

Гурьянов Игорь Олегович

**ИСПОЛЬЗОВАНИЕ СТАТИСТИЧЕСКИХ ПРОГРАММ ПРИ ИССЛЕДОВАНИИ
ФРАЗЕОЛОГИЧЕСКИХ ЕДИНИЦ КНИЖНОГО СТИЛЯ В РУССКОМ И АНГЛИЙСКОМ ЯЗЫКАХ**

Статья посвящена особенностям использования компьютерных программ, позволяющих собирать для дальнейшего анализа статистику употребления языковых единиц в диахроническом аспекте. Материалом исследования послужили фразеологические единицы книжного стиля из двуязычных словарей. Были изучены программы Google Ngram Viewer, поисковой механизм русского национального корпуса языка и британский национальный корпус. В ходе исследования выявлены недостатки современных статистических программ.

Адрес статьи: www.gramota.net/materials/2/2015/2-2/16.html

Источник

Филологические науки. Вопросы теории и практики

Тамбов: Грамота, 2015. № 2 (44): в 2-х ч. Ч. II. С. 63-65. ISSN 1997-2911.

Адрес журнала: www.gramota.net/editions/2.html

Содержание данного номера журнала: www.gramota.net/materials/2/2015/2-2/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: phil@gramota.net

ON THE CONTENT OF THE CHAPTERS WITH BOTANICAL VOCABULARY OF THE ANCIENT CHINESE DICTIONARY «ERYA»

Gur'yan Natal'ya Viktorovna, Ph. D. in Philology
School of Education of the Far Eastern Federal University
guryan-natalia@yandex.ru

The article is devoted to analyzing the content of the two chapters of the ancient Chinese dictionary –Erya” (III century B.C.) containing botanical vocabulary – “Explanation of herbs” and “Explanation of trees”. The author presents the translation of the fragments of the mentioned chapters; analyzes the categories of plants represented in the dictionary. The paper identifies what realia of the flora are registered in the dictionary (nominations of the varieties and species, fruits and parts of the plants, forms of vegetation, related concepts). The researcher also focuses on the naïve botanical conceptions of the ancient Chinese represented in the dictionary.

Key words and phrases: ancient Chinese dictionary –Erya”; botanical vocabulary; “Explanation of herbs”; “Explanation of trees”; naïve classification of plants.

УДК 8; 81'373.47

Филологические науки

Статья посвящена особенностям использования компьютерных программ, позволяющих собирать для дальнейшего анализа статистику употребления языковых единиц в диахроническом аспекте. Материалом исследования послужили фразеологические единицы книжного стиля из двуязычных словарей. Были изучены программы Google Ngram Viewer, поисковой механизм русского национального корпуса языка и британский национальный корпус. В ходе исследования выявлены недостатки современных статистических программ.

Ключевые слова и фразы: фразеология; корпус; статистика; трансформации; книжный стиль.

Гурьянов Игорь Олегович

Казанский (Приволжский) федеральный университет
igor.goor@gmail.com

ИСПОЛЬЗОВАНИЕ СТАТИСТИЧЕСКИХ ПРОГРАММ ПРИ ИССЛЕДОВАНИИ ФРАЗЕОЛОГИЧЕСКИХ ЕДИНИЦ КНИЖНОГО СТИЛЯ В РУССКОМ И АНГЛИЙСКОМ ЯЗЫКАХ[©]

Вопрос применения математических и статистических методов исследования в лингвистике является давним вопросом на стыке наук. На возможность использования математических методов при исследовании языковых особенностей указывал И. А. Бодуэн де Куртенэ. Несмотря на это, только в начале XX века с появлением структурной лингвистики ученые начали экспериментировать с применением математических методов для анализа языковых процессов. В. А. Звегинцев выделяет следующих ученых – математиков В. Я. Буныковского, А. А. Маркова. Языковые элементы рассматривались ими как подходящий иллюстративный материал для построения количественных методов или для статистических теорем. Ученые совершенно не задавались вопросом, представляют ли интерес результаты такого исследования для лингвистов [2, с. 115].

Современные компьютерные технологии дают большой простор для проведения различных исследований. Большие вычислительные мощности позволяют обрабатывать большие массивы данных. Нами была поставлена задача оценить возможность применения данных статистических методов для поиска трансформаций фразеологических единиц. В качестве материала исследования использовались фразеологические единицы книжного стиля, отобранные из двуязычных словарей. В ходе исследования были использованы следующие ресурсы: «Google Ngram viewer» [8], «Ruscorpora» – корпус русского языка [5], «BNC» – британский национальный корпус [7].

Применение методов количественной лингвистики во фразеологии является сложной задачей. Это объясняется тем, что основной предмет изучения – фразеологическая единица – обладает следующими характеристиками: воспроизводимость, устойчивость, сверхсловность (раздельнооформленность), принадлежность к номинативному инвентарю языка.

Современные российские и зарубежные ученые рассматривают использование фразеологических единиц в контексте. Обзор литературы по данному вопросу показал, что большинство ученых выделяют два вида употребления фразеологизмов: узуальный (словарная форма единицы) и окказиональный (трансформация единицы на одном из языковых уровней). Ученые не определили свое отношение к окказиональному способу употребления фразеологизмов: «С одной стороны, подобные преобразования рассматриваются как отклонения от нормы и, следовательно, подвергаются открытой критике. С другой стороны, контекстуальное обыгрывание фразеологических единиц приветствуется как неисчерпаемый источник творческой фантазии носителей языка» [1, с. 4].

В ходе нашего исследования нами были отобраны 23 фразеологические единицы книжного стиля из англо-русского словаря под редакцией А. В. Кунина и 30 фразеологических единиц книжного стиля из русско-английского словаря под редакцией С. И. Лубенской.

Книжный стиль речи считается наиболее давним видом речи, так как возник вместе с первыми памятниками письменности. Именно поэтому книжные фразеологизмы являются наиболее связными и реже поддаются изменениям, в сравнении с фразеологическими единицами других стилей.

Э. М. Солодухо в своих работах по изучению интернациональных фразеологизмов отмечал принадлежность некоторых фразеологизмов как к интернациональному, так и к книжному пласту фразеологии. Эта позиция подтверждается тем, что наиболее ранними памятниками европейской письменности считаются религиозные труды, включая Библию [6, с. 35].

Также отмечается, что большое количество книжных фразеологических единиц относятся к устаревшей лексике.

Для проверки этой гипотезы мы обратились к двум ресурсам, которые используются при обработке статистических данных языка.

Первая программа *Google Ngram Viewer* позволяет отслеживать частотность употреблений не только отдельных фраз, но и словосочетаний, а также фразеологических единиц. Программа производит поиск по обширной библиотеке *Google Books* и отмечает количество употреблений рассматриваемой единицы. Отличительная особенность ресурса состоит в том, что на выбор представляется большое количество языков: английский, русский, испанский, немецкий, китайский. С помощью данной программы нами был подтвержден тезис о том, что большинство фразеологических единиц книжного стиля относятся к устаревшим выражениям. В английском языке у всех отобранных единиц пики частности употребления относились к XIX веку и к началу XX. Приведем примеры:

The wings of Azrael – пик употребления – 1923 год [3, с. 18].

Appeal to Caesar – пик употребления – 1824 год [Там же, с. 59].

Обетованная земля – пик употребления – 1847 год [4, с. 287].

Почивать на лаврах – пик употребления – 1824 год [Там же, с. 170].

В английском языке выявилась следующая особенность – при добавлении артикля в конструкцию фразеологизма пик употребления может сместиться в сторону.

The burden of years – время лет [3, с. 47], пик приходится на 1856 год;

The burden of the years – время лет, пик приходится на 1916 год [Там же].

The cask of Danaides – бочка Данаид [Там же], пик приходится на 1912 год.

The cask of the Danaides – бочка Данаид [Там же, с. 68], пик приходится на 1823 год.

При исследовании частотности употребления русских фразеологических единиц возникла проблема правильного написания фразеологических единиц, так как в 1917-1918 году произошла реформа орфографии. База данных *Google Ngram Viewer* состоит из оригинальных письменных источников, переведенных в электронный вид, что осложняет выявление частотности употребления фразеологических единиц.

Стоит отметить особенность использования программы *Google Ngram Viewer*. В окно поиска могут вводиться только узуальные образцы единиц, поиск трансформаций, которым подвергаются фразеологизмы, практически невозможен.

Согласно классификации Е. Ф. Арсентьевой, существуют следующие типы трансформации:

1. Фразеологический каламбур – обыгрывание переносного значения фразеологизма и буквального значения его компонентов.

2. Замена лексического компонента/компонентов на лексему, связанную с окказионально варьируемым компонентом системными отношениями (синонимией, антонимией, принадлежностью к одному лексико-семантическому полю или одной тематической группе).

3. Вклинивание – 1) вклинивание, в котором вклинивающиеся слова синтаксически связаны с компонентами фразеологической единицы, и 2) вклинивание, в котором вклинивающиеся слова синтаксически не зависят от компонентов фразеологизма.

4. Добавление переменного компонента/компонентов к концу или началу оборота.

5. Эллипсис – сокращение количества компонентов фразеологизма.

6. Фразеологический повтор.

7. Расширенная метафора – создание нескольких ассоциативных метафор, связанных с базовым значением фразеологизма.

8. Фразеологическое насыщение контекста – простой тип представляет собой одновременное использование двух или более фразеологизмов; второй тип подразумевает использование сразу нескольких типов контекстуальных трансформаций фразеологических единиц [1, с. 8].

Если рассматривать данные трансформации как материал для использования в программе *Google Ngram Viewer*, то возникают следующие проблемы:

1. Программа не может отличить буквальное значение от переносного, поэтому стоит ограничиться лишь структурными изменениями фразеологических единиц, такими как вклинивание, лексическая замена, эллипсис и добавление компонента.

2. При вклинивании фразеологизма его компоненты могут быть разнесены в пределах предложения, но встречаются случаи, когда разрыв выходит за пределы предложения. В результате, программа может пропустить данную трансформацию.

3. В русском языке возникает сложность при поиске единиц в дореволюционных источниках.

Основным источником сбора статистической информации для исследования является национальный корпус русского языка (ruscorpora.ru). Преимущество отечественного корпуса состоит в том, что частотность употребления каждой единицы можно рассмотреть как в синхронии, так и в диахронии. Наиболее важной особенностью следует считать возможность настраивать расстояние удаленности компонентов друг от друга, что позволяет осуществлять поиск единиц, трансформированных с помощью вклинивания, разрыва, лексической замены. Приведем примеры трансформаций, найденных в корпусе.

Пример вклинивания в структуру фразеологической единицы местоимения «какой-нибудь»: *На смену романтическим порывам энтузиастов приходят здравомыслящие работяги, пытающиеся возвести на своих шести сотках **башиню** из какой-нибудь **кости**, а то и цельный воздушный замок* (Дмитрий Бавильский. Новые стихи) [5].

*Но **черное газетное дело** давно уже опостылело ему* (А. И. Куприн. Яма) [Там же]. В данном примере мы наблюдаем расширение смысла фразеологической единицы с помощью вклинивания смыслового компонента, выраженного прилагательным «газетное».

Немаловажным аспектом, влияющим на удобство обработки и пользования программой, является вкладка «статистика» и «распределение по годам». Система строит графики частотности употребления данной единицы по временной оси. Во вкладке «статистика» мы наблюдаем статистику по метатрибутам, стилистическое направление текста, пол автора, что делает интересным исследование с точки зрения гендерной лингвистики, и другие показатели.

Британский национальный корпус языка (natcorp.ox.ac.uk) [Там же] не располагает столь подробной статистической информацией при поиске языковых единиц, что отчасти компенсируется сторонними приложениями – *BNC Web at Oxford*, *BNC Web at Lancaster University*. Для использования этих приложений необходимо зарегистрироваться во внутренней сети этих университетов, что в рамках исследования пока невозможно.

Стоит отметить, что, несмотря на большое количество статистических данных, полученные результаты требуют тщательной проверки, так как среди единиц с переосмысленным значением могут быть и свободные сочетания слов.

Современное программное обеспечение позволяет облегчить поиск фразеологических единиц и способов их использования. Основная проблема при поиске фразеологизмов и их трансформаций заключается в невозможности для программы разделить прямое и переносное значение слова и словосочетания. При дальнейшем развитии искусственного интеллекта и создании самообучающихся систем поиска данная проблема может быть решена.

Программы по выведению статистики при изучении фразеологических единиц значительно облегчают труд ученого по исследованию фразеологических единиц в диахроническом аспекте.

Тезис о том, что книжные фразеологические единицы принадлежат к устаревшему пласту лексики, подтвердились лишь в английском языке, где пики использования 23 отобранных единиц пришлись на XIX и первую половину XX века; в русском языке, напротив, была отмечена тенденция на растущую популярность некоторых книжных фразеологизмов.

Список литературы

1. **Арсентьева Е. Ф.** Контекстуальное использование фразеологических единиц. Казань: Хэтер, 2009. 168 с.
2. **Звегинцев В. А.** Очерки по общему языкознанию. М.: МГУ, 1962. 382 с.
3. **Кунин А. В.** Англо-русский фразеологический словарь. Изд-е 4-е, перераб. и доп. М.: Рус. яз., 1984. 945 с.
4. **Лубенская С. И.** Русско-английский фразеологический словарь. 2-е изд-е. М.: АСТ-ПРЕСС КНИГА, 2004. 1080 с.
5. **Национальный корпус русского языка** [Электронный ресурс]. URL: <http://ruscorpora.ru/> (дата обращения: 17.11.2014).
6. **Солодухо Э. М.** Теория фразеологического сближения. Казань: Изд-во КГУ, 1989. 296 с.
7. **British National Corpus** [Электронный ресурс]. URL: <http://www.natcorp.ox.ac.uk/> (дата обращения: 18.11.2014).
8. **Google Ngram viewer** [Электронный ресурс]. URL: <https://books.google.com/ngrams> (дата обращения: 15.11.2014).

USING STATISTICAL PROGRAMMES WHILE INVESTIGATING PHRASEOLOGICAL UNITS OF A BOOKISH STYLE IN THE RUSSIAN AND ENGLISH LANGUAGES

Gur'yanov Igor' Olegovich
Kazan (Volga region) Federal University
igor.goor@gmail.com

The article is devoted to the specifics of applying computer programmes aimed to gather for further analysis the statistics of using linguistic units in the diachronic aspect. Phraseological units of a bookish style from bilingual dictionaries served as a material for research. The author analyzed the programmes Google Ngram Viewer, search engine of the Russian national corpus and British national corpus. The analysis identified the shortcomings of the modern statistical programmes.

Key words and phrases: phraseology; corpus; statistics; transformations; bookish style.