

А. Н. Баранов

*Институт русского языка им. В. В. Виноградова РАН
(Россия, Москва)*

baranov_anatoly@hotmail.com

Д. О. Добровольский

*Институт русского языка им. В. В. Виноградова РАН
(Россия, Москва)*

dobrovolskij@gmail.com

ПОНЯТИЕ КОРПУСНОЙ МОДЕЛИ ИДИОСТИЛЯ¹

Корпусная модель идиостиля писателя формируется как результат отображения особенностей частотного распределения лексических и синтаксических конструкций в корпусе текстов автора в теоретический конструкт, или, говоря современным языком прикладной лингвистики, онтологию, создаваемую средствами метаязыка лингвистической теории.

Естественно предъявлять некоторые требования к такой модели, ориентируясь на существующие возможности корпусной лингвистики как особого метода описания функционирования языка. В самом общем плане корпусная модель языковой системы или ее подсистемы должна иметь объясняющую и предсказующую силу. Объяснение состоит в выявлении тех закономерностей, которые характеризуют выбранный представительный корпус данных, в нашем случае – корпус текстов. Предсказание реализуется в том, что выявленные закономерности должны распространяться на новые данные, не вошедшие в исходный корпус. В случае идиостиля писателя и объясняющая, и предсказывающая сила оказывается ограниченной. Репрезентативность корпуса текстов конкретного автора достижима далеко не для всех феноменов языковой системы. Если для уровня морфологии и синтаксиса репрезентативность обычно достигается, то уровень лексики и особенно фразеологии требует значительного массива текстов, который может просто отсутствовать для конкретного писателя. Даже если репрезентативность на уровне лексики в авторском корпусе достигается, то предсказывающая сила ограничена тем, что при уже фиксированном корпусе авторских текстов распространение закономерностей на новые тексты уже невозможна.

В проекте, который осуществляется коллективом авторов на базе отдела экспериментальной лексикографии ИРЯ РАН, объектом исследования выступает корпус текстов Ф. М. Достоевского, сформированный на основе полного собрания сочинений под редакцией Г. М. Фридлендера (общий объем корпуса – порядка 2 900 000 словоупотреблений). Тем самым, мы имеем дело с закрытым корпусом, который априори следует считать репрезентативным. Иными словами, наша модель носит скорее объяснительный характер.

В нашем случае корпусная модель идиостиля Достоевского ограничивается лексическим уровнем, включающим систему фразеологизмов. Пока контуры этой онтологии только намечаются. Ясно, однако, что она

¹ Работа выполнена при поддержке РФФИ, грант № 18-012-90025.

должна учитывать существенные частотные отклонения от стандарта – как в сторону увеличения, так и в сторону уменьшения частоты. При этом стандарт должен устанавливаться с учетом фонового корпуса, репрезентативного относительно письменных дискурсивных практик соответствующей эпохи. Тем самым, еще одним объектом исследования оказываются репрезентативные корпуса писателей-современников Достоевского, а именно Л. Н. Толстого, И. А. Гончарова, М. Е. Салтыкова-Щедрина и И. С. Тургенева, тексты которых составили фоновый корпус.

Общий объем фонового корпуса – 6 822 134 словоупотребления. Объем каждого из подкорпусов таков: И. А. Гончаров – 915 497; М. Е. Салтыков-Щедрин – 2 113 573; Л. Н. Толстой – 2 685 209; И. С. Тургенев – 1 490 565.

Основными компонентами корпусной модели идиостиля в данном проекте являются **дискурсивные единицы, квантитативная модель тезауруса идиоматики языка Достоевского и лексические острова**. Проект основан на предположении, что Достоевский был одним из тех авторов-новаторов, которые изменяли стиль русской художественной прозы. Он, наряду с некоторыми другими авторами, широко вводил в авторскую речь и речь персонажей дискурсивные слова, которые использовались для регулирования дискурса и отражали широкий спектр модусной семантики – отношение говорящего и субъекта пропозиции. То, что раньше передавалось совокупностью отдельных пропозиций, концентрировалось в относительно небольшом количестве фраз сложной синтаксической и семантической структуры с единицами типа *кстати, между прочим, одним словом, по крайней мере, по меньшей мере* и под. Особенности употребления дискурсивных слов составляют первый модуль корпусной модели.

Второй модуль – идиомы Достоевского. В нашем исследовании предпринимается попытка сопоставить тезаурус современной идиоматики с теми идиомами, которые представлены в корпусе текстов Достоевского. Выявленные сходства и различия, частота употребления идиом дадут важную информацию об особенностях идиостиля Достоевского.

Наконец, третий модуль – лексические острова. Это особенности употребления Достоевским отдельных слов и фразеологизмов, существенных для его творческого метода и видения мира.