

Рецензия на раздел «Диалектные и исторические корпуса» в сборнике «Труды международной конференции "Корпусная лингвистика-2019"»

Рецензируемый раздел сборника отражает корпусные разработки в области славянской и финноугорской диалектологии и диахронии. Семь статей, образующих раздел, посвящены как разным аспектам создания корпусов, так и исследованиям, проведенным на их основе.

В статье В.А. Баранова «Поиск и демонстрация данных в историческом корпусе «Манускрипт»» описывается возможность применения ряда статистических мер к многокомпонетным сочетаниям в рамках корпуса славянских памятников X-XV вв. «Манускрипт». В ходе создания модуля n-грамм и модуля статистики были решены также задачи построения подкорпусов, отвечающих заданным параметрам, и организации нечеткого поиска, позволяющего работать с текстами с неустоявшейся орфографией.

Статья И.В. Азаровой, Е.Л. Алексеевой, А.М. Лаврентьева, Е.А. Рогозиной, К.В. Сипунина «Представление и анализ библейских, святоотеческих и литургических цитат в корпусе СКАТ» представляет собой пилотную разработку в области разметки интертекстуальных связей в историческом корпусе, включающем древнерусские жития XV-XVIII вв. Размещение корпуса на платформе ТХМ (Textometrie) делает возможным применение к размеченным житиям комплекса текстометрических процедур, разработанных для этой платформы.

Авторы статьи «К проблеме создания размеченных корпусов текстов в графике XIX века» А. А. Лебедев, А. А. Рогов, К. А. Кулаков и Н. Д. Москин работают с русскими публицистическими текстами второй половины XIX в. в графике того времени (корпус СМАЛТ). Они предлагают решение таких важных задач, как обработка текстов с вариативной орфографией и представление в корпусе составных словоформ.

С.О. Савчук в статье «Инструментарий Национального корпуса русского языка в диахронических исследованиях» рассматривает, как изменился характер координации подлежащего и сказуемого в русском языке на протяжении последних двух столетий, используя данные и инструментарий НКРЯ.

В статье С.С. Земичевой «Новые темы диалектного дискурса (на материале Томского диалектного корпуса)» изучается влияние исторических изменений XX в. на материальные аспекты жизни крестьян и на их мироощущение. Исследование стало возможным благодаря тематической разметке текстов корпуса.

Статья А.А. Крижановского, Н.Б. Крижановской и И.П. Новак «Представление диалектов в Открытом корпусе вепского и карельского языков (ВЕПКАР)» посвящена проблемам создания диалектного корпуса и разработки его инструментария. Представляется очень интересным предложенный авторами способ отождествления семантически и фонетически близких лексем в разных диалектах. Работа над корпусом позволила также уточнить разделение карельского и вепского языков на наречия и диалекты.

И.С. Николаев в статье «Корпусное исследование топонимов в ижорских народных песнях» использует данные корпуса «Древние песни финского народа». Варьирование топонимов в

составе распространенной песни на сюжет «Большой дуб» позволяет сделать выводы о их значимости для носителей ижорского языка.

В целом все статьи рецензируемого раздела отражают широкий круг тенденций в разработке и использовании исторических и диалектных корпусов, соответствуют тематике конференции, написаны на высоком научном уровне и представляют несомненный интерес как для создателей корпусов, так и для исследователей, использующих их в своей работе.

Доцент кафедры математической лингвистики СПбГУ, к.ф.н.

Е.Л. Алексеева

Личную подпись заверяю
Документ подготовлен по личной
инициативе

Алексеева Е. Л.

11 ИЮН 2019

ТЕКСТ ДОКУМЕНТА РАЗМЕЩЕН В ОТКРЫТОМ
ДОСТУПЕ НА САЙТЕ СПбГУ ПО АДРЕСУ
[HTTP://SPBU.RU/SCIENCE/EXPERT.HTML](http://spbu.ru/science/expert.html)

Копию считаем не действительной
Семенов В. В. Семенов

