

ИДЕНТИФИКАЦИЯ ПАРАДИГМАТИКИ И СИНТАГМАТИКИ В КОРПУСНО-ОРИЕНТИРОВАННОЙ ЛИНГВИСТИКЕ.

<https://doi.org/10.5281/zenodo.7803302>

Исаева Зера Гаировна

zeraisayeva@mail.ru

Преподаватель филологического факультета ФерГУ, направление русской филологии.

Аннотация.

Лингвистические корпуса, предназначенные для работы с объёмами текстов, извлекая из них нужную информацию, позволили создать коллекцию текстов, специально отобранных, размеченных по различным лингвистическим параметрам и обеспеченных системой поиска во всех аспектах языкового яруса. Увеличение объема материала исследования потребовало применения новых методов анализа лингвистических данных, включая их статистическую обработку.

Ключевые слова.

корпусная лингвистика, функционально-коммуникативная парадигма, корпусное описание, конкорданс.

Изучая продукт функционирования языка и обращая внимание, главным образом, на построение речи в процессе общения возникает такое направление, как корпусная лингвистика, являясь естественным продолжением функционально-коммуникативной парадигмы. Всем известен тот факт, что взаимодействие языковых составляющих, формы которого предопределены как системой и нормой языка, так и определенными условиями коммуникации – это успешное образование языкового общения.

По словам Е. С. Кубряковой, «современное состояние теоретической лингвистики характеризуется выдвиганием в ней двух главных парадигм научного знания - когнитивной и коммуникативной» [1:12]. В этой связи вполне естественно, что, будучи одной из главных парадигм научного знания в современном языкознании, функционально-коммуникативная парадигма вызвала кардинальное изменение в области теоретической лингвистики и дала толчок пересмотру многих фундаментальных понятий с новых позиций и в новом свете [1:10].

Идеология корпусной лингвистики позволяет сделать и более сильное утверждение: при прочих равных условиях, должно быть предпочтительнее такое описание языка, которое будет ориентировано на явления, лучше представленные в корпусе. Иными словами, то, что является центральным в корпусе, должно быть центральным и в грамматике. Плодотворность для теории языка описания механизмов выражения таких смыслов, которые не являются востребованными реальными говорящими на данном языке, как минимум неочевидна. Как кажется, по крайней мере полезнее вначале описать способы выражения тех смыслов, которые для данного языкового коллектива являются приоритетными, - то есть те, которые легко найти в корпусе данного языка. Смысл корпусно-ориентированной лингвистики в том, что она позволяет изучать действительно существующие в языке, а не мнимые явления.

Сравнительно консервативная область языка, как морфология, демонстрирует достаточно много существенных отличий от той картины, которая представлена в нормативных описаниях. Ведь корпусное описание (в том числе и в силу необходимости обеспечить корректный автоматический анализ текста) не может уклониться от того, чтобы анализировать все встречающиеся в текстах формы. А таких форм заведомо больше, чем это предписывается нормативной грамматикой. Причем, что существенно, в число примеров русского «морфологического расширения» попадают далеко не только случайные ошибки или окказиональные употребления - это и частотные разговорные формы, пока (почти) не находящие отражения в грамматиках, это и частые случаи языковой игры, также имеющие свою «грамматику» и свои регулярные механизмы, это и такие устаревшие, диалектные и просторечные формы, которые, несмотря на исключение их из нормативных грамматик, продолжают существовать в языке, известны говорящим и могут употребляться в соответствующих контекстах.

Но в то же время реально встречающихся в текстах форм в каком-то смысле и меньше, чем в нормативных источниках. Существующие системы автоматического морфологического анализа, как правило, не ориентированы на частотность той или иной формы: если форма в принципе возможна, она присутствует в автоматическом словаре. Следствием такого подхода оказывается то, что порождается немалое количество форм, которые, хотя и возможны по правилам русской морфологии, в реальных текстах никогда не встречаются. Это «гиперпорождение» не всегда безобидно: специалистам по

автоматическому морфологическому анализу хорошо известен эффект паразитической омонимии, затрудняющий работу программ; этот эффект возникает в тех случаях, когда у распространенной словоформы появляется мифический двойник, формально законный, но реально в текстах отсутствующий. К таковым, например, относятся разбор предложения для как деепричастия от глагола длить или разбор формы презенса сеем как краткой формы страдательного причастия сеемый (даже полная форма этого причастия встречается в корпусе всего один раз); проблематично и порождение форм множественного числа у слов типа современность, и мн. др.

Таким образом, морфология, отражающая данные корпуса, одновременно и шире, и уже «стандартной» - она включает в себя в качестве легитимных многие не предусматриваемые стандартной грамматикой варианты, но зато исключает надежно прописанные в стандартной грамматике «потенциальные» (т. е. возможные лишь теоретически) формы.

Из числа наиболее интересных морфологических явлений, существующих в современных русских текстах, но недостаточно отражаемых нормативной грамматикой (или вовсе не отражаемых), можно упомянуть о «стяженных форм» винительного и дательного падежей у личных местоимений я и ты (мя, тя, те), по своим морфосинтаксическим свойствам отчасти напоминающих древнерусские местоименные энклитики; такие формы (существовавшие в диалектах и в просторечии) не только надежно фиксируются в устной речи всех слоев носителей русского языка, но и начинают в заметном количестве проникать в некодифицированную письменную речь (особенно в тексты электронной коммуникации); к этому же классу примыкает и частотная форма именительного и винительного падежа вопросительного местоимения чё. Также можно сказать и о существовании достаточно большого числа нестандартных форм деепричастий на -а / -я(типа положе) и на -ши (типа поживши); к ним добавляются заимствованные из диалектов, но получающие всё большее распространение варианты на -мши (типа выпимши, соврамши, не спамши и т. п.).

В работе [2:2] показана особая типичность «усеченных прилагательных» для языка русской поэзии практически во все периоды ее существования, вплоть до современной поэзии, где сочетания типа пластмассовы цветочки в

определенных типах текстов и у определенных авторов оказываются вполне обычны.

Рассмотрим еще один из элементов, представляющий особый интерес в изучении корпусной лингвистики как идентификация парадигматики – это конкорданс. Конкорданс – это список всех употреблений определенного языкового выражения в контексте и с ссылкой на источник при необходимости. В данной коннотации этот термин часто применяется в корпусной лингвистике. В исходном своем толковании этот термин также используется для обозначения списка ключевых слов книги или работы, градированных в алфавитном порядке, с их контекстами [4:3].

Первый конкорданс, к Вульгате, был создан в 13 веке Хьюго де Сент-Шером, которому помогали 500 монахов. Из-за огромного объема работы по составлению конкордансов в те времена, конкордансы были созданы для трудов, которые имели высокую важность – Вед, Библии, Корана или произведений Уильяма Шекспира.

Конкордансы, также, часто используются в прикладной лингвистике: при переводе, в лексикографии, при обучении и изучении языка, при анализе текста [5: 315].

Конкордансы применяются для решения следующих лингвистических задач:

- сопоставлять разные использования одного слова;
- создавать списки слов;
- искать перевод различных словосочетаний или слов;
- искать и исследовать фразы и идиомы;
- анализировать ключевые слова;
- анализировать частоту употребления какого-либо слова и словосочетания.

При изучении текста можно выделить поисковую, эвристическую, аналитическую функции конкорданса, а также функции индексации и сравнения.

Поисковая функция позволяет быстро находить нужный фрагмент текста, используя заданное слово или словосочетание [6:229].

Данная функция конкорданса используется при подборе цитат и их сверке, при изучении особенностей текста оригинала. Конкорданс может быть использован составителями хрестоматий. Он позволяет искать фразы и идиомы в тексте, осуществлять поиск терминов и идиом в памяти переводов.

Эвристическая функция конкорданса отличает его от именных указателей наличием контекста. Зачастую, контексты позволяют лингвисту или простому читателю увидеть новое трактование текста.

Аналитическая функция позволяет проводить анализ различных языковых показателей, таких как лексемы, ключевые слова, частота их употребления в тексте и т.д.

Функция сравнения применяется при сравнении в тексте всевозможных коннотаций и употреблений слова.

В отличие от словаря, который ставит во главу угла словарную статью, конкорданс приводит примеры – контексты словоупотреблений.

Конкорданс являет собой фундамент для созданий разнообразных дифференциальных, в том числе и толковых словарей.

Итак, корпусно-ориентированная лингвистика вырабатывает в себе некоторые универсальные принципы в развитии различных технологий обработки текстовой информации, создавая специальные инструменты для работы с корпусом на любом языке и при выработке универсальных стандартов и технологий при использовании массивов языковых данных, предназначенных для лингвистических исследований в интересах широкого круга пользователей.

СПИСОК ЛИТЕРАТУРЫ:

1. Вестник ВГУ, серия лингвистика и межкультурная коммуникация, 2001, в. 14 теория языка Е.С. Кубрякова о когнитивной лингвистике и семантике термина «когнитивный».

2. Кулева А.С. «Усеченные прилагательные в языке русской поэзии» - Кулева А.С.. -2008.

3. Нагель Ольга Васильевна Корпусная лингвистика и ее использование в компьютеризированном языковом обучении // Язык и культура. 2008. №4 С.53-59.

4. Бобунова Мария Александровна Конкорданс XXI В.: новая старая форма // Вопр. лексикографии. 2016. №2 (10) С.41-54.

5. Якубовский К.И., якубовская К.А. Обзор современных лингвистических технологий и систем // Вестник МГУП. 2015. №2 С.315-319.

6.Труфанова Мария Владимировна «Применение цифрового словаря с корпусом для работы над лингвистическими проектами по немецкому языку // Социально-экономические явления и процессы.» 2011. №8 С.229-232.