

## СРАВНИТЕЛЬНЫЙ АНАЛИЗ ОРГАНИЧЕСКОГО И СТАТИСТИЧЕСКОГО КОРПУСОВ ТЕКСТОВ

Национальные корпуса текстов созданы как инструменты для научных исследований лексики и грамматики языка, а также для предоставления всевозможных справок, относящихся к указанным областям (лексика, грамматика, акцентология, история языка). В обычной жизни любой носитель языка не сталкивается с национальным корпусом и не использует его в своей языковой практике. Но такие корпуса, несомненно, остаются полезными для изучающих язык как иностранный и для исследователей данного языка.

Несмотря на очевидность этой пользы, ее причины и механизм реализации не исследованы в достаточной степени. Предлагается выполнить анализ в указанном направлении в категориях органического и статистического корпусов текстов.

Носители определенного языка также являются членами некоторого дискурсивного сообщества. Впервые о таком сообществе упоминал Дж. Суэйлз, в своей работе в 1990 году. Каждый носитель языка в течение своей жизни порождает два корпуса текстов:

- порожденные тексты (все сказанное или написанное индивидом за срок жизни);
- воспринятые тексты (все фактически услышанные или прочитанные за срок жизни индивида тексты).

Оба таких корпуса связаны и возникают именно в том дискурсивном сообществе или сообществах, участником которых является носитель языка. Данное сочетание корпусов, присущее каждому носителю языка, будем называть *органическим корпусом текстов*.

Если посмотреть на органический корпус с точки зрения онтологии, то мы увидим целостную, антропоцентричную коллекцию текстов. Она будет свободна от любых количественных и статистических категорий, не нуждаясь ни в каких критериях целостности. А идеальным, неограниченным в размерах, национальным корпусом является совокупность органических корпусов текстов всех носителей данного языка.

На основе данной модели сделаем вывод, что органический корпус текстов носителя языка «оказывается замкнутым в пределах определенного дискурсивного сообщества» (А. Ю. Мордовин). Это означает, что для понятия органического корпуса текстов основное значение имеют такие характеристики личности, как ее национальность, профессиональный, социальный, образовательный, возрастной, коммуникативный статус, характер ее межкультурных, межъязыковых и географических контактов. Для воспринимаемых текстов значение имеет фактическая пропорция различных форматов и жанров текстов.

Поэтому можно сделать вывод, что ни один национальный корпус текстов по определению не может в полной мере соответствовать языковому опыту каждого из носителей данного языка. С другой стороны, любой национальный корпус является статистическим, т.е. точки зрения носителя языка – деперсонифицированным. Основываясь на определении А. Ю. Мордовина, укажем, что под «деперсонификацией следует понимать такую предметную и контекстуальную разнесенность текстов, которая, в пределах разумного допущения, не предполагает возможность их «соседства» в органическом корпусе текстов». Однако мы не можем трактовать деперсонификацию корпуса как хаотичное и бессистемное наполнение его текстами различных авторов, поскольку именно непрогнозируемое или только отчасти прогнозируемое восприятие различных авторов в процессе жизнедеятельности индивида является нормальным.

На основе логического анализа можно сделать вывод, что до широкого распространения печати и массовых коммуникаций основными авторами текстов, входящих в органический корпус определенного индивида, всегда были люди из ближайшего окружения, которые также состояли в общем дискурсивном сообществе. Даже нормативные тексты (законы, религиозные тексты) предполагали определенное авторство.

И лишь различные произведения устного народного творчества можно назвать текстами без известного авторства.

Таким образом, по мнению Мордовина, если сравнивать статистический и органический корпусы текстов, следует отметить, «что последний может включать лишь незначительное количество «случайно» прочитанных текстов, формат, тематика и способ размещения которых не соответствуют исходному дискурсивному сообществу индивида; в то время как большинство воспринятых текстов будут соответствовать уместному кругу устного общения и принятых для данного дискурсивного сообщества массмедийных ресурсов».

Вследствие этого наполненность текстами в органическом корпусе, практически определяется «занимаемым местом в дискурсивном сообществе, или соответственным “нарушением” его “границ”» (А. Ю. Мордовин).

Следовательно, существование органического корпуса текстов как совокупности порожденных и воспринятых текстов определенным индивидом – членом дискурсивного сообщества, является эмпирически обоснованным и подтвержденным на практике. И каждый наивный носитель языка сталкивается именно с таким корпусом. Однако такая реальность не отменяет полезность и целесообразность составления национальных корпусов.

Иначе говоря, для носителя языка национальный корпус текстов выступает в качестве справочной системы, которая при всей своей пользе не связана и не обязана быть связанной с фактически существующим органическим корпусом текстов носителя.

Таким образом наиболее оправданными стратегиями и методами при обучении языку являются те из них, которые учитывают, что их содержимое имеет ограничения относительно дискурсивных сообществ обучаемых и целевых иноязычных дискурсивных сообществ.

Показательно в данной связи высказывание А. Ю. Мордовина о важности для лингвистического или филологического исследования языковых/речевых явлений таких параметров корпусов, которые мы называем «статистическими». Именно они позволяют «опосредованно пересекать границы дискурсивных сообществ, обычно труднопреодолимые в естественном языковом опыте и интроспекции, что и обуславливает новизну корпусного подхода».