

*Н.В. Ильина
А.Ф. Леонтьева*

РОЛЬ КОРПУСНОЙ ЛИНГВИСТИКИ В ОБРАЗОВАТЕЛЬНОМ ПРОЦЕССЕ

Статья посвящена анализу использования возможностей корпусной лингвистики при изучении родных и иностранных языков на материале Британского национального корпуса и корпуса удмуртского языка. Анализируются способы изучения семантической структуры слова, принадлежности к частям речи, частотность использования языковых единиц на материале указанных корпусов.

Ключевые слова: корпусная лингвистика, классификация корпусов, Британский национальный корпус, корпус удмуртского языка, методика преподавания языков.

Одним из перспективных направлений современной лингвистики и лингводидактики является изучение языков через использование лингвистических корпусов. Данная статья посвящена описанию возможностей использования корпусной лингвистики в образовательном процессе на материале удмуртского и английского языков.

Термин «*корпусная лингвистика*» используется для обозначения отрасли компьютерной лингвистики, которая занимается разработкой общих принципов построения и использования лингвистических корпусов (корпусов текстов) с использованием компьютерных технологий. Под корпусом текстов (а также лингвистическим или языковым корпусом) понимается большой унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач [1, с. 7]. Цель корпусной лингвистики заключается в изучении языка и его закономерностей через массивный объем примеров использования языковых единиц, действительно функционирующих в письменной или устной форме. Сегодня существуют различные виды языковых корпусов (устные/письменные, одноязычные/многоязычные, аутентичные и др.). Наиболее крупными считаются корпуса русского языка (Национальный корпус русского языка, <http://www.ruscorpora.ru>, содержит 500 млн слов русского языка), английского языка (British National Corpus, <http://corpus.byu.edu/bnc/>, содержит 100 млн слов британского варианта английского языка; Corpus of Historical American English (COHA), содержит 450 млн слов американского варианта английского языка), немецкого языка (корпус Берлинской Бранденбургской академии наук (DWDS), <https://www.dwds.de/>, содержит 106 млн слов немецкого языка).

Создаются также корпуса языков малочисленных народов. Так, например, существует корпус удмуртского языка, разработанный Йормой Луутоненом, Леонидом Ившиным и Кари Аро в университете г. Турку (Финляндия). В

корпус входит около 11000 текстов из удмуртских газет и журналов «Удмурт дунне», «Кенеш», «Вордскем кыл», «Инвожо», «Зечбур» и «Кизили». Также существует платформа корпуса удмуртского языка, которая была разработана Тимофеем Архангельским, Марией Медведевой и Екатериной Георгиевой в рамках проекта в Школе лингвистики национального исследовательского университета «Высшая школа экономики» в 2014 году. На веб-странице данного корпуса предложены три вида корпусов: основной корпус литературного удмуртского языка, в котором содержатся тексты из прессы, блогов и небольшой объем текстов художественной и научной литературы, корпус удмуртских социальных сетей и звуковой корпус удмуртских диалектов.

Одним из ключевых понятий корпусной лингвистики является «конкорданс», под которым понимают специальную программу, которая помогает анализировать большие массивы текстов и выявлять в них закономерность применения тех или иных языковых единиц. Конкорданс осуществляет поиск заданного словосочетания, сочетания морфем, синтаксической структуры в корпусе текстов. На основе полученных данных делается вывод о значениях, выполняемых функциях и особенностях употребления той или иной языковой единицы в конкретном языке [2, с. 130]. Роль конкорданса особенно важна для изучения устойчивых сочетаний, фразеологии, связи слов в словосочетании.

В образовательном процессе языковой корпус также широко используется для формирования и развития лексико-грамматических навыков учащихся в практике преподавания языков, при изучении лингвистических дисциплин, а также для сбора фактического материала при выполнении исследовательских проектов. Изучение языка с опорой на языковой корпус активизирует индуктивный подход к изучению языка, что, в свою очередь, позволяет формулировать правила, анализируя реально существующие тексты. Опираясь на данные языковых корпусов, студенты могут определить частотность использования и функционально-стилевую принадлежность слова, описать его семантическую и грамматическую структуру. Так, например, анализ данных, полученных в результате работы с Британским национальным корпусом, показывает, что прилагательное *smart* используется в значениях «нарядный», «замысловатый», «опрятный», «умный»: англ. He remembered Charlie Hatton's brand-new flat, the pretty young wife with her *smart* clothes. – рус. Он вспомнил совершенно новую квартиру Чарли Хаттона, симпатичную молодую жену с ее *нарядной одеждой* (здесь и далее – перевод авторов статьи); англ. He was *smart* and he went down well with the customers. – рус. Он был *умен* и хорошо общался с клиентами; англ. New objects: closed curves, *smart* arrows, freehand objects, rounded rectangles, regular polygons. – рус. Новые объекты: замкнутые кривые, *замысловатые* стрелки, произвольные объекты, скругленные прямоугольники, правильные многоугольники.

Данные языкового корпуса могут быть также использованы для определения принадлежности слова к частям речи. В английском языке слово *water* может быть представлено в предложениях различными частями речи:

- существительное: англ. The leopard now switched his attention to the open, sandy space leading to the *water*. – рус. Теперь леопард переключил свое внимание на открытое песчаное пространство, ведущее к *воде*; англ. Needs of the human population: *water*, food, firewood, clothing, shelter, plus health, educational and transport facilities as well as luxuries. – рус. Потребности населения: *вода*, еда, дрова, одежда, жилье, медицинские, образовательные и транспортные средства, а также предметы роскоши;

- прилагательное: англ. They are still abundant at present, although a little less common in shallow *water* sites than they were in the Palaeozoic and Mesozoic, but they are conspicuous and varied components of deep *water* faunas. – рус. Они все еще многочисленны в настоящее время, хотя на *мелководных* участках встречаются намного реже, чем в палеозое и мезозое, но они являются заметными и разнообразными компонентами *глубоководных* фаун;

- глагол: англ. *Water* plants first thing in the morning, or last thing at night, so that it doesn't evaporate, but goes straight to the roots. – рус. *Поливайте* растения первым делом утром или вечером, чтобы вода не испарялась, а шла прямо к корням; англ. Continue to *water* ornamental plants in containers and hanging baskets as necessary. – рус. При необходимости продолжайте *поливать* декоративные растения в контейнерах и подвесных корзинах.

Значительна роль корпусного анализа при определении коллокаций, фразовых глаголов, употребления предлогов. Ярким примером может быть анализ использования предлогов в словосочетаниях *in the street* «на улице» (949 случаев употребления) и *on the street* «на улице» (345 случаев употребления). Анализ частотности использования этих словосочетаний показывает превалирование варианта *in the street*; анализ источников позволяет при этом предполагать, что вариант *on the street* чаще встречается в разговорной речи и художественной литературе, а также в таблоидах.

Проиллюстрируем возможности использования языкового корпуса, разработанного Т. Архангельским, М. Медведевой и Е. Георгиевой, в практике преподавания удмуртского языка. Так, например, изучив примеры со словом *йыр*, студенты могут определить полисемию этого слова: а) удм. Снарядъёс *йыр* йылтй лобо. – рус. Снаряды над *головой* пролетают; б) удм. Кӧс турынэн бергес кылим – *сизьйыр* вазь вуиз. – рус. Запоздали мы с уборкой сена – *клевер* рано расцвел; в) Нош алигес веросьёсын но выжыкыльёсын «Кубиста *йыр*» нимо книгаез «Удмуртия» издательствоын печатласькиз. – рус. Недавно в издательстве «Удмуртия» вышла книга с рассказами и сказами под названием «*Виток* капусты». Изучив примеры, обучающиеся смогут сделать следующие выводы: 1) слово *йыр* используется в своем прямом значении «голова» (предл. 1); 2) слово *йыр* в сочетании со словом *сизь* образует новое слово «клевер» (предл. 2); 3) слово *йыр* в сочетании со словом *кубиста* используется в переносном значении «виток капусты» (предл. 3).

Как и в случае с английским языком, корпус удмуртского языка может быть использован для определения принадлежности слова к частям речи. Анализ демонстрирует, что в удмуртском языке слово *чебер* может быть представлено в предложениях разными частями речи:

- прилагательное: удм. Кытысь сыће *чебер* нылкышноез шедьтйды? – рус. Откуда вы нашли такую *красивую* девушку?; удм. Макем *чебер* но шулдыр гурт бырыны кулэ шат? – рус. Разве должна исчезнуть настолько *красивая* и прекрасная деревня?;

- наречие: удм. Нош кыће чылкыт, *чебер* кылйське. – рус. Как чисто, *красиво* звучит; удм. Ярам ай, беризь *чебер* сяськая, но со пöсьтолэзе гинэ усьтйське. – рус. Ладно еще липа *красиво* цветет, но она только в июле расцветает;

- существительное: удм. Нокытын но валэктэмын öвöл, ма герза удмурт антропологиез но *чебер/чебертэмлык* валатоньёсты. – рус. Нигде нет объяснения того, что объединяет удмуртскую антропологию и понятия *красота/некрасивость*; удм. Тй соосты *чебер* понна-а яке мукет мугьёсын-а вордйськоды? – рус. Вы их для *красоты* или для других целей разводите?

Таким образом, использование лингвистического корпуса в образовательном процессе усовершенствует процесс изучения языков, позволяя студентам работать с аутентичными материалами. Ценность использования корпусной методики в практике преподавания языков заключается в том, что она включает творческую работу студента с эмпирическими данными, репрезентативно воспроизводящими существующие в языке модели.

СПИСОК ЛИТЕРАТУРЫ

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов. Иркутск: Изд-во Иркут. гос. лингв. ун-та, 2011. 161 с.
2. Чернякова Т.А. Использование лингвистического корпуса в обучении иностранному языку [Электронный ресурс] // Язык и культура. Томск: Изд-во Нац. исслед. Томского гос. ун-та, 2011. №4. С. 127-132. URL: <http://www.lib.tsu.ru/mminfo/000349304/16/image/16-127.pdf> (дата обращения: 1.11.2019).
3. <http://udmurt.web-corpora.net/> (дата обращения: 2.11.2019)
4. <http://bncweb.lancs.ac.uk/cgi-bin/bncXML/BNCquery.pl?theQuery=search&urlTest=yes> (дата обращения: 2.11.2019)

Ильина Наталия Владимировна
кандидат филологических наук, доцент, кафедра лингвистической типологии и лингводидактики
Удмуртский государственный университет
E-mail: natilina@yandex.ru

Леонтьева Александра Филипповна
магистрант, специалист по учебно-методической работе, кафедра лингвистической типологии и лингводидактики
Удмуртский государственный университет
E-mail: sandra199608@rambler.ru