

Национальный корпус русского языка и проблемы гуманитарного образования.
Материалы международной научной конференции. Москва 19-20 апреля 2007. М.
2007. С. 58-60.

Роль корпусов в лингвистических исследованиях и преподавании языков

Арто Мустайоки
Хельсинкский университет
arto.mustajoki@helsinki.fi

По определению, корпус является систематически собранной и особым образом упорядоченной совокупностью текстов. Поскольку корпус реального языкового употребления представляет собой результат применения языка носителями, его можно считать самым естественным источником для научного исследования языка. Однако это якобы очевидное утверждение далеко не бесспорно в свете истории языкознания. Наоборот, целью приложения сил лингвистов считалось и все еще часто считается не то, как люди говорят или пишут (*parole*, речь, *performance*, узус), а то, что наличествует в сознании носителей (*langue*, язык, *competence*, норма). Никто, конечно, не может отрицать существование речи, но многие лингвисты считают ее вторичной по отношению к *langue*. Из такого взгляда следует, что в изучении языка нельзя опираться на речь, корпуса.

Несмотря на скепсис относительно пользы корпусов в лингвистических исследованиях, все более обширные корпуса производятся благодаря быстрому развитию компьютерных технологий. Они же и широко используются в изучении разных языков. На что они годятся? С точки зрения общей цели лингвистического интереса к языку, можно различать два подхода к корпусам. Согласно первому из них, с помощью корпусов можно описать речь, узус, то, как люди говорят и пишут, но только это. Согласно второму – данные, основывающиеся на корпусах, могут быть также использованы при описании *langue*. Далее можно, естественно, продолжить дискуссии о роли корпусов, задаваясь вопросом: что, в конце концов, должны описать лингвисты, особенно если речь идет об «официальном», каноническом описании языка в виде грамматик и толковых словарей. Вместо ответа рассмотрим некоторые конкретные примеры исследований, которые основываются на использовании корпусов и которые в таком виде были бы невозможны без них.

Сначала два примера, касающихся грамматических явлений. Первый из них относится к так называемому второму родительному падежу типа *чаю*. Грамматики русского языка, как правило, комментируют его судьбу весьма

лаконично: второй родительный вытесняется и практически уже вытеснился первым родительным падежом (*чая*) даже в разговорной речи. Языковой материал (узус, корпус), однако, не подтверждает такого взгляда (ср. Мустайоки & Пуссинен 2006). В некоторых контекстах типа «чашка чая/чаю» второй родительный падеж, действительно, практически уже исчез, но в некоторых других случаях он употребляется широко. Есть контексты, в которых он явно более предпочтителен, чем первый родительный падеж (например, *народу было много, набралось много народу, хочешь чаю*).

Второй пример относится к русской «стихийной» конструкции типа *Лодку унесло ветром*. Ее изучению посвящен ряд серьезных научных исследований. Они страдают, однако, от того обстоятельства, что в них повторяется весьма ограниченное количество примеров. Таким образом, и выводы односторонни. Собрав более обширный материал (ок. 3000 примеров), нам удалось пересмотреть представления о сущности данной конструкции (Мустайоки & Копотев 2005; Мустайоки 2006). Оказалось, например, что каузатором в этих конструкциях только в одной трети случаев является стихийная, природная сила. Также выяснилось, что вопреки ожиданиям, конструкция употребляется не только в прошедшем времени совершенного вида, но и в настоящем, будущем и прошедшем времени несовершенного вида.

Если приведенные выше примеры представляют традиционные грамматические вопросы, то следующий пример относится к таким прагматическим явлениям, которые в грамматиках не отмечаются. Другими словами, роль корпуса здесь иная, чем при изучении грамматических явлений. Можно предположить, что в «докорпусную» эру языковеды вообще не интересовались такими вопросами. Наш пример касается случаев, в которых говорящий поступает против обычных максим диалога, делая вид, что понимает или не понимает. При использовании обширного корпуса и особой техники поиска (в одном предложении должно встретиться и выражение *делать вид*, и глагол *понимать*) можно достаточно легко собрать представительный материал такого рода. На основе собранных примеров можно предложить их классификацию, а также проанализировать причины такого на первый взгляд противоречивого коммуникативного поведения (Mustajoki 2006).

Когда мы говорим об использовании корпусов в преподавании языка, первое наблюдение будет относиться к уровню обучения. Поскольку корпуса состоят из материалов естественного употребления языка, тексты и примеры, взятые из них, в среднем достаточно сложны и из-за этого вряд ли могут быть использованы на начальном этапе обучения. Кажется, что самый большой вклад корпуса могут внести в обучении филологов. На Отделении славистики и балтистики Хельсинкского университета накоплен большой

опыт использования корпусов русского языка при написании различных курсовых и дипломных работ. Весьма обширный материал по газетам и журналам в базе данных Интегрум позволяет достаточно легко и быстро проводить анализ типа: встречаемость слова *бренд/брэнд* в современном русском (когда оно появилось, в каких контекстах употребляется, как пишется); метафоры жизни (*жизнь это вода / дорога / океан / повесть / чудо / дар / школа / дар / колесо / игра* и т.д.).

Национальный корпус, позволяющий осуществлять поиск и по грамматическим параметрам, используется нами в духе *problem based learning* (обучение, основывающееся на решении проблем). Это методическое направление практикуется, например, на некоторых медицинских факультетах, где студенты знакомятся с анатомией человека на основе изучения конкретных случаев болезни пациентов. В преподавании иностранных языков разные правила изучаемого языка (скажем, парадигмы слов или структурные схемы предложений) зачастую приводятся преподавателем в готовом виде, без активной умственной работы учащихся. Противоположный, проблемный подход вызывает у самих учащихся желание делать собственные выводы на основе отобранного языкового материала. Так, студентов просят составить правило склонения одушевленных существительных среднего рода в единственном и множественном числах. Поиск можно сделать с помощью Национального корпуса, задав соответствующие грамматические параметры.