

ИСПОЛЬЗОВАНИЕ КОРПУСНЫХ ТЕХНОЛОГИЙ ДЛЯ СБОРА ЛИНГВО- СТАТИСТИЧЕСКОЙ ИНФОРМАЦИИ

Яцевич Е.А.

*Белорусский государственный университет,
Беларусь, Минск*

В статье описываются основные возможности по применению корпусных технологий для сбора линво-статистической информации на основе анализа методически ориентированного корпуса языковых учебных текстов, а также рассматриваются базовые принципы его создания.

Ключевые слова: *корпусные технологии, методически ориентированный корпус языковых текстов*

USE OF CORPUS TECHNOLOGIES FOR COLLECTING LINGVO-STATISTICAL INFORMATION

Jacevich K.A.

*Belarusian State University,
Belarus, Minsk*

The article describes the main opportunities for use of corpus technologies for collecting linvo-statistical information on the basis of the analysis of methodically oriented corpus of language study texts in the specialty and discusses the basic principles of its creation.

Key words: *corpus technologies, methodically oriented corpus of language texts*

Активное применение корпусных технологий в области лингводидактики стало следствием бурного развития информационных технологий и использования их в сфере компьютерной лингвистики, одним из направлений которой является корпусная лингвистика.

Лингвистический, или языковой, корпус представляет собой большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач [1, с. 7]. В то же самое время, разработчики НКРЯ уже дают определение корпусу как информационно-справочной системе, основанной на собрании текстов на некотором языке в электронной форме. Национальный корпус представляет данный язык на определенном этапе (или этапах) его существования и во всём многообразии жанров, стилей, территориальных и социальных вариантов и т. п. [2]. Указанные возможности лингвистического корпуса обеспечивают

его лингводидактический потенциал и создают условия использования корпуса как перспективного инструмента при обучении языку инофонов в теоретическом и практических аспектах.

Применение корпуса позволяет избежать интуитивности, а также субъективности преподавателя при формировании учебно-методических материалов, при составлении лексических и грамматических минимумов в ходе работы и над профессиональным модулем, и над модулем общего владения, учебных программ разного уровня и обеспечить их релевантность учебным задачам, причем при применении корпуса увеличивается и скорость обработки лингвистического материала, и достоверность выборки, что в комплексе повышает эффективность деятельности специалиста в области преподавания изучаемого языка. Данной проблеме посвящены работы таких исследователей, как Н.Б. Гвишиани, А.В. Зубова, В.П. Захарова, В.В. Рыкова, В. А. Плунгяна, Л.В. Рычковой, Т. Джонса, М. Льюса, С. Конрада, М.Хоуи, Д. Байбера и многих других.

В связи с тем, что данные корпуса используются как доказательная основа при оценке необходимости рассмотрения того или иного языкового явления для анализа потребности в его дальнейшем изучении, важно соблюдать принципы формирования корпуса текстов. Немаловажным является понимание того, что большой национальный корпус не может использоваться для решения всех лингводидактических целей одновременно, в силу чего в последнее время ведутся работы по созданию именно методически ориентированных языковых корпусов для обеспечения работы над лингводидактическими задачами применительно к языку специальности конкретной научной и профессиональной области. Именно анализ методически ориентированного языкового корпуса позволяет сфокусировать внимание на действительно важных и нужных языковых явлениях и не упустить их при определении содержания обучения языку специальности, а также при работе над профессиональными минимумами. Под профессиональным минимумом мы понимаем вслед за Б.А. Лapidусом «достаточную для обеспечения задач курса в области данного вида речевой деятельности, особым образом отбираемую и обязательную для усвоения совокупность языковых явлений (или паралингвистических, в широком смысле, явлений, или системных знаний об изучаемом языке, или фоновых знаний), каждое из которых предназначено для использования в предусмотренном программой общении и требует либо введения и отработки в упражнениях, либо только введения (демонстрации)» [3, с. 23]. Значимой является и возможность при составлении методически ориентированного языкового корпуса учитывать уровень владения изучаемого языка инофоном и формировать корпус в зависимости от предполагаемых к решению лингводидактических задач. Немаловажным в аспекте учета уровня владения языком инофоном является различение активного и пассивного языкового материала и, как следствие, влияние данного критерия на принципы отбора лексических единиц. Данный вопрос глубоко исследовался в трудах таких ученых, как Г. Пальмера, М. Уэста, Л.В. Щербы, А.С. Герда, Р.Г. Пиотровского, Р. Ладло, Б.А. Лapidуса и многих других.

Среди базовых принципов формирования методически ориентированного языкового корпуса следует назвать следующие, выступающие как основные для корпусов такого типа [4, с. 123-124]:

- 1) принцип аутентичности в широком понимании;
- 2) критерий профессиональной релевантности (т.е. репрезентативности);
- 3) критерий лингвострановедческой ценности;
- 4) критерий лингвистической адекватности и насыщенности;
- 5) критерий лексической насыщенности (насыщенность лексическими средствами общения, как понятийными словами и коллокациями, так и функциональными средствами реализации коммуникативных стратегий);
- 6) принцип проблемности (тексты и материалы, вызывающие интерес и желание высказаться);
- 7) принцип соответствия интересам обучаемых;
- 8) принцип доступности.

При учете указанных принципов, соотносимых с методическими и дидактическими принципами, создание методически ориентированного языкового корпуса будет обладать практически направленным характером и соответствовать реальным потребностям специалиста по преподаванию языка специальности инофонам.

В качестве примера востребованности и продуктивности данного вида деятельности приведем разработанный нами методически ориентированный языковой корпус – массив языковых учебно-научных текстов по экономике на русском языке для иностранных студентов, владеющих русским языком в объеме уровней минимальной коммуникативной достаточности / пороговой коммуникативной достаточности (МТЭИ₁), в который были включены тексты из учебно-методических пособий по научному стилю (экономический профиль) для подготовительных отделений вузов РБ (в частности БГУ и БГЭУ).

Результатом работы явилось формирование на основе полученных после анализа материала корпуса таких лингво-информационных ресурсов, как «Тезаурус по экономике для изучающих русский язык», а также баз данных, отражающих лексический состав таких частей речи, как имена существительные, глаголы, предлоги и их грамматические характеристики в статистическом аспекте. Кроме того, были разработаны базы знаний, одна из которых предназначена для отражения информации о значениях предложно-падежных форм имен существительных, а вторая – информации относительно основных деривационных моделей (словообразовательные пары, цепочки, парадигмы) для инофонов, владеющих русским языком в объеме уровней минимальной коммуникативной достаточности / пороговой коммуникативной достаточности.

Нельзя сказать, что главным принципом при работе с корпусом выступает принцип частотности. Как справедливо отмечает Л.Н. Беляева, при необходимости в получении статистических данных необходимо четко представлять, с какой лингводидактической целью

проводится анализ информации по корпусу, так как «получение неких числовых данных без опоры на лингвистическую постановку задач и точное понимание, что, собственно, описывается конкретной формулой, делают такие экзерсисы бессмысленными и бесперспективными» [5, с. 43].

Именно корпусные технологии на основе анализа текстов, включенных в их состав в соответствии с базовыми принципами, способствуют отбору релевантных лингводидактическим задачам, соотносимым в свою очередь с реальными потребностями инофонов для осуществления продуктивной речевой деятельности на изучаемом языке в профессиональной сфере, лексико-грамматических данных, наиболее репрезентативных в функциональном аспекте речевой деятельности. Полученные лексико-грамматические данные, представленные в систематизированном виде посредством использования предлагаемых разработчиками различных программных средств, организуются в определенную информационную систему, которая обеспечивает хранение данных, их обновление при необходимости и предоставляет возможность продуцирования новых знаний на основе уже содержащихся в информационной системе.

Комплексное применение информационных технологий в области лингводидактики позволяет формировать информационное образовательное пространство в соответствии с мировыми тенденциями в области образования.

Литература

1. Захаров, В.П. Корпусная лингвистика: учебник для студентов гуманитарных вузов / В. П. Захаров, С.Ю. Богданова. – Иркутск: ИГЛУ, 2011. – 161 с.
2. Что такое Корпус? [Электронный ресурс]. – Национальный корпус русского языка. – Москва, 2003 – 2017. – Режим доступа : <http://ruscorpora.ru/corpora-intro.html>. Дата доступа : 10.09.2017.
3. Лapidус, Б.А. Проблемы содержания обучения языку в языковом вузе / Б.А. Лapidус. – М. : Высшая школа, 1986. – 144 с.
4. Горина, О. Г. Использование технологий корпусной лингвистики для развития лексических навыков студентов-регионоведов в профессионально-ориентированном общении на английском языке : диссертация ... кандидата педагогических наук : 13.00.02 / Горина Ольга Григорьевна; [Место защиты: Моск. гос. ун-т им. М.В. Ломоносова. Филол. фак.]. – Москва, 2014. – 325 с.
5. Беляева, Л.Н. Лингвистические ресурсы информационной образовательной среды: состав, структура, функции / Л.Н. Беляева // Известия РГПУ им.А. И. Герцена. – № 171. – СПб, 2014. – С. 47-52.

ⁱ Рекламный буклет. Перевод с греческого.

ⁱⁱ Соловьев Е.А. Достоевский. Его жизнь и литературная деятельность. М., 2015.

ⁱⁱⁱ Одной из трагических страниц в истории современной Греции является "исход" сотен тысяч греков-коммунистов в страны социалистического лагеря в 1947-1949 годах, в СССР в том числе, в надежде на светлое будущее в стране, где все равны и все счастливы. В конечном итоге они разделили судьбу

понтийских, мариупольских, цалкинских греков СССР, будучи сосланными советским режимом в удаленные районы Казахстана и Узбекистана, как потенциальные предатели.

^{iv} Джуха И.Г. Мариупольские, цалкинские, понтийские. - М.: Издательский дом Международного университета в Москве, 2016, С. 329.

v ΠΑΓΑ ΒΟΥΛΗ. Μόσχος & Αθηνά. Москов и Афина (Сказка, сочиненная учениками греческой школы города Москвы). - Муниципальное объединение Аспропиргос, 2016.

^{vi} Пусть меня кто-то и упрекнёт, назовем это так - в гумбольдтианстве, что я слишком акцентирую внимание на том, что при помощи языка идет постижение мира другого народа. Но я полностью согласна с тем положением неоднозначной, а иногда и спорной теории немецкого ученого, которое утверждает, что, овладевая другими языками, человек расширяет «диапазон человеческого существования»^{vi}.