К ОПЫТУ ФОРМИРОВАНИЯ АННОТИРОВАННОЙ БАЗЫ ДАННЫХ В СОЦИАЛЬНО-СЕТЕВОМ ДИСКУРСЕ ИНТЕРНЕТА

Р.К. Потапова, Л.Р. Комалова
Московский государственный лингвистический университет
Москва

На фоне активных политических и геополитических преобразований конца XX – начала XXI веков возникает потребность в разработке действенных механизмов распознавания агрессивного (в том числе речевого) поведения.

В рамках исследования семантического поля «агрессия» в фокусе внимания находятся средства и механизмы реализации и побуждения негативных эмоций и переживаний реципиента, в том числе, продуцируемые в ходе взаимодействия в социальных сетях в мировой сети Интернет. Анализ содержания социально-сетевого дискурса (по Р.К. Потаповой [3]) подтверждает предположение о расширении семантического поля «агрессия» и о снижении порога критичности восприятия (см., например, [2, 7]).

Доступность текстов в среде Интернет и современные программные разработки позволяют проводить мониторинг интернет-контента и выявлять тексты, провоцирующие агрессивное поведение. Весьма актуальным и перспективным является разработка баз данных и баз знаний, позволяющих кластеризировать вербальные средства актуализации деструктивного (в том числе речевого) поведения в совокупности с возможностью оценки их воздействующей силы, в том числе потенциала побуждения реципиента к разным формам агрессии (см., например, [1, 5]).

Разрабатываемая база данных представляет собой совокупность особым образом аннотированных самостоятельных дискурсивных единиц, которые являются полилогическими единствами (полилогами) с выделением опорно-тематических блоков (тем, относящихся к семантическому полю «Политические преобразования»). Единицей базы данных выступает аннотированное полилогическое единство (АПЕ).

Специфической особенностью полилогических единиц описываемой БД является их прагматическая составляющая, репрезентирующая негативную прагматическую установку коммуникантов (участников общения) по отношению друг к другу, к внешней среде, событиям, связанным с политическими преобразованиями в стране проживания, в других странах, а также в мире в целом. Основными маркерами воздействующего дискурса в данном случае являются оценочные лексические единицы и неологизмы, которые в полилогах размечены при помощи тэгов.

Речевой материал для составления базы данных набирался методом сплошной выборки из социально-сетевого дискурса (по Р.К. Потаповой) за временной период: октябрь 2015г. – март 2016г. Полилоги (в авторской редакции) представлены на русском языке с элементами речи на украинском языке. В качестве коммуникантов выступают русскоговорящие пользователи таких открытых социальных сетей как «ВКонтакте» и «Livejournal», а также пользователи, комментирующие новостные сообщения на сайтах информационных агентств и сетевых СМИ.

Аннотированная база данных формировалась в рамках экспериментального исследования по определению негативной прагматической установки коммуникантов в процессе обмена мнениями, в котором участвовали 30 испытуемых — носителей русского языка, дифференцированных по гендерному и возрастному цензам (21 женщина, 9 мужчин; в возрасте 20-23 лет).

База данных разрабатывается как открытая система, которую можно пополнять, а также расширять список параметров аннотации. В настоящее время база данных включает 937 аннотированных русскоязычных полилогических единств на тему «Политические преобразования».

Параметры аннотирования БД разрабатывались на основе предшествующего опыта формирования аннотированных полнотекстовых корпусов на русско-, англо- и испаноязычном материале текстов сетевых СМИ, в содержании которых присутствуют образцы семантического поля «агрессия» [4, 6, 8].

Изменение типа единицы базы данных с авторского монологического текста (подготовленный письменный дискурс, композиционно организованный в законченное целое) на полилогическое единство социально-сетевого дискурса (ССД) потребовало преобразования структуры аннотации. Несмотря на то, что инициальный пост (условная реплика) задает тему дискуссии, наличие большого количества участников (коммуникантов) размывает рамки первоначальной темы и придает полилогу характер политематического. При этом в зависимости от активности определенных участников дискуссии центр обсуждения переносится к тому или иному тематическому пику. В связи с этим в аннотации был снят блок, предлагающий выбор из перечня стереотипных тем коммуникации (сформулирован в ходе предшествующего экспериментального исследования), и введено два окна, в которых оператор формулирует а) ведущую тему полилога, б) сопутствующие(ую) темы(у). В аннотации также был снят блок количественных данных в отношении числа словоформ и графем в полилоге и введен блок квантитативных сведений об участниках полилога.

По сравнению с аннотацией монотекста сетевых СМИ аннотация ССД-полилога характеризуется репрезентацией особенностей речевой продукции так называемого коллективного автора и обращенностью к двум фокусам суггестии: внутреннему — воздействие на непосредственных участников коммуникации (включенный реципиент) и внешнему — воздействие на «случайных» наблюдателей (сторонний реципиент). В связи с этим в аннотации АПЕ особое внимание уделяется оценке воздействующей силы полилога (аффективная и прагматическая компоненты) на реципиента, оценке непосредственных участников полилога по их речевой активности, а также в аннотацию вводится блок сведений о стороннем наблюдателе, в роли которого выступают испытуемые.

В аннотации каждой АПЕ учитываются следующие сведения: сведения об источнике речевого материала; определение темы полилога; оценка воздействующей силы полилогического единства (на основе субъективных оценок группы испытуемых); сведения об испытуемом (аннотаторе); сведения об участниках полилога (гипотетические данные на основе восприятия испытуемым текста полилога и параметрических данных, указанных коммуникантом в своей учетной записи в социальной сети). Каждому аннотированному полилогическому единству присвоен специальный идентификационный номер, который указан в аннотации. При нажатии кнопкой мыши на данный номер в новом окне открывается word-файл с АПЕ, в котором помимо краткой аннотации (указаны формальные сведения) приводится полный исходный текст «ветки» сообщений за указанный период.

Текст полилога размечен при помощи специальных тэгов, позволяющих идентифицировать текстовые фрагменты, которые испытуемые воспринимали как репрезентацию прагматической и эмоциональной видов установки коммуникантов на конфронтацию. Тэгами отмечаются вербальные фрагменты полилога, в которых коммуниканты используют оценочные лексические единицы (_Л _*Л) и неологизмы (_Н _*Н).

При желании оператор имеет возможность вносить корректировку в разметку и аннотацию каждой АПЕ и сохранять внесенные правки.

База данных применима на практике для проведения научных исследований оценочной лексики и синтаксиса, направленных на воздействие с учетом эмоционального состояния как участника коммуникации, так и стороннего наблюдателя посредством сообщений в социально-сетевом дискурсе в сети Интернет. База данных может быть использована для проведения тематического лексико-морфолого-синтаксического анализа с целью выявления структуры текстов, порождающих отрицательные эмоциональные состояния испытуемого, а также в учебных целях в составе учебно-методических комплексов дисциплин «Лингвоконфликтология», «Лингвокриминалистическая экспертиза текста», «Фундаментальное и прикладное речеведение», «Социолингвистика», «Новые информационные технологии и лингвистика».

ЛИТЕРАТУРА

1. Комалова Л.Р. Агрессогенный дискурс: типология мультилингвальной вербализации агрессии. М.: Спутник+, 2017.
2. Кошкарова Н.Н. Конфликтный дискурс: психологические и лингвистические аспекты. Челябинск: ЮУрГУ, 2008.
3. Потапова Р.К. Социально-сетевой дискурс как объект междисциплинарного исследования // Дискурс как социальная деятельность: приоритеты и перспективы: материалы Второй международ. научн. конф. (Москва, 16 – 17 октября 2014 г.). М.: МГЛУ, 2014. С. 20 – 22.
4. Потапова Р.К., Комалова Л.Р. База данных русскоязычных текстов, содержащая единицы семантического поля «агрессия» // Вестн. Моск. гос. лингвист. ун-та. 2014. № 19 (705). Часть I. С. 112 – 121.
5. Потапова Р.К., Комалова Л.Р. Вербальная структура коммуникативного акта агрессии: тематический толковый словарь. М.: ИНИОН РАН, 2015. Вып. 1..
6. Потапова Р.К., Комалова Л.Р. Формирование аннотированных баз данных семантического поля «агрессия» (на материале русскоязычных и англоязычных электронных интернет-источников) // Писменото наследство и информационните технологии: материали от V Международна науч. конф. (Варна, 20 – 24 сентября 2014 г.). София; Ижевск, 2014. С. 68 – 72.
7. Шаховский В.И. Диссонанс экологичности в коммуникативном круге: человек, язык, эмоции. Волгоград: изд-во ИП Поликарпов И.Л., 2016.
8. Potapova R., Komalova L. On principles of annotated databases of the semantic field “aggression” // A. Ronzhin, R. Potapova, V. Delič (Eds.): SPECOM 2014, LNAI 8773. Springer International Publishing Switzerland, 2014. P. 322 – 328.