

ВВЕДЕНИЕ

Вопрос о внутренней структуре связующих средств русского языка, в том числе коннекторов, безусловно не является новым. Хорошо известен тот факт, что многие из них состоят более чем из одного графического слова, а многие характеризуются, кроме того, высокой степенью формального варьирования. Также хорошо известны предлагаемые решения — как теоретические, так и лексикографические, которые, однако, решают вопрос о составе связующих средств русского языка, как правило, путем создания их списка, размеры которого могут значительно отличаться: от чуть более 150 в «Русской корпусной грамматике» до почти 700 в «Русской грамматике» 1980 г. Такой подход характеризует, впрочем, и другие лингвистические традиции.

С развитием информационных технологий, созданием корпусов текстов, в том числе параллельных и аннотированных, вопрос о линейных границах неоднословных языковых единиц приобретает особую актуальность. Это объясняется, с одной стороны, необходимостью аннотировать такие единицы, чтобы дать затем возможность пользователю найти их в корпусе текстов, а с другой стороны, необходимостью определить место многочисленных вариантов с общим формантом в системе языка. Например, для коннекторов с формантом *не только* в экспериментальном массиве, сформированном на основе текстов параллельного французского подкорпуса Национального корпуса русского языка (НКРЯ) и использованном в процессе подготовки монографии, насчитывается почти 60 вариантов.

В монографии предлагается новый подход к решению этих проблем, который стал возможен благодаря созданию и развитию нового лингвистического ресурса: надкорпусной базы данных (НБД) коннекторов, в которую были загружены тексты параллельного французского подкорпуса НКРЯ. Объем текстов НБД, составляющий почти 5 млн словоупотреблений, позволяет лингвистам аннотировать употребления коннекторов по заданной ими системе параметров и хранить результаты этой работы в виде структурированных двуязычных аннотаций; НБД позволяет также получить статистику, характеризующую структуру коннекторов и тот или иной аспект их функционирования согласно заданным параметрам.

В предлагаемом нами подходе единицей анализа и аннотирования коннекторов становится речевая реализация, т. е. та форма, в которой коннектор употребляется в конкретном контексте. Это позволяет дать более последовательное, многоаспектное и структурированное описание неоднословных коннекторов, обладающее большей объяснительной силой, чем предлагаемое в рамках традиционной русской грамматики.

Предлагаемому подходу и лингвистическим методам описания коннекторов с применением НБД, разработанному терминологическому аппарату, классификации неоднословных коннекторов посвящена первая глава монографии. В четырех описательных главах данный подход используется для исследования формального варьирования коннекторов как показателей отношения непосредственного предшествования одного события другому, отношения неединственности, а также коннекторов *а то* и *хотя*. Помимо количественного метода анализа, обеспечивающего дополнительные аспекты и новое измерение семантического и структурного анализа коннекторов, в некоторых главах используется также новый контрастивный метод исследования (получивший название унидирекционального, или монофокусного), согласно которому форма, значение и функциональные особенности переводного эквивалента русского коннектора дают дополнительную информацию о его свойствах.

В шестой главе монографии описаны методы и средства информатики для описания структуры неоднословных коннекторов, а также устройство НБД, ее структурные схемы и наполнение полей, статистические функции, используемые в процессе исследования структуры неоднословных коннекторов. Благодаря разработанной системе перекрестных кластеров лингвисты могут эффективно анализировать и многоаспектно описывать неоднословные коннекторы, их состав, сочетаемость составляющих их компонентов и элементов, расширяя в случае необходимости число исследуемых аспектов.

Заключительная, седьмая, глава монографии показывает, что статистические данные о структуре коннекторов могут стать удобным стилеметрическим инструментом при определении авторства текста. Для оценки близости между текстом X и текстами претендентов на авторство используется сумма абсолютных величин разностей между частотами в исследуемом тексте X и средней частотой по данному коннектору в текстах претендентов. В подсчете участвуют около 300 наиболее частотных коннекторов, статистические данные о которых приводятся в приложении.

Предлагаемая читателю монография может быть интересна не только специалистам по корпусной и компьютерной лингвистике. В силу поставленных в ней вопросов и предлагаемых решений она представляет, на наш взгляд, интерес также для специалистов в области теоретической грамматики и лексикографии служебных слов и открывает новые пути исследования неоднословных языковых единиц.

Ольга Инькова