

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
им. М. Т. КАЛАШНИКОВА

Информационные технологии и письменное наследие

El'Manuscript-2012

Материалы IV международной научной конференции
Петрозаводск, 3–8 сентября 2012 года

Петрозаводск, Ижевск
2012

УДК 004.9
ББК 81.11+81.2-0
И741

Изданы при поддержке гранта РФФИ (проект № 12-06-06061-г),
гранта РГНФ (проект № 12-04-14154-г) и в рамках реализации комплекса
мероприятий Программы стратегического развития ПетрГУ на 2012-2016 г.

Ответственные редакторы:

В. А. Баранов, д-р филол. наук, проф.

А. Г. Варфоломеев, канд. физ.-мат. наук, доц.

Информационные технологии и письменное наследие [Текст] :
И741 материалы IV междунар. науч. конф. (Петрозаводск, 3–8 сентября
2012 г.) / отв. ред. В. А. Баранов, А. Г. Варфоломеев. — Петрозаводск ;
Ижевск, 2012. — 328 с.

ISBN 978-5-8021-1402-5

Сборник содержит материалы конференции, посвященной современ-
ным электронным средствам хранения, описания, обработки, исследования
и публикации памятников письменности и исторических источников.

УДК 004.9
ББК 81.11+81.2-0

ISBN 978-5-8021-1402-5

© Петрозаводский государственный
университет, 2012
© Ижевский государственный технический
университет им. М. Т. Калашникова, 2012

ЧАСТОТНЫЕ ИССЛЕДОВАНИЯ ТЕКСТОВ ИСТОЧНИКОВ ГРАЖДАНСКОЙ ПЕЧАТИ XVIII ВЕКА

А. Ю. Филиппович

МГТУ им. Н.Э.Баумана, Москва

Представленное исследование базируется на результатах проектов¹. В рамках исследования были рассмотрены различные источники гражданской печати XVIII века. Материалами послужили электронные факсимильные издания, представленные в свободном доступе в сети Интернет. Основные ресурсы:

- Электронная библиотека: старопечатные книги Российской государственной библиотеки, режим доступа: <http://elibrary.rsl.ru/>.
- Некоммерческая электронная библиотека «ImWerden», режим доступа: <http://imwerden.de/>
- Электронная библиотека, режим доступа: <http://www.infanata.org/>
- Электронная библиотека Google книги, режим доступа: <http://books.google.ru/>
- Электронная библиотека на Сайте Екатерины Кисловой, режим доступа: <http://ekislova.ru/>

Основные типы документов, определяющие лексический состав текстов: законодательные документы (указы, манифесты, законы и т. п.), юридические документы, исторические и географические описание городов и областей, исторические известия, художественная литература (путешествия, сочинения), описания родословий, уставы организаций, словари.

Были подобраны фрагменты источников объемом 25 страниц (ок. 25000 знаков), осуществлен символьный ввод их текстов. Всего было рассмотрено 52 источника, большая часть которых относится к последней четверти XVIII в.

В результате количественных исследований текстов были построены частотные словники фрагментов текстов, выявлены параметры функции распределения частот. Для этого использовались различные локальные ресурсы: Interlex; Andrew Tools; WinDialex, СУБД и on-line ресурсы:

- <http://advego.ru/text/seo/>
- <http://www.online-utility.org/text/analyzer.jsp>
- <http://www.samborsky.com/tools/textanalyse/>

Параметры и функциональные зависимости распределения в модели “ранг-частота” являются важнейшими атрибутами элементов частотного словника. Ранг — это порядковый номер слова в упорядоченном по убыванию частоты словнике; частота — формальный параметр слова (словофор-

¹ Проект РГНФ «Интегрированная инструментальная информационно-программная среда для автоматизации исследований Словаря Академии Российской 1789—1794 гг.» <http://www.it-claim.ru/Projects/ESAR/ESAR.htm>; Проект Гранта Президента РФ «Разработка словарных компонентов интегрированной информационной технологии переиздания печатных источников XVIII – нач. XIX вв.» <http://www.it-claim.ru/Projects/DicXVIII/DicXVIIImain.htm>

мы) в словнике, характеризующий частоту его (ее) появления в исходном тексте.

Полученные данные были использованы для сопоставительного исследования лексики рассматриваемых источников и Словаря Академии Российской 1789–1794 гг. (далее САР). Основой для исследования послужила созданная ранее база данных САР [Филиппович, 2008; Черкасова и Филиппович], включающая расширенный словник заголовочных слов [Филиппович, 2010а, 2010б]. Сопоставление лексики САР и источников XVIII в. осуществлялось в двух направлениях: 1) сопоставление словников фрагментов источников и расширенного словника САР; 2) сопоставление словников фрагментов источников и словника цитатного материала САР.

В результате сопоставительных исследований лексики было выявлено следующее: около 40% лексики источников рассматриваемого исторического периода не входит в состав современного русского языка в частности в словари-спеллеры текстовых редакторов. При этом для разных типов изданий 10–20% лексический состав совпадает. Созданный расширенный словник Словаря Академии Российской 1789–1794 гг. покрывает до 30–40% лексики.

Результаты проведенных исследований могут быть использованы для разработки систем распознавания исторических текстов, решения практических задач электронного и полиграфического издания древних памятников на базе квантитативных и лексических моделей текстов XVIII века.

Список литературы

Филиппович, 2008 — Филиппович А.Ю. Словарь Академии Российской (1789–1794): информационная технология переиздания. Вступительная статья М.И.Чернышевой. М., 2008.

Черкасова и Филиппович — Черкасова Г.А., Филиппович А.Ю. Информационная технология подготовки текста САР для создания его лингвистической базы данных. [Текст] // Альманах исследований Словаря Академии Российской. [Электронный ресурс]. Режим доступа: <http://www.philippovich.ru/Projects/ESAR/Almanah.htm>, свободный.

Филиппович, 2010а — Филиппович А.Ю. Информационная технология формирования лексического ядра языка печатных источников XVIII – нач. XIX вв. на основе Словаря Академии Российской 1789–1794 гг. // Горизонты прикладной лингвистики и лингвистических технологий. [Электронное издание: компакт-диск] / Доклады межд. науч. конф. Симферополь, 2010.

Филиппович, 2010б — Филиппович А.Ю. Разработка словарных компонентов интегрированной информационной технологии переиздания печатных источников XVIII – начала XIX века. // Печатные средства информации в современном обществе. Материалы науч. межвуз. конф. Сборник тезисов докладов. М.: 2010. С. 82–85.