

КОРПУСНАЯ ЛИНГВИСТИКА КАК ИННОВАЦИОННАЯ ТЕНДЕНЦИЯ РАЗВИТИЯ ФИЛОЛОГИЧЕСКОГО ОБРАЗОВАНИЯ

Мирошник С.О.

Корпусная лингвистика - одна из методик, наиболее стремительно развивающихся в современном теоретическом и прикладном языкознании. Это относительно новое лингвистическое течение, которое начало свое активное влияние в 60-х годах XX века в связи с интенсивным развитием компьютерных технологий. Термин «корпусная лингвистика» прочно вошел в научный обиход в последние десятилетия XX века с публикацией в 1983 году сборника научных трудов “Corpus Linguistics: Recent Developments in the Use of Computer Corpora in English Language Research” на материалах конференции ICAME “Conference on the Use of Computer Corpora in English Language Research”.

В лингвистике, корпус - тщательно подобранная и обработанная по определенным правилам совокупность текстов, которая используется для статического анализа и проверки статических гипотез, подтверждение отдельных лингвистических правил в данном языке. Корпус - основное понятие и фундамент корпусной лингвистики. Анализ и обработка различных типов корпусов является предметом большинства работ в области компьютерной лингвистики (например, извлечение ключевых слов), распознавания речи и машинного перевода, в которых корпуса часто применяются при создании скрытых маркированных моделей для маркировки частей речи и других задач. Корпуса и частотные словари могут быть полезны в обучении иностранным языкам. Классифицировать корпуса можно по различным признакам: цель создания корпуса, тип языковых данных, «литературность», жанр, динамичность, тип разметки, объем текстов и т.д. По критерию параллельности, например, корпуса можно разделить на одноязычных, двуязычных и многоязычных. Многоязычные и двуязычные делятся на два типа: параллельные - множество текстов и их переводов на один или несколько языков, или сопоставимые (псевдопараллельные) - оригинальные тексты на двух или

нескольких языках. Многочисленные исследования в области современного корпусного языкознания происходят в двух магистральных направлениях. Первое направление сосредоточено на разработке проблем, касающихся теории и практики создания корпусов, другими словами, концептуализации корпуса, типология корпуса, его назначение, объем, параметризация предметной области, репрезентативность, структурирование и принципы отбора базовых единиц, хранения и т.д. Второе направление направлено на исследование именно лингвистических корпусов, то есть, изучению языка с помощью корпусных методов. Однако, следует заметить, что четкой границы между указанными направлениями не существует, ведь практически все составители корпусов в то же время осуществляют и лингвистические исследования на их основе. Такая двухвекторность корпусной лингвистики обусловлена двойственной природой объекта ее исследования - текстового корпуса, который, с одной стороны, выступает в качестве исходного речевого материала для корпусной лингвистики, а с другой, является результатом деятельности этого языковедческого направления. Предметом корпусной лингвистики выступают теоретические основы и практические механизмы создания и эксплуатации языковых корпусов. Необходимо отметить, что корпусная лингвистика как система методов и принципов использования корпусов для изучения языка и для изучения / обучения языку имеет теоретические основы, но теоретические основы еще не являются научной теорией. Корпусная лингвистика применяется в методологии изучения языка с широкими возможностями во многих лингвистических направлениях и теориях. Так как корпусная лингвистика пользуется в своих исследованиях большими по объему и репрезентативными выборками текстов естественного языка, значит, существует несколько типов программного обеспечения, которые могут использоваться при исследовании: коммерческие компьютерные программы – конкордансеры (LEXA, MonoConc, MicroConcord, TACT, WordSmith, WordCruncher, Manatee (Bonito), IMS Corpus Workbench (CQP), XAIRA, LEXA, Virtual

Corpus Manager (VMC), EXMARaLDA Corpus-Manager (Co-Ma)), а также компьютерные программы, разработанные для особых процессов анализа. Эти типы программ применяются для сложных грамматических моделей. Работа с корпусом осуществляется с помощью специальных программных средств - конкордансеров (более простой тип программ) и корпусных менеджеров, предоставляющих различные возможности по получению необходимой информации из корпуса. Корпусный менеджер - это специальная поисковая система, которая задействует программные средства для поиска данных в корпусе, получения статистической информации и предоставления результатов пользователю в удобной форме. Результаты этой процедуры представляются в виде горизонтальных строк с поисковым словом посередине. Этот процесс называется KWIC (Key Word In Context).

Появление электронных корпусов существенно изменило способ исследования языковых явлений. Репрезентативные, сбалансированные, значительные по объему корпуса предоставляют информацию о динамике процессов изменения лексического состава языка, функционирование грамматических и синтаксических единиц, позволяющих осуществлять анализ употребления языковых единиц в разных стилях, жанрах и т.д. Если ранние корпусные исследования были сосредоточены главным образом только на применении языковых единиц (частота употребления определенных существительных, глаголов и прилагательных), то с появлением мощных компьютерных технологий исследователи получили возможность определять и анализировать более сложные лексические, грамматические и синтаксические явления.

Привычной практикой стало использование корпусов для укладки справочной литературы - грамматик, учебных словарей и справочников как для носителей языка, так и для тех, кто изучает.

Использование корпусов становится все более популярным в обучении и при изучении языка. Преподаватели и студенты все чаще обращаются к корпусу для осуществления быстрого разноуровневого поиска и анализа

информации, а конкорданс признается удобным инструментом для эффективного освоения языка. Корпуса текстов служат неоценимым ресурсом реальных примеров употребления тех или иных лексических и грамматических единиц, при условии, что преподаватели и студенты имеют доступ к корпусам и знакомы с техникой работы с корпусом.

Литература

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов / В.П. Захаров, С.Ю. Богданова. – Иркутск: ИГЛУ, 2011.
2. Демська–Кульчицька О. Основи національного корпусу української мови [монографія] / О. М. Демська-Кульчицька – Київ, 2005.
3. MacEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice / T. MacEnery, A. Hardie. – Cambridge University Press, 2012.
4. Meyer Ch. P. English Corpus Linguistics. An introduction / Ch.P. Meyer. – Cambridge University Press, 2004.

