

КОРПУСНЫЕ МЕТОДЫ В ИССЛЕДОВАТЕЛЬСКОЙ ДЕЯТЕЛЬНОСТИ СТУДЕНТОВ-ФИЛОЛОГОВ

И.Г. Гомонова
(Гомель, Беларусь)

В статье рассматривается применение корпуса текстов в исследовательской деятельности студентов-филологов. Корпусные методы характеризуются как эффективный способ решения лексикографических, вариантных, словообразовательных и фразеологических задач.

Прошло не так много времени с момента появления корпуса (он был создан в 1963 г.), а лингвистические первого в мире корпуса текстов (Брауновский корпус) существенно изменили исследовательскую и преподавательскую практику лингвистов. С помощью лингвистических корпусов решаются многочисленные исследовательские и дидактические задачи [см. об этом: 1].

Изучая дисциплину «Корпусная лингвистика», студенты-филологи знакомятся с принципами работы корпуса текстов и вырабатывают у себя навыки использования корпусных методов в исследовании и преподавании языка. С этой целью студентам предлагается как выполнить готовые задания, так и составить свои задания, разрабатывая которые можно анализировать разнообразие языковых фактов и обучать языку.

В первую очередь на базе корпусов решают лексико-графические задачи: получение дополнительного иллюстративного материала; обнаружение семантических и стилистических изменений, происходящих в слове, и др. Так, осуществляя на базе Национального корпуса русского языка (далее – НКРЯ) [2] верификацию толкования слова *вакация*, студенты приходят к выводу, что в значении данного слова произошли некоторые семантические сдвиги. В текстах XVIII в. – первой половины XX в. начиная с М.В. Ломоносова (первое упоминание данного слова в текстах НКРЯ) существительное *вакация* употребляется преимущественно в значении „каникулы, перерыв в учебном процессе“ : *Вакация или успокоение от трудов академических по примеру прочих академий имеет продолжиться с 1 мая по пятнадцать июня, в которое академики от своих трудов увольются и имеют свободу на сие время отъехать куда пожелают в пределах Российского государства* (М.В. Ломоносов. Проект регламента Академии Наук (1762)); *Школьники, возвращающиеся домой на вакацию, не с большею радостью приветствуют колокольню родного села* (И.И. Лажечников. Басурман (1838)); *Молодежи в поезде вообще много: едут домой на летние вакации из университетов, из институтов* (С.Д. Мстиславский. Крыша мира (1905)). Наибольшая активность употребления данного слова, по данным НКРЯ, приходится на конец XIX – начало XX в.в. С 1910 г. встречаемость существительного *вакация* резко снижается. В современных лексикографических источниках это слово характеризуется как устаревшее. Однако анализ употребления данного существительного в текстах XXI в. дает основание отнести его к «возвращенной лексике» и констатировать некоторую активизацию значения „перерыв в общественно-политической деятельности, отпуск политического деятеля“ : *В случае же с отпусками Путина таких вопросов не возникло с самого начала. Однако нынешняя вакация ВВП позволяет сделать вывод о том, что в ближайшее время требования к интенсивности и качеству работы госаппарата будут ужесточены* (С. Юрьев. Что делал Путин в Сочи? Он работал // Комсомольская правда, 2002.09.24). Решение компенсаторных лексикографических задач (получение дополнительного иллюстративного материала) актуально прежде всего для редких слов. Анализируя контексты со словом *цата* („золотая или серебряная полукруглая подвеска, прикрепляемая к окладу иконы возле шеи изображенной фигуры“ [3]), студенты формируют более полное представление о значении и условиях употребления данного существительного и делают вывод о том, что в настоящее время оно используется в основном как специальное, характерное для текстов по искусствоведению: *Цата – полукруглая или сердцеобразная металлическая подвеска у икон под ликом, прикрепленная к краям венца* (П.И. Мельников-Печерский. В лесах. Книга первая (1871–1874)); *То есть сначала украшались лишь поля иконы, затем стали закрываться нимбы и свет, затем – появились венцы, цаты (полукруглые украшения по углам венцов), гривны – уже закрывалась живопись: наконец, оказалось закрыто всё, кроме лика* (С.А. Еремеева. Лекции по русскому искусству (2000)).

Самый многочисленный тип задач, решаемых на базе корпусов текстов, – исследование грамматической вариантности. Корпус при этом может выступать и как источник контекстов, иллюстрирующих варианты форм, в том числе ошибочные (такие примеры можно использовать в качестве учебного материала на занятиях, посвященных изучению норм литературного языка), и как база для исследования, направленного на выяснение соотношения вариантов.

Как известно, распространенной морфологической ошибкой является смешение разных способов образования форм степеней сравнения имен прилагательных и наречий. Оформив в корпусе текстов соответствующий запрос (*более* + простая форма сравнительной степени; *самый* + простая форма превосходной степени), студенты получают контексты с ошибочными формами, делают вывод о типичности таких ошибок и вырабатывают навыки их исправления: *Ну, ведь известно, что чем более фантастичней выдумка, чем она менее имеет под собой известных оснований, тем она ближе к истине* (А. Иличевский. Бутылка (2005) // «Зарубежные записки», 2008); *Но самое наиглавнейшее – печать дара на нем была* (Б. Евсеев. Евстигней // «Октябрь», 2010).

Полезными для студентов в теоретическом и практическом отношении являются задания, связанные с определением соотношения вариантов. Так, выявляя соотношение вариантов форм повелительного наклонения глаголов *ехать* / *поехать*, студенты наблюдают значительное преобладание нормативного варианта над ненормативными: *поезжай* –

1329 вхождений, *езжай* – 464 вхождения, *ехай* – 51 вхождение, *едь* – 10 вхождений (данные по Основному корпусу НКРЯ). В то же время отмечается факт преобладания варианта *езжай* в текстах Газетного корпуса (182 вхождения против 97 вхождений варианта *поезжай*), что может свидетельствовать об отсутствии строгого контроля за соблюдением норм литературного языка в текстах современной публицистики. Анализируя данные варианты, студенты отмечают также тот факт, что только вариант *поезжай* представлен в НКРЯ текстами XVIII века, варианты *езжай* и *едь* отражены в корпусе начиная с текстов XIX века, *ехай* – начиная с текстов XX века. При этом последний вариант, представляющий собой грубое нарушение морфологической нормы, используется, как правило, в характерологической функции – отражает разговорно-просторечный характер речи персонажа или его иностранное происхождение: *Цельный день собирались. Теперь вот ночью ехай* (Н. Дубов. Мальчик у моря (1966)); *«Ехай, ехай, – торопил он, – ищи свой баба, кенязь»* (Б. Окуджава. Путешествие дилетантов (Из записок отставного поручика Амирама Амилахвари) (1971–1977)).

С опорой на корпус текстов можно сделать интересные наблюдения в отношении практически всех вариантов, представленных в морфологической системе русского языка. На основе лингвистического корпуса целесообразно также анализировать синтаксическую вариантность. Объектом такого исследования могут выступать варианты организации словосочетаний со связью управления. Выполняя подобные задания, студенты, в частности, анализируют выбор падежных форм при употреблении производных предлогов. Например, для предлога *сообразно* нормой является сочетание с формами дательного и творительного падежа (вариант *сообразно с*). По данным НКРЯ, более востребован первый вариант (981 вхождение против 515 вхождений в текстах Основного и Газетного корпусов), преобладание которого особенно характерно для текстов конца XX – начала XXI в.в. Представлены также в текстах корпуса ненормативные варианты употребления предлога *сообразно* с формой родительного падежа. Такие контексты подлежат анализу с целью предупреждения подобных ошибок в речи: *Переменный воздушный конденсатор для настройки имеет максимальную емкость 250-300 см, и включается **сообразно** схемы* (Н. Кузнецов. Детекторный приемник с выключением мертвых витков // «Радио Всем», 1927).

С использованием корпусных методов можно анализировать и деривационную вариантность, обусловленную наличием в русском языке синонимических морфов. Результатом такого исследования могут быть выводы о предпочтительности того или иного варианта, об особенностях сочетаемости вариантов и т. д. Исследуя, например, адъективные дериваты *апельсиновый* и *апельсинный*, которые в словарях подаются без всякой дифференциации, студенты делают вывод о значительном преобладании первого адъектива в текстах НКРЯ (в Основном корпусе соотношение 865 и 204; в Газетном – 687 и 2; даже с поправкой на неснятую омонимию, вследствие которой в конкордансе с прилагательным *апельсиновый* представлены формы существительного *апельсинов*, разница существенная). Перевес в пользу адъектива *апельсинный* наблюдается только в Поэтическом корпусе (32 и 48), что, возможно, объясняется его лучшей ритмизацией: *Скользит гондола длинная По бархатной гряде, А корка **апельсинная** Качается в воде...* (Вс.А. Рождественский. Венеция (1923- 1926)). Что касается сочетаемости прилагательных *апельсиновый* и *апельсинный*, то, по выражению М. Королевой, они, «разобрали себе существительные, спокойно распределили их между собой и выступают с ними исключительно попарно» [4]. Действительно, предпочтения в сочетаемости данных адъективов несомненны: *апельсиновый* – *сок, напиток, нектар, джем, масло, ликер* и др. („приготовленный из апельсина); *дерево, роцца, плантация* („относящийся к апельсинам; состоящий из них); *свет, блузка, ранец, утюг, машина* и др. („цвета апельсина); *апельсинный* – *корка, долька, кожура, цедра* („относящийся к плоду апельсина). В то же время нельзя согласиться с заявлением об исключительности такой сочетаемости. Анализ контекстов употребления данных адъективов убедительно демонстрирует «обмен» существительными: *апельсиновый* – *корка, кожа, долька* и др.; *апельсинный* – *сок, сироп, ликер, варенье, крем, духи* и др.; *дерево, роцца, сад* и др.; *цвет, пушок*.

Подобные тенденции (и в количественном отношении, и в плане сочетаемости) характерны и для дериватов *мандариновый* и *мандаринный*, *ананасовый* и *ананасный*. Пара адъективов *абрикосовый* и *абрикосный* представлена в корпусе преимущественно первой единицей, второе прилагательное входит только 3 раза в тексты Поэтического корпуса: *Апельсинный, **абрикосный**, Лейся, сок души роскошный...* (М.И. Цветаева. Встреча вторая [Царь-девица, 5] (1920)); *В стране неведомой увлечь Могла « меня ты И блеском **абрикосным** плеч, И губ медвяным ароматом* (Н.С. Тихонов. Паладин (1914–1921)).

Исследованием деривационной вариантности не ограничиваются словообразовательные задачи, которые решаются с использованием корпусных методов. С помощью корпуса текстов можно выявлять продуктивность словообразовательных моделей, уточнять состав производных слов, изучать их сочетаемость, анализировать окказиональные слова и др.

В качестве источника контекстов употребления, в том числе включающих трансформированные варианты, корпус текстов выступает в исследованиях, объектом которых являются фразеологические и паремиологические единицы. Приведем примеры использования корпуса в студенческих исследованиях такого рода. Выполняя работу на тему «Сложные предложения- паремии с сопоставительными отношениями в русском языке», исследователь делает в корпусе запрос на пословицу *Чем дальше в лес, тем больше дров*. Чтобы получить как можно больше контекстов ее употребления, запрос оформляется отдельно на первую и вторую части (можно также опустить союз). В разных корпусах НКРЯ фиксируется более 70 вхождений анализируемой пословицы, представляющих собой богатый материал для исследования ее функций, условий встраивания в текст, типов трансформации и т.д. Полученные контексты демонстрируют, что в большей их части пословица *Чем дальше в лес, тем больше дров* подвергается преобразованиям разного типа. Наиболее типичное из них – замена второй части пословицы: *Но чем дальше в лес, тем больше гнилушек* (В. Кичин. Целеустремленное кино // «Известия», 2002.07.26); *Чем дальше в лес – дорожка назад труднее* (Г.Н. Владимов. Генерал

и его армия (1994)) и др. Реже встречается контаминация двух пословиц: *Я врал убедительно, красочно, аранжируя свою болтовню цифрами, подробностями и пословицами типа «чем дальше в лес, тем своя рубаха ближе к телу»* (А. Рубанов. Сажайте, и вырастет (2005)); *Чем дальше в лес, тем больше дров, а где лес рубят, там щепки летят* (В.П. Авенариус. Юношеские годы Пушкина (1888)). Представлены также усечение состава: – *Какие новости, чем дальше в лес! Чего только не узнаешь случайно о старинных друзьях* (Д. Симонова. Сорванная слива (2002)); расширение состава: *Но чем далее* - пословица имеет структуру сложноподчиненного предложения с союзом *чем...тем*, однако встречается также вариант бессоюзного сложного предложения: *В трех соснах запутался. Дальше в лес – больше дров* (Л. Яхнин. Слова знакомые и незнакомые: Метафоры // «Мурзилка», 2002). Несмотря на тип трансформации, в большинстве случаев пословица сохраняет сопоставительные отношения между частями и компаративы в двух. Во фразеологических словарях данный оборот *же из них совершенно выбился и устал как сивый мерин*

Охарактеризованные с разной степенью полноты исследовательские задачи, в ходе решения которых корпус выполняет функцию моделирования естественной языковой среды, не охватывают всего круга корпусных исследований, но, возможно, послужат стимулом для обращения к корпусу текстов при выполнении подобных студенческих работ.

Анализ трансформаций паремиологических и фразеологических единиц – один из аспектов их исследования на базе корпуса текстов. С использованием корпусных методов можно решать и другие задачи, способствующие разностороннему и достоверному изучению фразеологизмов и паремий. Например, при проведении исследования, посвященного устойчивым выражениям, содержащим цветообозначения, студент определяет с помощью НКРЯ, описание каких действий и характеристик человека сопровождается сравнением *как сивый*. Определение содержания курса синтаксиса сложно-подчиненного предложения в учреждениях общего среднего образования, минимизация лингвистического материала осложняется дискуссионностью многих вопросов синтаксической теории. Остановимся на некоторых спорных моментах классификации сложноподчиненного предложения. В истории русской лингвистической науки широкое распространение получили логико-семантическая и структурно-семантическая классификации. Логико-семантическая классификация (А.Х. Востоков, С.И. Абакумов, Ф.И. 1840)); *Узнал себя, свои манеры... бездарен, как сивый мерин* (А.П. Чехов. Записные книжки (1891–1904)).

Литература

1. Образовательный портал Национального корпуса русского языка [Электронный ресурс]. – Режим доступа: <http://studiorum.ruscorpora.ru/>.
2. Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru/>.
3. Толковый словарь Ефремовой / Т.Ф. Ефремова; 2000 [Электронный ресурс]. – Режим доступа: <http://dic.academic.ru/contents.nsf/efremova/>.
4. Королева, М. Как это по-русски? Апельсиновый или апельсиновый? / М. Королева // «Российская газета» [Электронный ресурс]. – Режим доступа: <http://www.rg.ru/2013/03/14/koroleva.html>.
5. Большой фразеологический словарь русского языка. Значение. Употребление. Культурологический комментарий / отв. ред. В.Н. Телия. – М.: АСТ-ПРЕСС КНИГА, 2006. – 784 с.