перспективной корпусной разработке для фразеологической единицы осуществляется подбор контекстов, которые влияют на узус фразеологизма, в том числе меняют коннотацию последних. Таким образом выявляются оценочные параметры фразеологизмов на контекстуальном уровне. И, наконец, указанный количественный параметр конкорданса каждой ФЕ позволит просмотреть динамику употреблений и предоставить компьютерный учет частотности

фразеологизмов.

Таким образом, алгоритм работы с мультязыковым фактическим материалом ведется по следующему принципу согласно языковой группе: ранжирование согласно тематической принадлежности, лексикографический, контекстуальный и концептуальный анализ семантики каждой ФЕ. Результаты такой лингвистической процедуры помещаются в таблицы соответствующего вида. Впоследствии данные таблиц ложатся в основу баз данных по каждой языковой группе. Учитывая то обстоятельство, что в процесс работы в качестве библиографической базы включены лексикографические источники, каждая таблица снабжена списком источников используемой литературы. При организации корпуса это позволит поисковому движку выдавать сведения о цитируемых источниках.

Отметим, что формат настоящей публикации не позволяет полноценно описать работу с каждой фразеологической единицей в отдельности, поскольку указанная процедура является весьма трудоемким и объемным процессом. Поэтому видится целесообразным проиллюстрировать работу с фактическим материалом на нескольких примерах фактического материала и лапидарно осветить результаты работ.

В качестве иллюстративного материала выберем фразеологические единицы с соматическим компонентом. Итак, например фразеологизм *from head to toe*, отобранные нами в качестве составляющей исследовательского корпуса, имеет следующие значения в разных толковых словарях:

Macmillan Dictionary: used for emphasizing that you mean all of someone's body: They were covered in mud from head to toe [10];

Oxford Dictionary: all over one's body: I was shaking from head to toe (Oxford, 2017);

Cambridge Dictionary: completely covering your body: The dog was covered in mud from head to toe [10];

Longman Dictionary: if a person is dressed or covered in something from top to toe, they are completely dressed or covered in it: They were covered in mud from top to toe [10].

Если проанализировать словарные дефиниции, возможно определить, что рассматриваемая фразеологическая единица имеет мотивированную номинацию, то есть ее общее значение вытекает из значения одного из ее элементов. В данном случае знаменательных фразеобразующих элементов два, более того, они оба передают прямое значение, которое в совокупности при условии составления фразеологической единицы приобретают вторичный смысл. Также рассмотрение словарных дефиниций показывает способность элемента *head* и элемента *toe* иметь субституты и выступать во фразеологических единицах в вариантах: *from head to foot, from top to toe*. Однако они являются абсолютными синонимами и это никоим образом не влияет на их коллокацию.

Анализ частотности фразеологической единицы, связанный с рекуррентностью этой единицы в современной устной речи, показывает, что относительно других фразеологических единиц, имеющих более выраженное метафорическое значение, эта единица достаточна частотна. Думается, что

простота ее восприятия за счет актуальной мотивированности значения как раз обуславливает данное обстоятельство.

Анализ функционального значения говорит в пользу прецедентности единицы, то есть способности без каких-либо других языковых средств полноценно передавать смысл.

Например, название салона красоты в предложении: *From Head To Toe is a spa studio that provides customized services for women and men using natural and organic products (http://fromheadtoeola.com/about/)* позволяет передавать положительную коннотацию и посредством контекста передавать значение завершенности и полноты образа. Это значение отлично от словарных данных в примерах, где «с ног до головы», как правило, ассоциировано с отрицательным модусом.

С опорой на выявленное в результате контекстуального анализа функциональное значение и опыт изучения лексикографических толкований, а также принимая во внимание совокупное экстралингвистическое знание, переходим к иллюстрации ключевых фрагментов анализа в таблицы для последующей организации ее в базы данных.

Проиллюстрируем полученные данные, оформленные для иллюстративности в виде таблицы.

Таблица
Результат лингвистического анализа фразеологизмов английского языка с соматическим компонентом. Английский язык
Table

The result of the linguistic analysis of English phraseological units with a somatic component. The English Language

Фразеологическая единица	Результат лексикографического анализа	Результат контекстуального анализа
from head/top to toe/foot	положительная коннотация -	положительная коннотация +
	отрицательная коннотация -	отрицательная коннотация +
	мотивированность значения +	мотивированность значения +

Итак, на основании вышеизложенного, в результате контекстуального, лексикографического анализа становится очевидным тот факт, что фразеологизм соответствует определенным параметрам, которые позволяют ему отличаться от других фразеологических единиц. Такие параметры, способствующие уточнению исследовательского корпуса, будут актуальны для всех рассматриваемых групп языков, поскольку составляют принципиальную основу последующего компаративного анализа.

На основе полученных данных, как было указано выше, будет организована база данных, которая представляет собой структурированный исследовательский корпус фразеологизмов английского языка, накопленный при помощи сплошной выборки из соответствующих аутентичных источников фактического материала, на основании способности таких словосочетаний отвечать перечисленным критериям. В ней также содержатся сведения о верификации отобранных единиц на предмет соответствия объекту исследования посредством лексикографического, контекстуального анализа каждой фразеологической единицы.

Приведем приблизительное описание перспективной базы данных, которая станет одним из этапов формирования мультязыкового фразеологического корпуса. Итак, остановимся на описании основных сущностей и их атрибутов базы данных

«БИБЛИОГРАФИЯ» – сущность, идентифицирующая источники, из которых производилась выборка лексем.

Код_Библ – код источника, Целый, первичный ключ.

Авторы – ФИО авторов источника, Текстовый (255), не обязательный.

Название – название литературного источника, Текстовый (255), обязательный.

Тип – тип литературного источника, Текстовый(255), обязательный

Доп_Свед – поле, содержащее дополнительные сведения о переводчиках, редакторах и издании, Текст (255), не обязательный.

Город – название города издания источника, Текст (100), не обязательный.

Издательство – название издательства источника, Текст (100), не обязательный.

Год – год издания источника, Целый, не обязательный.

Стр – количество страниц в источнике, Целый, не обязательный.

URL – гиперссылка на источник с указанием режима доступа, Текстовый (255), не обязательный;

Тип_Библ – вид источника, Целый, внешний ключ.

«ТИПЫ_БИБЛ» – сущность, идентифицирующая виды источников для выборки.

Код_Типа – код вида источника, Целый, первичный ключ.

Название – название вида источника, Текстовый (100), обязательный.

«АНГЛИЙСКИЕ_ФЕ» – сущность, идентифицирующая фразеологические единицы английского языка.

Код_ФЕ – код фразеологической единицы, Целый, первичный ключ.

ФЕ – фразеологическая единица, Текстовый (255), обязательный.

«АНГЛИЙСКИЕ_ЛЕК_А» – сущность, идентифицирующая результаты лексикографического анализа фразеологической единицы английского языка.

Код_ЛА – код результата лексикографического анализа, Целый, первичный ключ.

Результат – содержание результата анализа, Текстовый (255), обязательный.

Код_ФЕ – код фразеологической единицы, к которой принадлежит результат анализа, Целый, внешний ключ.

Код_Библ – код источника, из которого извлечен результат анализа, Целый, внешний ключ.

«АНГЛИЙСКИЕ_КОНТ_А» – сущность, идентифицирующая результаты

контекстуального анализа фразеологической единицы английского языка.

Код_КА – код результата контекстуального анализа, Целый, первичный ключ.

Результат – содержание результата анализа, Текстовый (255), обязательный.

Код_ФЕ – код фразеологической единицы, к которой принадлежит результат анализа, Целый, внешний ключ.

Код_Библ - код источника, из которого извлечен результат анализа, Целый, внешний ключ.

Мотивированное значение – флаг, определяющий задействует ли данная фразеологическая мотивированное значение, Логический, обязательный.

Немотивированное значение – флаг, определяющий задействует ли данная фразеологическая единица немотивированное значение, Логический, обязательный.

Положительная коннотация – флаг, определяющий задействует ли данная фразеологическая единица положительную коннотацию, Логический, обязательный.

Отрицательная коннотация – флаг, определяющий задействует ли данная фразеологическая единица отрицательную коннотацию, Логический, обязательный.

Заключение:

Как видно из описания параметров базы данных и ее контента, для создания мультязыкового параллельного корпуса фразеологизмов требуется оптимальная организация результатов анализа эмпирического материала и четкого соответствия критериям. В противном случае при несоответствии критериям фразеологизмов разных языковых групп в мультязыковом параллельном корпусе будет наблюдаться диссонанс, что существенно осложнит процедуру компаративного анализа и доставит неудобства в последующей исследовательской работе.

Информация о конфликте интересов: авторы не имеют конфликтов интересов для декларации.

Information of conflict of interests: authors have no conflicts of interests to declare.

Список литературы

1. Алефиренко Н.Ф. Язык, познание и культура. – Волгоград: Перемена, 2006. – 227 с.
2. Вежбицкая А. Язык. Культура. Познание / А. Вежбицкая. – Издательство: М.: Русские словари, 1996. – 416 с.
3. Водяха А.А. Эмотивный потенциал фразеологических единиц // Взаимодействие языковых уровней в сфере фразеологии; доклад – Волгоград, перемена, 1996. – С. 30 – 32.
4. Вакуров В.Н. Основы стилистики фразеологических единиц. На материалах советского фельетона / В. Н. Вакуров. – М.: Изд-во МГУ, 1983. – 175 с.
5. Кунин А.В. Курс фразеологии современного английского языка. – М.: Высшая школа, 1986. – 336 с.
6. Тер-Минасова С. Г. Язык и межкультурная коммуникация ... / С.Г. Тер-Минасова. – М.: Слово, 2000 — 624 с.
7. Никитин М.В. Курс лингвистической семантики. Учебное пособие. 2-е издание, дополненное и исправленное. СПб.: Изд-во РГПУ им. А.И.Герцена, 2007. – 819 с.
8. Шанский Н.М. Фразеология современного русского языка [Учеб. пособие для филол. фак. ун-тов] / Н.М. Шанский. – Москва Высш. школа, 1963. – 156 с.
9. Cambridge Dictionary / Cambridge University Press. – Cambridge, 2017. – URL: <http://dictionary.cambridge.org/ru/> (дата обращения 15 апреля 2018).
10. The Longman Dictionary of Contemporary English / Pearson Education Limited. – Edinburgh, 2016. – URL: <http://www.ldoceonline.com/> (дата обращения 15 апреля 2018).
11. Macmillan Dictionary / Macmillan Publishers Limited. – London, 2009- 2017. – URL: <http://www.macmillandictionary.com/> (дата обращения 15 апреля 2018).
12. Oxford Dictionaries: Language matters / Oxford University Press. – Ashburn (Virginia), 2017. – URL: <http://www.oxforddictionaries.com/> (дата обращения 15 апреля 2018).

References

1. Alefirenko, N. F. (2006), *Jazyk, poznanie i kultura* [Language, Knowledge and Culture], Peremena, Volgograd, Russia. [in Russian].
 2. Vezhbitskaya, A. (1996), *Jazyk. Kultura. Poznanie* [Language, Culture, Knowledge], Russkie slovari, Moscow, Russia. [in Russian].
 3. Vodyah, A. A. (1996), “The Emotional Potential of Phraseological Units”, *Vzaimodeistvie jazykovyh urovneii v sfere frazeologii*, Peremena, Volgograd, Russia, P. 30 – 32. [in Russian].
 4. Vakurov, V. N. (1983), *Osnovy stilistiki frazeologicheskikh edinit. Na materialah sovetskogo feletona* [Basics of Stylistics of Phraseological Units. On the Materials of the Soviet Feuilleton], MGU, Moscow, Russia. [in Russian].
 5. Kunin, A. V. (1986), *Kurs frazeologii sovremenno angliiskogo jazyka* [Course of Phraseology of Modern English], Vysshaja shkola, Moscow, Russia. [in Russian].
 6. Ter-Minasova, S. G. (2000), *Jazyk i mezhkulturnaja kommunikatsija* [Language and International Communication], Slovo, Moscow, Russia. [in Russian].
 7. Nikitin, M. V. (2007), *Kurs lingvisticheskoj semantiki. Uchebnoe posobie. 2-e izdanie, dopolnennoe i ispravlennoe* [Course of Linguistic Semantics. Tutorial. 2nd edition, supplemented and amended], RGPU im. A. I. Gercena, Saint Petersburg, Russia. [in Russian].
 8. Shansky, N. M. (1963), *Frazeologija sovremenno russkogo jazyka* [Phraseology of Modern Russian], Vysshaja shkola, Moscow, Russia. [in Russian].
 9. Cambridge Dictionary, Cambridge University Press (2017), [Online], available at: <http://dictionary.cambridge.org/ru/> (Accessed April 15, 2018).
 10. The Longman Dictionary of Contemporary English (2016), Pearson Education Limited, Edinburgh [Online], available at: <http://www.ldoceonline.com/> (Accessed April 15, 2018).
 11. Macmillan Dictionary, Macmillan Publishers Limited (2017), London, [Online], available at: <http://www.macmillandictionary.com/> (Accessed April 15, 2018).
 12. Oxford Dictionaries: Language matters (2017), Oxford University Press, Ashburn (Virginia) [Online], available at: <http://www.oxforddictionaries.com/> (Accessed April 15, 2018).
- Гэ Цзиншэнь**, старший преподаватель, кандидат филологических наук, Китайский Океанский Университет.
- Ge Jingshen**, Assistant Professor, PhD in Philology, Ocean University of China (OUC).
- Куприева Ирина Анатольевна**, доктор филологических наук, профессор кафедры английской филологии и межкультурной коммуникации Института межкультурной коммуникации и международных отношений ФГАОУ ВО «Белгородский государственный национальный исследовательский университет».
- Irina A. Kuprieva**, Doctor of Philology, Professor, Belgorod State National Research University.