СПЕЦИФИКА ИСПОЛЬЗОВАНИЯ КОРПУСОВ ТЕКСТОВ В ИССЛЕДОВАНИЯХ ЯЗЫКОВОЙ СЕМАНТИКИ

Астапкина Е.С.

Белорусский государственный университет

Аннотация. Данная статья посвящена рассмотрению возможностей корпусной методологии при проведении лингвистических исследований. Внимание уделяется корпусу текста как эффективному инструменту отбора, систематизации и анализа современной речевой практики, а также изучению нововведений и верификации лексики при обновлении словарей.

Ключевые слова: корпус, лингвистические информационные ресурсы, компьютерно-опосредованная среда, корпусная методология, лексические изменения.

Современная лингвистика характеризуется системным подходом: только сплошной анализ материала позволяет получить полную, системную картину состояния исследуемого объекта, которым и является тот или иной национальный язык, постоянно развивающийся и представленный колоссальным разнообразием случаев его употребления.

Проблема языковых изменений вполне традиционна и не поражает новизной, однако она относится к тем вечным темам, неисчерпаемость которых зиждется на неисчерпаемости самого языка, его динамизме и постоянном развитии. Это тем более справедливо в последние десятилетия, когда изменения в лексической системе языков буквально галопируют перед нашими глазами.

К лексическим изменениям относят пополнение словарного состава языка благодаря образованию новых слов на базе имеющихся, заимствование отдельных лексем из других языков, а также изменение лексических значений имеющихся в языке слов.

К сожалению, языковые изменения, уже достаточно широко представленные в речевой практике, не зафиксированы в лексикографических источниках: фиксация языковых новообразований в словарях обычно опаздывает за развитием языка.

Так, при исследовании современного состояния языка лингвист не всегда может получить объективные результаты исследования при обращении к печатным лексикографическим источникам ввиду их ограниченности в фиксации новых элементов лексической системы. Также следует отметить, что традиционный механический сбор лингвистического материала для такого рода исследований часто представляет собой трудоемкий и достаточно длительный этап, принимая во внимание, что объект изучения лингвистики в настоящее время – время компьютерно-опосредованной коммуникации – создается, хранится и реализуется в недоступном для изучения в «докомпьютерные» времена объеме [2, с. 9].

Однако данные трудности преодолеваются благодаря возможностям сетевых лингвистических ресурсов – в данном случае корпусов национальных языков.

Под лингвистическими ресурсами понимаются «средства лингвистической поддержки, которые используются для представления информации о соответствующем естественном языке (словари, онтологии, тезаурусы и пр.), а также для представления данных в системе обработки информации и для решения задач извлечения эмпирической информации. Кроме того, к лингвистическим ресурсам относятся собственно языковые ресурсы – полнотекстовые базы данных, национальные

и специализированные корпуса текстов, формирующие источник знаний о языках. Использование лингвистических ресурсов возможно как в «ручном» режиме, так и при решении задач автоматической обработки текста» [3, с. 49].

Лингвистика, как и другие области науки, автоматизируется. Современный «лингвоинформационный» формат данных облегчает систематизацию и анализ речевой практики. Так, «стали осязаемыми новые возможности изучения специально-предметных «выборок» языковых фактов, с которыми мы в большинстве случаев и имеем дело при исследовании семантики. Появившись в 60-е гг. XX в., корпусная лингвистика создала принципиально новую методологическую основу для репрезентации речевой практики» [2].

А.Н. Баранов в своей работе «Корпусная лингвистика» приводит следующее определение корпуса: «Корпус текстов – это вид корпуса данных, единицами которого являются тексты или их достаточно значительные фрагменты, включающие, например, какие-то полные фрагменты макроструктуры текстов данной проблемной области» [1, с.115].

В данном контексте можно привести пример следующих национальных корпусов текстов: Национальный корпус русского языка [6], British National Corpus [10]), оснащенных специализированной поисковой системой и представляющих мощный инструмент для проведения объективных исследований.

В.П. Захаров в «Корпусной лингвистике» упоминает, что «корпусы могут использоваться для получения разнообразных справок и статистических данных о языковых и речевых единицах. В частности, на основе корпусов можно получить данные о частоте

словоформ, лексем, грамматических категорий, проследить изменение частот и контекстов в различные периоды времени, получить данные о совместной встречаемости лексических единиц и т.д. Представительный массив языковых данных за определенный период позволяет изучать динамику процессов изменения лексического состава языка, проводить анализ лексико-грамматических характеристик в разных жанрах и у разных авторов, и т.д. Корпусы призваны служить также источником и инструментом многоаспектных лексикографических работ по подготовке разнообразных исторических и современных словарей» [5, с. 3].

Так, в компьютерно-опосредованной среде проверка лексики языка при обращении к корпусным данным помогает согласовать работу лингвистов с требованиями времени. Подобная стратегия успешно реализуется при составлении и обновлении словарей серии *COLLINS*. Так, для ввода новой леммы в словник, а также для иллюстрации ее современного употребления привлекаются данные «Bank of English» – репрезентативной части созданного Джоном Синклером корпуса *COBUILD*, или *Collins Birmingham University International Language Database* (англ. – ‘Международная языковая база данных «Коллинз» и Бирмингемского университета’).

Характерно следующее наблюдение Н. Д. Голева: «В последнее время лингвисты, не довольные интуитивным решением проблемы формирования лексического состава словарей, все чаще обращаются к Национальному корпусу русского языка и данным Интернета, а также к лингвистическому эксперименту как более надежным источникам» [4, с. 185].

Таким образом, корпусные инструменты решают следующие проблемы при проведении лингвистического исследования:

1) Отбор материала. При проведении лингвистического исследования, не требующего сплошной выборки языкового материала, отбор единиц осуществляется при помощи специально сформированных запросов (исходя из возможностей корпуса, его разметки), что уменьшает трудоемкость процесса сбора и обработки материала.

2) Представительность данных. При обращении к корпусу языка возможны квантитативные и статистические исследования.

3) Изучение редких/новых языковых явлений. Корпусные исследования дают возможность проследить за динамикой языковых изменений в разные временные отрезки. При этом наряду с окказиональными явлениями корпусная методология позволяет интерпретировать и частотные явления.

4) Объективность исследования: «...Использование корпусов делает возможным объективизировать лингвистику, найти более веские аргументы применительно к фактам, создать ситуацию повторяемости, что, по мнению А. Мустайоки, является важнейшим элементом науки» [6]. Так, именно возможность верификации результатов корпусного исследования гарантирует его эффективность, достоверность, эмпирическую релевантность. По мнению В.А. Плунгяна, «Лингвистика корпусов позволяет нам понять, каков язык на самом деле, а не каким мы хотим, чтоб он был» [7], что соответствует дескриптивному подходу к изучению и описанию языковых явлений.

Так, возможна, например, корпусная верификация степени освоенности англицизмов в русском языке. Такое исследование

можно провести как на базе всех текстов Национального корпуса русского языка, так на базе корпуса публицистических текстов, принимая во внимание то, что язык публицистики наиболее открыт к нововведениям. В рамках изучения заимствований корпусная методология позволяет выявить, употребляется ли слово в том же виде, в каком оно существует в своем языке (звуковой состав, структура слова); приспособилось ли слово к системе заимствующего языка (изменение слова по правилам заимствующего языка)¹; насколько часто слово употребляется в письменной речи; какова его словообразовательная активность.

Использование лингвистических информационных ресурсов в исследованиях лексики и семантики обладает существенным потенциалом, что обусловлено, с одной стороны, недостаточным задействованием такого вида ресурсов в исследовательской практике, а с другой стороны, высокой эффективностью данного подхода.

Корпусная методология, которая характеризуется универсальностью и эффективностью, позволяет системно исследовать языковую проблематику благодаря точности и многоаспектности полученных результатов, а также позволяет абстрагироваться от “субъективности” исследователя и приблизиться к объективному изучению языка.

Корпус языка фактически предназначен для точного изучения языковых свойств, особенностей и явлений, наблюдаемых в языке.

¹ В корпусе [6] предоставлена возможность лексико-грамматического поиска с указанием части речи, число, падеж, род слова и т.д.

ЛИТЕРАТУРА

1. Баранов, А.Н. Корпусная лингвистика / А.Н. Баранов // Введение в прикладную лингвистику: учебное пособие. — М.: Едиториал УРСС, 2003. — С. 112–137.
2. Баркович, А.А. Корпусная лингвистика: специфика современных метаописаний языка / А.А. Баркович // Вестник Томского государственного университета. — 2016. — № 406. — С. 5–13.
3. Беляева, Л.Н. Сетевые ресурсы в технологии перевода / Л.Н. Беляева // Вестник СПбГУ. Язык и литература. — 2016. — №4. — С. 45–55.
4. Голев, Н.Д. Лексическая реализация как функциональная характеристика словообразовательной системы русского языка и количественные параметры ее описания / Н.Д. Голев // Осмь десять : сб. науч. ст. к 80-летию И.С. Улуханова ; отв. ред. М.А. Малыгина. — М., 2015. — С. 183–193.
5. Захаров, В.П. Корпусная лингвистика: Учебно-методическое пособие / В.П. Захаров. — СПб., 2005. — С. 3.
6. Национальный корпус русского языка [Электронный ресурс]. — Режим доступа: <http://www.ruscorpora.ru>. — Дата доступа: 24.11.2020.
- 7.3. Плунгян, В.А. Почему современная лингвистика должна быть лингвистикой корпусов [Электронный ресурс]. — Режим доступа: <http://polit.ru/article/2009/10/23/corpus/>. — Дата доступа: 15.08.2022.
8. Плунгян, В.А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики / В.А. Плунгян // Русский язык в научном освещении. — 2008. — № 16 (2). — С. 7–20.
9. Barkovich, A. Informational Linguistics: The New Communicational Reality / A. Barovich. — Newcastle upon Tyne: Cambridge Scholars Publishing, 2020. — 265 p.
10. British National Corpus [Electronic resource]. — Mode of access: <http://www.natcorp.ox.ac.uk>. — Date of access: 24.11.2020.