

## К ВОПРОСУ О СОЗДАНИИ КОРПУСА ТЕКСТОВ ПСКОВСКОЙ ДЕЛОВОЙ ПИСЬМЕННОСТИ<sup>1</sup>

*В статье рассматривается возможность создания лингвистического корпуса на основе хозяйственных книг Псково-Печерского монастыря XVII-XVIII века. Также обоснована актуальность работы над созданием виртуальной библиотеки и определен план работы по подготовке материалов для электронной библиотеки исторических текстов, что позволит обеспечить доступность псковских памятников деловой письменности для широкого круга исследователей.*

**Ключевые слова:** корпус, корпусная лингвистика, электронная библиотека, исторический текст.

J.V. Kovalykh

## ON THE ISSUE OF CREATING THE LINGUISTIC CORPUS OF THE PSKOV BUSINESS WRITTEN LANGUAGE

*The article considers the possibility of creating the linguistic corpus of the Pskov business written language on the basis of the XVII-XVIII centuries' commercial books of Pskovo-Pechersky Monastery. The author proves the urgency of establishing the virtual library and outlines the actions aimed at preparing materials for the electronic library of historical texts, which will ensure the availability of the Pskov artifacts of the business written language for a wide range of researchers.*

**Key words:** corpus, Corpus Linguistics, electronic library, historical text.

Для начала следует выяснить, что такое корпус, корпус текстов.

Разработкой теории корпуса, созданием и использованием текстовых корпусов занимается **корпусная лингвистика** — лингвистическое направление, появившееся в 60-е годы XX века. В 80-е годы повышению интереса к этому разделу языкознания способствовало стремительное развитие информационных технологий.

Исследователи дают разные определения корпуса (от лат. corpus – “body”), но все они сходятся в одном: «Корпус текстов — это определенным образом организованное множество, элементами которого являются тексты. Организация корпуса может быть самая разная, в зависимости от прагматических целей его создателя или пользователя. Тексты, которые суть составляющие элементы корпуса, могут представлять собой целое оригинальное словесное произведение или какую-либо его часть. Как правило, предполагается, что весь корпус записан на магнитном (машинном) носителе и расположен компактно в одном месте» [6, с. 89]. В.В. Рыков определяет лингвистический корпус как некоторое собрание текстов, в основе которого лежит логический замысел, логическая идея, объединяющая эти элементы. Логическая идея воплощается в правилах организации текстов в корпус; алгоритме и программе анализа корпуса текстов; сопряжённой с этим идеологии и методологии [6].

Аннотирование корпусов, или разметка, — это описание каждой единицы текста с помощью специальной системы отметок, которая осуществляется программными средствами.

<sup>1</sup> Исследование осуществляется в рамках в реализации федеральной целевой программы «Научные и научно-педагогические кадры инновационной России» на 2009-2013 годы по теме «Язык и культура в коммуникативном пространстве региона как отражение национального единства в его многообразии». (Государственный контракт от «15» сентября 2010 г. № 14.740.11.0228. Шифр «2010-1.1-304-125»).

Лингвистическая разметка подразумевает присвоение словам особых кодов (*тэгов*, от англ. tag – ‘ярлык, метка’). Каждому коду соответствует определенный набор грамматических признаков, характеризующих данное слово. Однако полностью автоматизировать такую работу не представляется возможным, так как нередко возникают спорные вопросы, связанные с морфологической омонимией и т.п., когда исследователь должен сам принять решение.

Таким образом, лингвистический корпус представляет собой некоторую совокупность текстов, объединенных по какому-либо принципу (в соответствии с замыслом исследователя), оформленных определенным образом и сопровождаемых системой помет в соответствии с общепринятыми требованиями. Корпус должен быть обеспечен специализированной поисковой системой.

В настоящее время многие российские университеты уже включились в работу по созданию различных лингвистических корпусов, в частности в создание Национального корпуса языка. Участие в этой работе принимают как преподаватели, так и студенты. Например, В.И. Подлеская в статье «Современные компьютерные методы в изучении и преподавании лингвистических дисциплин: корпусная лингвистика» отмечает: «Поскольку современный профессионально подготовленный лингвист обязан уметь эффективно пользоваться корпусами, а при необходимости, уметь самостоятельно сформировать корпус с нужными параметрами, в 2003/2004 учебном году для специальности «Теоретическая и прикладная лингвистика» введен обязательный курс «Корпусная лингвистика» [5].

Создание корпуса сделало бы работу с текстами, например, письменных памятников более эффективной. Хозяйственные книги Псково-Печерского монастыря могли бы составить отдельный лингвистический корпус. Эти памятники объединены следующими признаками:

- 1) языком;
- 2) они в основном отражают хозяйственную сторону жизни монастыря;
- 3) принадлежат к ограниченному временному пространству (XVII - XVIII век);
- 4) являются иллюстрацией делового языка носителей определенной территории (ныне это территории Печорского, Псковского, Гдовского, Островского, Пыталовского, отчасти Пушкиногорского районов).

Как и в других случаях, потенциальный корпус мог бы быть использован для выявления частотности употребления того или иного слова и его словоформ и не только. Если все единицы (слова и словоформы) сопроводить грамматическим комментарием, это позволит в значительной мере облегчить работу по сбору материала для исследования грамматического строя языка псковских хозяйственных книг и т.д. Кроме того, такой корпус может стать богатым источником данных для лексикографии и лексикологии.

Таким образом, перед создателем корпуса встанут следующие проблемы:

- 1) транслитерация скорописных текстов в соответствии с требованиями публикации памятников письменности, выработанными в лингвистическом источниковедении и сформулированными С.И. Котковым [2, 3];
- 2) выбор стандарта аннотирования текстов и осуществление лингвистической разметки;
- 3) компьютерная обработка текстов, применение корпусного менеджера, что предполагает хорошую подготовку в сфере компьютерных технологий или возможность привлечения к работе специалистов в области информационных технологий.

Из представленных вопросов только первый может быть решен на данном этапе. Два других пока представляют собой далекую перспективу для научной работы, к которой можно было бы подключить и студентов.

Создание электронной библиотеки исторических текстов сделает псковские памятники письменности доступными для широкого круга исследователей. Ю.Н. Филиппович в 2006 г. справедливо отмечал: «Несмотря на то, что электронные издания являются только нарождающейся формой распространения информации, они представляются перспективными, в том числе и для решения проблем сохранения древних письменных источников» [1, с. 333].

Работа над созданием электронной библиотеки производится при сотрудничестве кафедры русского языка и проблемной научно-исследовательской лаборатории региональных филологических исследований ПГПУ.

В настоящее время в нашем распоряжении имеются тексты, опубликованные кандидатом филологических наук, доцентом Н.Д. Сидоренской:

- переписная книга Псково-Печерского монастыря 1652 г. (опубликована также в журнале «Псков») [4],
- нетные книги [7].

Кроме того, производится транслитерация текста приходо-расходной книги 1674-1675 гг. В дальнейшем предполагается публикация и других памятников псковской деловой письменности.

При подготовке текстов соблюдается четкое следование требованиям публикации памятников письменности. Учитывается опыт лингвистов, работающих с древнерусскими источниками, издающих рукописные памятники [1].

Создание виртуальной библиотеки сделает доступными тексты псковских хозяйственных книг для научно-исследовательской работы не только преподавателей и студентов ПГПУ, но и для самого широкого круга специалистов — историков и филологов.

## Литература

1. Астахина Л.Ю. Слово и его источники. Русская историческая лексикология: источниковедческий аспект. С послесловием Ю.Н. Филипповича. М., 2006.
2. Князевская О.А., Котков С.И. Правила лингвистического издания памятников древнерусской письменности. М., 1961.
3. Котков С.И. Источниковедческие исследования и научное издание памятников в области русского языка // Источники по истории русского языка XI-XVII вв. М., 1991.
4. Переписная книга Псково-Печерского монастыря 1652 г. /Подгот. текста к публик. Н.Д. Сидоренской // Псков. 1996. № 4. С. 128-132;  
Переписная книга Псково-Печерского монастыря 1652 г. /Подгот. текста к публик. Н.Д. Сидоренской // Псков. 1996. № 5. С. 105-108;  
Переписная книга Псково-Печерского монастыря 1652 г. /Подгот. текста к публик. Н.Д. Сидоренской // Псков. 1997. № 6. С. 122-128;  
Переписная книга Псково-Печерского монастыря 1652 г. /Подгот. текста к публик. Н.Д. Сидоренской // Псков. 1998. № 8. С. 156-161;  
Переписная книга Псково-Печерского монастыря 1652 г. /Подгот. текста к публик. Н.Д. Сидоренской // Псков. 1999. № 10. С. 191-196;  
Переписная книга Псково-Печерского монастыря 1652 г. /Подгот. текста к публик. Н.Д. Сидоренской // Псков. 1999. № 11. С. 187-196;  
Переписная книга Псково-Печерского монастыря 1652 г. /Подгот. текста к публик. Н.Д. Сидоренской // Псков. 2001. № 14. С. 159-165;  
Переписная книга Псково-Печерского монастыря 1652 г. /Подгот. текста к публик. Н.Д. Сидоренской и Е.В. Ковалых // Псков. 2001. № 15. С. 185-191;  
Переписная книга Псково-Печерского монастыря 1652 г. /Подгот. текста к публик. Н.Д. Сидоренской и Е.В. Ковалых // Псков. 2002. № 16. С. 211-217;  
Переписная книга Псково-Печерского монастыря 1652 г. /Подгот. текста к публик. Е.В. Ковалых и Н.Д. Сидоренской // Псков. 2002. № 17. С. 208-213.
5. Подлесская В.И. Современные компьютерные методы в изучении и преподавании лингвистических дисциплин: корпусная лингвистика // <http://www.rsuh.ru/article.html?id=398>.
6. Рыков В.В. Прагматически ориентированный корпус текстов // Тверской лингвистический меридиан. Теоретический сборник / Ред. Сусов И.П. Тверь, 1999. Вып. 3. С. 89-96.
7. Сидоренская Н.Д. Нетные книги Псково-Печерского монастыря 1682 г.: Монография / Публикация текста и исследование Н.Д. Сидоренской; составление Указателей слов и личных имен Е.В. Ковалых; отв. редактор Н.В. Большакова. Псков, 2010.