

ПРИМЕНЕНИЕ ДАННЫХ КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ ПРИ АНАЛИЗЕ СЕГМЕНТНОЙ ФОНОСТИЛИСТИКИ ФИННО-УГОРСКИХ САКРАЛЬНЫХ ТЕКСТОВ

Андуганова М. Ю.

ГОУ ВПО «Югорский государственный университет», Ханты-Мансийск, Россия (626011, Ханты-Мансийск, ул. Чехова 16/4 - 343), e-mail: Sem-Marianna@yandex.ru

Статья дает общую информацию о предмете компьютерной лингвистики в современной науке. В ней также определяется понимание сакральности текста разными исследователями и степень ее изученности в финно-угроведении. Основной целью является демонстрация способов применения данных компьютерной лингвистики для обработки фоностилистики текста на сегментном уровне, наглядно показано, что можно производить подсчеты элементов фоностилия, в частности: аллитерации, ассонанса, консонанса, рифмы, что позволит выявлять наиболее активные фонемы в том или ином тексте. Такие данные могут принести неоценимую помощь не только в исследованиях по лингвистической стилистике, но и психолингвистике, а особенно при анализе звуко-символизма сакральных текстов на финно-угорских языках.

Ключевые слова: компьютерная лингвистика, сакральный текст, фоностилика, сегментный уровень.

DATA APPLICATION OF COMPUTATIONAL LINGUISTICS IN ANALISING PHONOSTYLISTICS OF FINNO-UGRIC SACRED TEXTS ON THE SEGMENTAL LEVEL

Anduganova M. Y.

Ugra State university, Russia (628011, Khanty-Mansiysk, street Chekhova, 16/4 - 3430), e-mail: Sem-Marianna@yandex.ru

The article gives basic information about the subject of computational linguistics on modern science. It also defines interpretation of sacred texts by different researchers. The paper shows the degree of studies in the field of Finno-Ugric sacred texts by language experts. The basic aim is to demonstrate methods of application of phonostylistics on segmental level of texts processing by computational linguistics data. There are shown methods of computation of stylistic devices such as alliteration, assonance, consonance, rhyme. The results will be useful not only in the field of linguistic stylistics, but also in psycholinguistics and in analysis of sound and symbol of Finno-Ugric sacred texts.

Keywords: computational linguistics, sacred text, phonostylistics, segmental level.

Информатизация современного общества заметно повлияла на укрепление значимости компьютерной лингвистики, которая наряду с вычислительной и инженерной имеет объединенный и обобщенный термин – прикладной лингвистики. Эти три направления появились на стыке глубоко гуманитарной науки лингвистики (языковедения), изучающей законы развития и использования языком как могучим средством мышления и коммуникации, и компьютерных технологий, с помощью которой машине передается все большая часть технического и интеллектуального труда, выполняемого ранее человеком. Компьютерная лингвистика тесно связана с лингвистическим обеспечением информатики. Исследования в данной области опираются на основные единицы языковых уровней, таких как слово, словосочетание, а также на семантику текста (речи). Определить и описать их роль и значимость в современной машинной обработке текстов и гипертекстов не всегда представляется возможным. В узком смысле проблематика компьютерной лингвистики часто

связывается с междисциплинарным прикладным направлением для решения задач лингвистического обеспечения информатики.

Основной целью является демонстрация способов применения данных компьютерной лингвистики для обработки фоностилистики текста на сегментном уровне. При ее реализации были применены не только общенаучные методы и приемы, но и собственно лингвистические и математические (контекстологический, количественный, типологический, описательный и т.д.). Объектом нашего внимания стали сакральные тексты (СТ) в общем и религиозные сакральные тексты в частности. Это связано с тем, что на современном этапе лингвистических исследований сакральных текстов выявляется разный подход к трактовкам этого понятия. Так, одни исследователи под СТ понимают лишь тексты библейского происхождения [15, 13, 14, 8, 3, 10, 4, 6, 2] и посвящают им свои лингвистические исследования как при анализе словарного состава, так фразеологии и даже дискурса.

Другие как, например, Н. И. Коновалова сакральность понимают в более широком ее значении и считают, что СТ является отражением устной народной традиции в виде парологической, языческой, дохристианской картины мира [5: 8].

В финно-угристане проблеме сакральности мало уделено внимания в работах лингвистов, вернее данный вопрос практически не исследован в должной мере. Это справедливо объясняется тем, что искомое направление является совершенно новым как в отечественном и зарубежном языкознании [5:6], так и в финно-угроведении. К такой проблематике обратились лишь в 90-х гг. XX века [5:5]. Мы не ставим цели раскрывать детально историю вопроса, поскольку она может быть освещена отдельно. В рамках данной статьи весьма уместно могло бы быть упоминание фоностилистики как частного направления исследований стиля, которое подразделяется на два уровня – сегментного и суперсегментного. Ввиду того, что нами определено освещение вопроса «применения данных компьютерной лингвистики в сегментной фоностилистике финно-угорских СТ», то, следует заметить, что сегментный уровень является первым, т.е. элементарным в фоностиле [1:73]. Результаты компьютерной лингвистики на этом уровне могут применяться в части автоматической обработки языка, которую выделяет Е. П. Соснина. По ее мнению, при определении границ предметной области компьютерной лингвистики существуют два аспекта: 1) автоматическая обработка языка (Language Processing), включающая задачи анализа и моделирования языковой структуры; 2) прикладное направление.

К автоматической обработке языка относятся: графематический / фонематический анализ языка; морфологический анализ; лексико-грамматический анализ языка; синтаксический анализ или парсинг; анализ и моделирование семантической структуры;

задача синтеза языковых элементов, в т.ч. генерация текстов; автоматическая лингвостатистика.

Прикладным направлением является: машинный перевод; распознавание и синтез речи; разработка и использование искусственных языков; разработка и использование языков программирования; разработка и использование языков информационных систем; компьютерная лексикография и терминография; лингвистические основы информационного поиска; автоматическое индексирование, реферирование и классификация текстов; автоматический контент-анализ; авторизация текстов; гипертекстовые технологии представления текста; корпусная лингвистика; компьютерная лингводидактика [12, 16].

Поскольку сегментный фоностиль основан на экспрессии повторов фонем, их частоты, периодичности и т.д., то многие данные можно выявить благодаря незамысловатым операциям и использованию Word. Так, благодаря простой замене можно выявить количество слов в тексте, начинающихся с той или иной фонемы. Это важно для того, чтобы показать потенциально возможную частоту аллитерации и реальную ее частоту. Особенно при анализе текстов в финно-угриктике, поскольку аллитерация является одной из основополагающих стилистических средств устного народного творчества [11:131]. Заметим, что для проведения такой операции не требуется дорогостоящий компьютер или сверхновая программа и операционная система, например:

Используемый компьютер: 486 DX, операционная оболочка WINDOWS 3.11, текстовый редактор WORD 6.0.

Задача: Произвести подсчет графемы в набранном тексте на марийском языке (например: сколько звуков “а”, “ж” и т.д.), а также подсчет количества сочетаний конкретных графем (например: сколько сочетаний “аб”, “ак” и т.д.).

Нами был осуществлен анализ 65 текстов марийских языческих молитв, в котором применялся данный подход. Помимо обработки каждой молитвы был получен обобщенный результат графемной представленности в текстах молитв из разных источников.

Таблица 1

Частотность употребления графем в позиции начала слова

Графема	Кол-во	Графема	Кол-во	Графема	Кол-во	Графема	Кол-во
А	1062	Ж	24	О	779	ӱ	338
Ӓ	6	И	788	Ӗ	55	Х	3
Б	20	Й	982	П	3613	Ч	328
В	1588	К	5003	Р	223	Ш	2411
Г	1107	Л	614	С	1194	Ы	876
Д	858	М	1395	Т	2357	Э	248

1. Андуганова, М. Ю. Потенциал дискурса в исследовании фоностилистики устного народного творчества: Гуманитарные науки и образование / М. Ю. Андуганова. – Саранск, 2012. – С. 73–75.
2. Антропова, В. В. Фразеологизмы с теологическими, демонологическими компонентами и их дериватами в современном русском языке: Семантический и лингвокультурологический аспекты: автореф. дис. ... канд. филол. наук / В. В. Антропова. – Челябинск, 2004. – 23 с.
3. Бетехина, Е. Н. Фразеологизмы с библейскими именами: (в русском и английском языках). – СПб. : Изд-во Санкт-Петерб. гос. ун-та, 1999. – 171 с.
4. Гриненко, Г. В. Сакральные тексты и сакральная коммуникация. Логико-семантический анализ вербальной магии / Г. В. Гриненко. – М.: Новый век, 2000. 445 с.
5. Коновалова, Н. И. Сакральный текст как лингвокультурный феномен: дис. ... д-ра филол. наук: 10.02.01 / Н. И. Коновалова; Моск. гос. обл. ун-т. – Екатеринбург, 2007. – 354 с.
6. Мендельсон, В. А. Фразеологические единицы библейского происхождения в английском и русском языках: автореф. дис. ... канд. филол. наук / В. А. Мендельсон. – Казань, 2002. – 186 с.
7. Молданова, Т. А. Пелымский Торум – устроитель медвежьих игрищ. – Ханты-Мансийск: Полиграфист, 2010. – 224 с.
8. Никитина, Т. Г. Проблемы изучения этнокультурной специфики фразеологии / Т. Г. Никитина. – Псков, 1998. – 208 с.
9. Основы финно-угорского языкознания : вопросы происхождения и развития финно-угорских языков. – М.: Наука, 1974. – 484 с.
10. Родионова, И. В. Имена библейско-христианской традиции в русских народных говорах : автореф. дис. ... канд. филол. наук / И. В. Родионова. – Екатеринбург, 2000. – 24 с.
11. Семенова, М. Ю. Фонологические средства выразительности марийских языческих молитв и текстов карело-финских рун / М. Ю. Семенова.– Ханты-Мансийск, 2006. – 174 с.
12. Соснина, Е. П. О проблеме определения границ предметной области компьютерной лингвистики / Е. П. Соснина ; УлГТУ : Режим доступа : <http://ling.ulstu.ru/linguistics/chair/lecturers/sosnina/problemcl/> (дата обращения 9.04.2013).
13. Стижко, Л. В. Русское «черт» и чешское «čert»: значимость и коннотации // Язык и этнический менталитет / Л. В. Стижко [и др.]. – Петрозаводск, 1995. – С. 78–83.
14. Телия, В. Н. Русская фразеология. Семантический, прагматический и лингвокультурологический аспекты / В. Н. Телия. – М.: Языки и рус. культ., 1996. – 288 с.
15. Юдин, А. В. Христианский ономастикон в русских заговорах / А. В. Юдин // Русское языкознание: межвед. науч. сб. – Киев, 1992. – Вып. 25. – С. 130–137.

16. Ярославцева, Е. И. Компьютерная база данных «Языки мира» и ее возможные применения: дис. ... д-ра филол. наук : 10.02.21 / Е. И. Ярославцева. – М., 2005. – 305 с.

Рецензенты:

Долженко Н.Г., д.фил.н., проф. кафедры филологии Югорского государственного университета, г. Ханты-Мансийск;

Вафеев Р.А., д.фил.н., проф. кафедры филологии Югорского государственного университета, г. Ханты-Мансийск.