

Корпусное представление древнерусских памятников

Возможности корпусного представления древних славянских рукописей открыли новую эпоху в их научном издании и исследовании. В России эта деятельность осуществляется в рамках проекта «Манускрипт» и программы «Национальный корпус русского языка».

Национальный корпус русского языка (ruscorpora.ru) содержит четыре подкорпуса, представляющих древнерусские письменные источники: собственно древнерусский, корпус берестяных грамот, старорусский корпус и церковнославянский. Все представленные в них памятники имеют метатекстовую разметку, что позволяет вести поиск по авторам, жанрам, дате создания и др.

Древнерусский корпус включает оригинальные древнерусские произведения, выполненные на Руси переводы с греческого и памятники южнославянского происхождения, переписанные на Руси. Вошедшие в этот корпус тексты имеют грамматическую разметку. Это позволяет осуществлять детальный поиск по грамматическим параметрам, причем все словоформы имеют точное указание на их место в издании или рукописи, по которой выполнена разметка. У переводных памятников приводится греческое соответствие данной словоформы. Аналогично устроен корпус *берестяных грамот*.

Основу *церковнославянского корпуса* составляют современные богослужебные тексты (XIX-XX век) – около 60% объема корпуса, однако в корпусе представлены и более ранние тексты, относящиеся к XVII-XVIII вв.

Самым крупным является *корпус старорусских текстов*, который фактически охватывает весь древнерусский период – с XI-го по XVII в. В настоящее время корпус включает 4,5 тысячи документов общим объемом около 6 млн словоупотреблений. На нынешнем этапе работы в него включаются все существующие публикации древнерусских текстов разных жанров (летописи, памятники бытовой и деловой письменности, религиозная литература и др.). Поиск в этом корпусе возможен только по словам или частям слов. В дальнейшем этот корпус объединит все письменные источники XI–XVII вв.

Corpus Representation of Old Russian Services

Abilities of corpus representation of ancient Slavonic manuscripts have opened a new era in their editing and studying. In Russia this activity is carried within the framework of the project “Manuscript” and of “Russian National Corpus” program.

Russian National Corpus (ruscorpora.ru) includes four subcorpuses, representing Old Russian written heritage: Old Russian, birch bark manuscripts, Middle Russian and Church Slavonic. All represented services have metatextual marking, which allows you to carry out search by author, genre, date etc.

Old Russian corpus contains original Old Russian writings, written in Rus’ translations from Greek and South Slavonic services rewritten in Rus’. Texts included in this corpus have grammatical marking. This allows to carry out detailed search by grammatical parameters, moreover, all word forms have exact addresses from its publications or manuscripts which was marked up. In case of translated services corresponding Greek word form is provided. The birch bark manuscripts corpus is arranged in a similar way.

The main part of Church Slavonic Corpus consists of modern liturgical texts (XIX – XX centuries), which occupy approximately 60% of the corpus volume. However texts pertaining to the XVII-XVIII centuries are also present in the corpus.

The largest is the Middle Russian Text Corpus, which de facto covers the whole period from XI to XVII centuries. The corpus currently includes 4.5 thousand documents with overall volume of approximately 6 millions word usages. On the present stage of work all of existing publications of Old Russian texts of different genres are being included. Only search by keywords or by parts of keywords can be carried out in this corpus. Subsequently this corpus will contain all written sources dated XI-XVII centuries.