

МИНИСТЕРСТВО НА ОБРАЗОВАНИЕТО И НАУКАТА НА РЕПУБЛИКА БЪЛГАРИЯ
КИРИЛО-МЕТОДИЕВСКИ НАУЧЕН ЦЕНТЪР ПРИ БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ
ИЖЕВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. М. Т. КАЛАШНИКОВА
НАУЧНОЕ СООБЩЕСТВО “ПИСЬМЕННОЕ НАСЛЕДИЕ”
DIGITAL MEDIEVALIST SCHOLARLY COMMUNITY
ФОНДАЦИЯ „УСТОЙЧИВО РАЗВИТИЕ НА БЪЛГАРИЯ“

**Писменото наследство
и информационните технологии**

El’Manuscript–2014

Материали от V международна научна конференция
Варна, 15–20 септември 2014 г.

София · Ижевск
2014

Сборникът е издаден с финансовата подкрепа на Министерството на образованието и науката на Република България по процедура за подкрепа на международни научни форуми.

Отговорни редактори: проф. дфн В. А. Баранов
 доц. д-р В. Желязкова
 д-р А. М. Лаврентъев

Редактори: Нели Ганчева, Веселка Желязкова (български текст)
 О. В. Зуга, В. А. Баранов (руски текст)
 Кевин Хокинс (Kevin Hawkins) (английски текст)

Писменото наследство и информационните технологии [Текст] : материали от V международна науч. конф. (Варна, 15–20 септември 2014 г.) / отг. ред. В. А. Баранов, В. Желязкова, А. М. Лаврентъев. — София ; Ижевск, 2014. — 448 с.

Сборникът съдържа материали от конференция, посветена на разработването и създаването на съвременни средства за съхраняване, описване, обработка, анализ и публикуване на ръкописни и старопечатни книжовни паметници и исторически извори, а също и на въпросите за подготвянето на електронни ресурси в областта на хуманитаристиката и тяхното използване в научните изследвания и преподаването.

© Кирило-Методиевски научен център — БАН, 2014
© Ижевский государственный технический университет
им. М. Т. Калашникова, 2014
© Авторски колектив, 2014
© Лилия Тошкова — графичен дизайн на корицата, 2014

ISBN 978–954–9787–25–2

Конкорданс цитатного материала Словаря русского языка XI–XVII вв.

Ю. Н. Филиппович

Инфо-когнитивные технологии, историческая лексикография, словарь русского языка XI–XVII вв., рукописная древнерусская картотека, электронные издания, автоматическая обработка текстов

The Concordance of the Dictionary of the Russian Language of the 11th–17th Centuries

Yuriy Philippovich

This paper presents a “pilot” issue of the serial digital edition of a concordance of issues 1–21 of the Dictionary of the Russian Language of the 11th–17th Centuries. The concordance is based on quotations from the printed issues of the dictionary. There are 2,817 processed entries containing 10,837 word forms. The concordance is designed to support the authors of new issues of the dictionary (the still incomplete letters of the alphabet) and also for researchers in the field of Russian historical lexicography and lexicology.

Словарь русского языка XI–XVII вв. (далее — Словарь или СЛРЯ) — уникальный лексикографический объект, насчитывающий в настоящее время 29 выпусков, содержащих около восьмидесяти тысяч слов. В русской исторической лексикографии ему нет равных. Особенностью этого старейшего издания является темп выхода из печати его выпусков, что является следствием дружной и слаженной работы авторского коллектива — ученых и сотрудников отдела “Исторической лексикографии” Института русского языка им. В. В. Виноградова РАН (<http://www.ruslang.ru>).

Успех коллектива во многом предопределен тем, что, во-первых, подготовка к изданию очередных томов представляет собой четко отлаженную технологию; во-вторых, в основе их труда лежит материал Рукописной древнерусской картотеки, насчитывающей почти два миллиона карточек — выписок из более 3000 источников многовековой истории русской письменности и первокнижности [Астахина и др. 1998].

Средствами информационной технологии подготовки к изданию большинства выпусков Словаря (с 1-го по 21-й) являлись традиционные инструменты ручной работы — “перо и бумага”, а также “печатная машинка”. Начиная с 23-го выпуска (1996 г.) текст Словаря стал готовиться с использованием компьютерной техники и передаваться в издательство на машинных носителях (22-й выпуск также был подготовлен с использованием компьютеров, но позже 23-го). Технологиче-

ский процесс подготовки словарных статей на буквы *А – С* был организован так, что на момент начала работ по подготовке к печати очередного тома определенная и зачастую значительная их часть уже была написана, т. е. авторская работа была почти завершена, словарные статьи уже были составлены. Далее предстоял значительный труд по их дополнению, сверке цитат и описаний источников, научное и техническое редактирование и др. работы. При завершении буквы *С* и подходе к букве *Т* стали возникать организационные трудности, связанные с ресурсными ограничениями (в основном финансовыми и кадровыми), стала очевидной тенденция к замедлению темпа издания томов Словаря, появился риск не завершить его. Это потребовало поиска резервов и новых подходов к организации работ над Словарем и, в особенности, к работе авторов [Филиппович Ю., Чернышева 1999; Филиппович Ю. 2002].

Рукописная древнерусская картотека — основа Словаря, однако ее использование при подготовке к изданию тома ограничено в основном стадией авторской работы. Чаще всего только в начале работы над словарной статьей автор работает с картотечным материалом. Более того, этот материал так организован, что не позволяет эффективно (с малыми затратами времени) повторно анализировать уже “отработанный материал”, как вошедший в другие словарные статьи, так и не использованный. В связи с этим отметим следующее: во-первых, по оценке самих ученых, только четвертая часть карточек выбирается авторами в качестве иллюстративного материала словарных статей; во-вторых, на карточках, как правило, выделено (подчеркнуто) только одно слово, которое явилось основанием к размещению карточки на определенном алфавитном месте; в третьих, карточки, текст которых вошел в Словарь, не помечаются. Вместе с тем часть карточек написана так, что в тексте выделены не одно, а два или более слов. Кроме этого, для обогащения материала некоторых букв предпринимались попытки переноса карточек на новое алфавитное место.

В дискуссиях о темпе издания очередных томов, сроке завершения Словаря и содержании новой технологии подготовки словарных статей высказывались различные идеи-предложения, отметим две из них.

Первая состоит в том, что необходимо издать недостающие тома в форме картотеки, т. е. издать современным способом, с использованием компьютерных технологий на непечатных носителях содержание карточек. Данная идея может быть реализована в рамках задачи создания электронного дубля Картотеки, как уникального лексикографического памятника [Богатова 1994]. Технологически это представлялось в виде CD-ROM издания, подготовленного в процессе сканирования карточек Картотеки и записи электронных файлов изображений на компакт-диски. Для реализации этой идеи необходимо было начать сканирование с ящиков на те буквы, для которых отсутствовали авторские словарные материалы (буквы *Т – Я*).

Вторая идея состоит в том, чтобы автоматизировать работу авторов словарных статей в части подбора материалов, воспользовавшись опубликованным ци-

татным материалом, иллюстрирующим другие заголовочные слова. Для этого необходимо ввести печатные выпуски Словаря в компьютер, декомпозировать введенные словарные статьи, выделить цитатный материал и организовать его удобным для использования способом. Такой формой организации цитатного материала является конкорданс.

Конкорданс СЛРЯ представляет собой способ организации уже опубликованного цитатного материала, который реализует вторую из названных идей — идею автоматизации работы авторов-составителей словарных статей. Он задуман как сериальное издание. Каждый выпуск посвящен одной букве — от У до Я.

В Интернете по адресу <http://www.it-claim.ru> размещено электронное издание пробного выпуска серии переработанных с использованием компьютерных технологий материалов СЛРЯ. Это “экспериментальный образец” будущего много-томного издания. Оно содержит материалы только одного “типового” тома Словаря — словарные статьи Володѣнье – Вящшина (Вып. 3. М.: Наука, 1976). В издании представлены:

1. Частотный словник словоформ (пример Щ и Ы; Ъ)

2 щели, щолоку

1 щавей, щавью, щадить, щадя, щадятъ, щастие, щебню, щедротамъ, щедрь, щедрьи, щедя, щеки, щепения, щепу, щепы, щетину, щетию, щетки, щеткою, щит#, щитовья, щитом, щитъ, щиты, щоку, щудка-волотка, щук, щука, щукошнои, щукъ, щупаетъ, щупал, щупами, щюри

...

4 ыной, ыные

3 ыных

2 ыном, ыныхъ, ысподи

1 ыгсмь, ызбъ, ызбе, ызбу, ызвъстной, ызвестном, ызмъне, ызмънными, ылмънях, ынои, ысбы, ысковъ, ыскъ, ыстерпъ, ысцовыхъ

2. Индекс словоформ на буквы Т – Я¹

ыной ВПЕРВЫЕ (ВПЕРВЫЙ, ВПЕРВЫЯ) (3, 81); ВРОВНО (3, 112); ВРОВНО (3, 112); ВЫСКАЗЫВАТИ (3, 247);

3. Частотный указатель источников²

330 Мин. ноябрь Месяц ноябрь // Ягич И. В. Служебные минеи за сентябрь, октябрь и ноябрь. В церковнославянском переводе по русским рукописям 1095–1097 гг. Изд. ОРЯС АН. СПб., 1886. С. 267–512. 1097 г., XII в.

275 Назиратель Книга глаголемая “Назиратель”, сиречь “Уряд домовных детель”. — Рукоп. ГИМ, Барс. N 371, вт. пол. XVI в. / Пер. с лат. Crescentius Opus rurelium commodorum. 1471, 1538, 1548 [в Словаре цитируется по изд.: Назиратель / Под ред. С. И. Коткова. М., 1973].

¹ В индексе указываются заголовочные слова словарных статей, номера страниц и соответствующих выпусков Словаря.

² В соответствии с [Филиппович А. 1999; Филиппович А. 2000; Филиппович Ю., Филиппович А. 2002] — <Абсолютная частота>, <Шифр источника>, <Полное наименование источника>.

4. Фрагмент конкорданса на буквы У – Я³

фелоня ВРАТА (3, 100)

Долженъ же архимандритъ носить мандии простомонашеския и во олтарь царскими враты безъ фелоня и епитрахили не входити.

АЮБ I, 150. 1697 г.

фетиль ВЫЕМНОЙ (3, 201)

Фетилного дѣла заводу: 2 котла мѣдныхъ, что фетиль ворять... 7

въкишъ деревяныхъ.

ДАИ X, 100. 1682 г.

Материалы получены после обработки 2817 словарных статей. Общее количество обработанных словоформ — 10837, в том числе по буквам алфавита: Т — 4883, У — 2230, Я — 1021, Х — 887, Ц — 819, Ч — 461, Ш — 317, Ф — 76, Ю — 71, Щ — 39, Ы — 33.

Литература

- Астахина и др. 1998 — Астахина Л. Ю., Черкасова Г. А., Романова Г. Я., Чернышева М. И. Подготовка к изданию книги “История древнерусской рукописной картотеки (КДРС). Авторский состав. Указатель источников”. Отчет о НИР № 98–06–80252 (Российский фонд фундаментальных исследований).
- Богатова 1994 — Богатова Г. А. Словарь русского языка XI–XVII в.: продолжение составления и публикации, экспериментальное использование новых информационных технологий для автоматизированного ведения и издания. Отчет о НИР № 94–06–19171 (Российский фонд фундаментальных исследований).
- Словарь 2001 — Словарь русского языка XI–XVII вв. Справочный выпуск / Астахина Л. Ю., Державина Е. И., Романова Г. Я., Филиппович Ю. Н., Филиппович А. Ю., Чернышева М. И. М.: Наука, 2001. 814 с.
- Филиппович А. 1999 — Филиппович А. Ю. Информационно-поисковая система “Указатель источников” // Интеллектуальные технологии и системы. Сб. статей. Вып. 2. М.: Изд-во МГУП, 1999. С. 230–240.
- Филиппович Ю., Чернышева 1999 — Филиппович Ю. Н., Чернышева М. И. Историко-лексикологическое (тематическое) исследование: экспериментальный опыт на основе информационной технологии // Вопросы языкознания. 1999. № 1. С. 56–83.
- Филиппович А. 2000 — Филиппович А. Ю. Прямой и обратный словники Словаря русского языка XI–XVII вв. Отечественная лексикография. Вып. 4. М.: ИРЯ РАН, 2000. С. 83–86.
- Филиппович Ю. 2002 — Филиппович Ю. Н. Информационная технология электронного издания рукописных и первопечатных памятников древнерусской письменности // Издательское дело и редактирование: теория, методика, практика: межвед. сб. науч. тр. Вып. 6. М.: Изд-во МГУП, 2002.
- Филиппович Ю., Филиппович А. 2002 — Филиппович Ю. Н., Филиппович А. Ю. Электронный Указатель источников рукописной древнерусской Картотеки и Словаря русского языка XI–XVII вв. М.: МГУП, 2002. 423 с.

³ Приведен не полностью — иллюстрируется только одно первое вхождение словоформ