

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«ПЕРМСКИЙ ГОСУДАРСТВЕННЫЙ
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»
АССОЦИАЦИЯ «ИСТОРИЯ И КОМПЬЮТЕР»

ЦИФРОВАЯ ГУМАНИТАРИСТИКА: РЕСУРСЫ, МЕТОДЫ, ИССЛЕДОВАНИЯ

Материалы
Международной научной конференции
(г. Пермь, 16–18 мая 2017 г.)

Часть 1



Пермь 2017

УДК 004:3
ББК 32.97+6/8
Ц75

Цифровая гуманитаристика: ресурсы, методы, исследования: материалы Междунар. науч. конф. (г. Пермь, 16–18 мая 2017 г.): в 2 ч. / Перм. гос. нац. исслед. ун-т. – Пермь, 2017. – Ч. 1. – 175 с.
ISBN 978-5-7944-2908-4 (ч. 1)
ISBN 978-5-7944-2907-7

Сборник содержит материалы докладов, представленных на конференции, посвященной состоянию и основным проблемам развития Digital Humanities. В первой части рассматриваются социальные, институциональные, глобальные, языковые и мультикультурные аспекты цифровых гуманитарных наук, вопросы использования информационных технологий в области текстологии, текстометрии, лингвистики, компьютерного источниковедения и математических методов обработки информации.

УДК 004:3
ББК 32.97+6/8

*Издается по решению ученого совета
историко-политологического факультета и ученого совета
механико-математического факультета ПГНИУ*

Редакционный совет

д-р ист. наук С.И. Корниенко, д-р филол. наук В.А. Баранов,
д-р филол. наук К.И. Белоусов, канд. филос. наук Е.М. Березина,
канд. ист. наук Д.Б. Вершинина, канд. пед. наук Д.А. Гагарина,
д-р филол. наук Е.В. Ерофеева, д-р филол. наук С.Л. Мишланова,
канд. ист. наук Н.Г. Поврозник, А.Р. Харисова

Ответственный редактор

С.И. Корниенко

Издание осуществлено при поддержке гранта РФФИ, проект № 17-06-20071

ISBN 978-5-7944-2908-4 (ч. 1)
ISBN 978-5-7944-2907-7

© ПГНИУ, 2017

(the target audience). It is assumed, that different urban space fragments can be "tuned" to different types of consumers. As a result of the study a collection of urban locus language environment was formed. Besides, the most important features in gender correlation, socio-cultural and economic features, increased aspiration for integration with western economies, transitional condition of the Russian society self-identification at the regional and national-ethnic level, deficiency of Russian mythology potential to commercialization have been identified. These peculiarities permit to speak about the specific type of the customer for the localized urban topos: the predominance of the "western" type features over "regional" and "national-oriented" types. This makes it possible to discover natural specific features of modern urban onomasticon, to predict formation of its image and to regulate the process of economic and cultural development of modern society.

Key words: ergonym, ergonymikon, graphosemantic modeling, semantic graph, semantic map of the city, type of consumer.

УДК 801.6

ОБЪЯСНИМЫ ЛИ ОШИБКИ В.А. ЖУКОВСКОГО С ПОМОЩЬЮ АНАЛИЗА ДАННЫХ?¹

Орехов Б.В.

Национальный исследовательский университет «Высшая школа экономики», Россия, Москва, ул. Мясницкая, 20,
nevmenandr@gmail.com

В статье речь идёт о метрических отступлениях, которые мы наблюдаем в переводе «Одиссеи», выполненном В.А. Жуковским. Русский гекзаметр, конвенционально принятый для передачи античного метра, подразумевает обязательные 6 ударений в строке (соответствующих 6 долгим слога оригинала). Но в 49 случаях (на 12 с лишним строк поэмы) Жуковский ошибается и делает 5 или 7 ударений в строке. В статье изложены результаты поиска закономерностей между этими ошибками и другими параметрами текста: сюжетом, лексикой, некоторыми особенностями синтаксиса. Общий вывод неоднозначен. С одной стороны, мы действительно можем установить закономерности между песнями, в которых есть такие ошибки, и общим движением сюжета. Песни с ошибками приходится на ту часть поэмы, в которой Телемах находится в отъезде с Итаки. С другой стороны, более точной обусловленности, которая показывала бы, чем именно строка с ошибкой отличается от всех остальных строк, не обнаруживается.

Ключевые слова: поэтический перевод, В.А. Жуковский, анализ данных, метрические аномалии, русский гекзаметр.

© Орехов Б.В., 2017

¹ В данной научной работе использованы результаты проекта «Точки роста в языковой системе: корпусные исследования и моделирование», выполненного в рамках Программы фундаментальных исследований НИУ ВШЭ в 2017 году.

В представленном В.А. Жуковским варианте перевода поэмы Гомера есть 49 случаев особого рода, которые мы будем далее называть метрическими аномалиями. Речь идет в основном о тех же дактилических строках, длина которых, однако, отличается от гекзаметра на одну стопу как в сторону увеличения, так и в сторону уменьшения:

Многих он в помощь себе приведёт из песчаного Пилоса, многих (II, 326)

Как видно, в этих примерах Жуковский использует семистопный дактиль вместо привычного (и требуемого) шестистопного. Таких строк во всем переводе насчитывается 30.

В этой земле не знаком мне; скажи, где дорога (VI, 177)

Здесь строка оказывается на стопу короче, чем остальные, меньше, они исчерпываются 19 стихами.

Можно ли выявить внутритекстовые причины метрических аномалий, используя методы анализа данных? Даже не перечисляя конкретные методы, следует сказать, что это серьезный вызов. Анализ данных служит для обнаружения тенденций, характерных для больших массивов информации. В нашем случае ситуация обратная: мы пытаемся выявить механизм появления не тенденций, а исключений.

Метрические аномалии распределены по песням неравномерно. Больше всего их (по 6) в IX и XII песнях (обе переведены в 1844 г.). В песнях I, III, XVI, XIX–XXIV аномалий нет совсем. При этом песни с наибольшим числом аномалий представляют нам аномалии разного типа. В упоминавшихся IX и XII песнях число в 6 аномалий складывается из 3 семистопных и 3 пятистопных строк. В то же время V песнь, насчитывающая 5 аномалий, имеет совсем другое соотношение их типов: здесь 4 семистопные строки и только 1 пятистопная.

Как кажется, оперирование контекстом песни предполагает уместность подхода, при котором мы разделили все песни на две неравные группы. В первую попали те, в которых встречаются метрически аномальные строки, во второй оказались те, в которых таких строк не наблюдается. Таким образом, в первой группе будут II, IV–XV, XVII, XVIII песни; во второй – I, III, XVI, XIX–XXIV. Легко заметить, что первая группа захватывает, в основном, середину поэмы, а во второй оказываются песни «сюжетной рамки» начала и финала гомеровского текста.

Сначала попытаемся провести анализ с помощью традиционного герменевтического инструментария. Существуют ли видимые глазу отличия песен первой группы от песен второй группы на каком-либо из возможных уровней (прежде всего, конечно, сюжетном)?

В принципе, наиболее значимый сюжетный фактор, который делит весь текст «Одиссеи» на две части, – это состояние главного героя: половину повествования он находится на суше (Огигия, Итака), а

вторую половину – в морском путешествии (реальном или данном через нарратив). При этом рассказ о путешествиях Одиссея приходится именно на середину поэмы. Однако Одиссей покидает Огигию в середине V песни (метрические аномалии появляются уже во II и в IV) и прибывает на Итаку в XIII, в то время как метрические аномалии встречаются и в XIV, XV, XVII и XVIII песнях (хотя и единичны в двух последних). Таким образом, 6 песен из 15 нарушают предложенную модель распределения «аномальных» и «неаномальных» песен поэмы.

Существует несколько метрик, позволяющих оценить тематическую близость текстов. В этом исследовании мы будем использовать самую распространенную, построенную на основе вычисления TF-IDF. После того, как мы произвели все нужные подсчеты, мы приступили к собственно автоматической классификации, инструментом которой выступил самый распространенный алгоритм кластеризации, т. е. разбиения имеющегося множества объектов на группы, а именно алгоритм k-средних. Мы осуществили кластеризацию. В первый кластер попали песни, в которых мы ожидали бы увидеть метрические аномалии, во второй — песни, в которых аномалий в идеальном случае не было бы. Если посчитать меру согласия (каппа Коэна), то можно сделать следующий вывод: опыт кластеризации на основе распределения частотности слов в песнях (которое в конечном счете имеет в виду автоматическое определение тематики текста) показал, что кластеры хотя и не вполне строго, но в целом соответствуют делению на песни с метрическими аномалиями и без них.

Для анализа строк был выбран ряд текстовых характеристик, которые можно легко определить автоматически: 1) позиция в песне (влияет ли то, в начале, в середине или в конце песни находится аномальная строка; все песни разделены на 5 частей); 2) количество слов в строке; 3) набор ударных гласных в строке и т. д.

Лучший результат, который показал анализ атрибутов, нельзя назвать удовлетворительным, его мера качества оценивается алгоритмом в 0,05481. Эта цифра означает, что если потенциально и существует какой-то фактор, на 100 % объясняющий появление метрических аномалий, то выявленный набор атрибутов справляется с этой задачей только на 5,5 %. Фактически таким результатом можно пренебречь.

CAN WE EXPLAIN V.A. ZHUKOVSKY'S MISTAKES BY MEANS OF DATA ANALYSIS?

Orekhov B. V.

National Research University Higher School of Economics, 20,
Myasnitckaya st., Moscow, Russia, nevmenandr@gmail.com

In this article we are talking about metric deviations that we observe in the translation of the "Odyssey" made by V.A. Zhukovsky. Russian hexameter, the standard

adopted for transmission antique meter, implies 6 accents in the string (simply speaking, the respective 6 long syllables). But in 49 cases (more than 12 lines of a poem) Zhukovsky is wrong and makes 5 or 7 beats in a line. The article presents the results of a search of regularities between these errors and other settings of the text: plot, language, some peculiarities of the syntax. The overall conclusion is ambiguous. On the one hand, we can establish patterns between songs that have such errors, and the overall movement of the plot. Songs have mistakes on the part of the poem, in which Telemachus is departing from Ithaca. On the other hand, the more accurate determination, that would show what makes the line with an error different from all other lines, is not detected.

Key words: Poetic translations, Zhukovsky, data analysis, metric anomalies, Russian hexameter.

УДК 81

К ВОПРОСУ О ПРИНЦИПАХ ПОДАЧИ В ЭЛЕКТРОННОМ СЛОВАРЕ НЕОДНОСЛОВНЫХ ЕДИНИЦ, НАЗЫВАЮЩИХ МИФОЛОГИЧЕСКИХ ПЕРСОНАЖЕЙ¹

Русинова И.И., Гранова М.А.

Пермский государственный национальный исследовательский университет, Россия, Пермь, ул. Букирева, 15, irusinova@mail.ru

В работе обосновывается актуальность создания электронного «Словаря мифологических рассказов Пермского края». Наибольшую сложность при лексикографировании представляют однословные номинации духов-«хозяев» различных локусов (домового, дворового, банника, лешего и под.). Авторами приводится тематическая классификация таких единиц и ставится вопрос о принципах их подачи в указанном словаре. Анализ исследовательских работ и существующих словарей демонологической лексики позволяет сделать вывод об отсутствии четких принципов теоретического рассмотрения однословных номинаций мифологических персонажей и их словарного описания в современной науке. По нашему мнению, такие единицы должны включаться в словарь. Критериями для их отбора являются, во-первых, регулярная встречаемость во множестве текстов, а во-вторых, отражение в их семантике релевантных признаков персонажа.

Ключевые слова: мифологические рассказы Пермского края, демонологическая лексика, электронный словарь.

Фольклористами и этнолингвистами пермских вузов собран обширный корпус мифологических текстов. Оптимальной формой систематизации лексики, отражающей мифологические представления, может стать словарь.

© Русинова И.И., Гранова М.А., 2017

¹ Публикация подготовлена при финансовой поддержке РФФИ (грант № 17-14-59009а(р)) и гранта, выполняемого в рамках тематических планов по заданию Министерства образования РФ, № 34.1505.2017/ПЧ.