

6. ОБЩЕНИЕ НА ЯЗЫКАХ – ЯЗЫКИ ОБЩЕНИЯ

ИНСТРУМЕНТАЛЬНО-НЕЗАВИСИМЫЙ КОРПУС: НОВЫЕ РЕАЛИИ ЭМПИРИЧЕСКОЙ ПАРАДИГМЫ

Баркович А. А.

Белорусский государственный университет

Аннотация. Статья посвящена рассмотрению лингвистической специфики инновационного типа корпусов – инструментально-независимого корпуса. Инструментально-независимый корпус, как доступная для обработки любым корпусным менеджером совокупность текстов или иных артефактов коммуникации, являет собой актуальный феномен прикладного языкознания. В частности, на примере информационного ресурса данного типа *EuroNews Green Corpus* были исследованы аспекты интерпретации, структуризации и моделирования языковых данных англоязычного экологического дискурса. Практикой подтвержден существенный метаязыковой потенциал инструментально-независимого корпуса в контексте эмпирической парадигмы.

Ключевые слова: корпусная лингвистика, инструментально-независимый корпус, корпус-менеджер, статистика, метаданные.

Сегодня лингвистика стала чрезвычайно пластичной, демонстрируя практически полный охват коммуникации, в том числе ее технических и информационных аспектов [см., напр., 1]. Пожалуй, наиболее актуальным аспектом в данной связи является корректная метаязыковая *организация* корпусных исследований, о чем свидетельствует опыт использования инструментария нового поколения, позволяющего не только собирать

«очевидную» статистику, но и осуществлять ее сложную и комплексную обработку на уровне *метаданных* [2]. В данной связи продуктивны лингвистические исследования, посвященные комплексной характеристике референтной речевой практики, анализу ее концептуальных основ, моделированию ее метаструктуры и иным существенным характеристикам объектной проблемной области.

Опора на репрезентативный материал лингвистическом исследовании – один из ключевых критериев объективности его результатов. Любые отклонения от нормы в речи являются информативными в лингвистическом аспекте, однако, благодаря корпусам, не заслоняют языковой мейнстрим при рассмотрении особенностей языковой практики как таковой. При этом корпусный подход исключает поверхностность и фрагментарность исследований, актуализируя вопросы совместимости и преемственности разнообразной метаязыковой практики.

В целом корпусный формат обеспечивает возможности лингвистического рассмотрения как отдельных языковых единиц или текстов, так их совокупности, – что исключительно важно при исследовании коммуникации. Дополнительные возможности, появляющиеся при компьютерно-опосредованной обработке языковых данных, – весомый аргумент при верификации результатов. Корпусный анализ как методика признан специалистами эффективным инструментарием решения прикладных лингвистических задач [3]. Это обусловлено возможностью благодаря корпусам собирать, пополнять и изучать гигантские объемы языковых данных, для изучения которых еще не так давно требовались огромные материальные ресурсы и временные затраты. Показательна в данной связи позиция А. Н. Баранова, который говорит о преимуществе использования корпусных методов при обработке

статистических данных, зачастую оказывающихся решающим аргументом при оценке той или иной гипотезы [4, с. 135–137].

Для компьютерной поддержки лингвистических исследований сегодня создаются как узкоспециализированные корпус-менеджеры, так и универсальные программы, которые можно использовать для анализа текстовых ресурсов разного происхождения и состава. В настоящее время подобные программы, так называемые «корпусные оболочки», как правило, доступны посредством Интернета. Среди апробированных и пользующихся известностью инструментов такого рода можно отметить, в частности, *AntConc*, *Nooj*, *Tropes*, *Sketch Engine* и ряд других.

С корпусными оболочками могут быть использованы корпуса «открытой» архитектуры – инструментально-независимые. **Инструментально-независимый корпус** – это доступная для обработки любым корпусным менеджером совокупность текстов или иных артефактов коммуникации. Компьютерно-опосредованная форма текстов уже является показателем их атрибутированности той или иной искусственно-языковой разметкой. Это, в свою очередь, позволяет загружать инструментально-независимый корпус практически в любую корпусную оболочку или использовать любой доступный корпус-менеджер. Примером такого корпуса является, в частности, *EuroNews Green Corpus* [5]. Проведенная обработка данного корпуса корпусным менеджером *Sketch Engine* и другими корпусными менеджерами показала высокую эффективность подобной практики и, соответственно, лингвистическую состоятельность инструментально-независимого корпуса как класса метаязыкового инструментария [6].

Подобный подход обладает таким немаловажным преимуществом как полная доступность и языковая «прозрачность» самих текстов, что практически исключено в «классическом» корпусном формате.

Типологические особенности корпусов подобного типа – их открытая архитектура и автономность, принципиально совместимые с различными корпусными инструментами. Это позволяет при их задействовании сосредотачивать основные усилия на предметной лингвистической работе, не расходуя исследовательские ресурсы на уже в общем достаточно стандартизированную практику разработки и отладки программного обеспечения корпусных менеджеров. Корпусы инструментально-независимого типа – инновационный тип лингвистического инструментария и именно за ними, несомненно, будущее интердисциплинарных исследований.

В целом проведение лингвистического анализа больших массивов языковых данных с помощью современной корпусной методики обладает большим потенциалом. Корпусный инструментарий эффективен не только для автоматизации сбора вполне очевидных статистических данных, но и для выполнения достаточно сложной аналитической работы по интерпретации, структуризации и моделированию коммуникации. Так, посредством инструментально-независимых корпусов, как показывает практика, могут быть успешно интерпретированы и оценены такие метаязыковые категории как токены, ключевые слова и термы. Исследуемые текстовые данные могут быть соответствующим образом структурированы: в том числе, это касается таких категорий языковой «материи», как леммы, коллокации и *n*-граммы. На базе инструментально-независимых корпусов можно проводить и моделирование баз языковых данных в виде конкордансов, тезаурусов и частотных словарей.

Таким образом, проведение корпусных исследований посредством инструментально-независимого корпуса позволяет материализовать существенный потенциал для результативной лингвистической работы.

Список литературы

1. Рассказы о сновидениях. Корпусное исследование устного русского дискурса. Под ред. А. А. Кибрик, В. И. Подлесская. – М.: Языки славянской культуры, 2009. – 736 с.
2. Баркович, А. А. Корпусная лингвистика: специфика современных метаописаний языка / А. А. Баркович // Вестник Томского государственного университета. – 2016, № 406. – С. 5–13.
3. Баркович, А. А. Лингвистические корпуса китайского языка: функциональный аспект / А. А. Баркович, Ван Цин // Вестник МГЛУ. Сер. 1. Филология. – 2015, № 5 (78). – С. 105–113.
4. Баранов, А. Н. Введение в прикладную лингвистику / А. Н. Баранов. – М.: URSS. 2021. – 368 с.
5. EuroNews Green Corpus: иллюстративный корпус текстов экологического дискурса / О. А. Рипинская, А. А. Баркович, available at: <http://surl.li/fmiip> (30.03.2023).
6. Баркович, А. А. Лингвистический потенциал корпусного исследования англоязычного экологического дискурса / А. А. Баркович, О. А. Рипинская // Весці БДПУ. Сер. 1. Педагогіка. Псіхалогія. Філалогія. – 2023, № 2. – С. 84–88.