

УДК 81

# ПОСТРОЕНИЕ РЕЧЕВОГО ПОРТРЕТА УЧАСТНИКА ЭКСТРЕМИСТСКОГО ФОРУМА С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ КОРПУСНОЙ ЛИНГВИСТИКИ<sup>1</sup>

**ЛИТВИНОВА Татьяна Александровна,**

кандидат филологических наук, научный сотрудник  
Регионального центра русского языка;

**ЛИТВИНОВА Ольга Александровна,**

младший научный сотрудник научно-исследовательской лаборатории корпусной идиолектологии,  
Воронежский государственный педагогический университет

**АННОТАЦИЯ.** Статья посвящена актуальной проблеме речевого портретирования социальной группы и возможностям применения современных методов корпусной лингвистики. В статье проводится сопоставительное исследование корпусов текстов лиц – участников экстремистского форума «КавказЧат» и комментариев интернет-пользователей на одну и ту же тему – теракт в московском метро 2010 года. С использованием комплекса программ WordSmith Tool выявляются особенности речевого портрета участника экстремистского форума.

**КЛЮЧЕВЫЕ СЛОВА:** речевой портрет, интернет-форум, экстремистский текст, корпус текстов, корпусная лингвистика.

**BUILDING A SPEECH PORTRAIT OF AN EXTREMIST FORUM PARTICIPANT USING CORPUS LINGUISTIC METHODS**

**LITVINOVA T. A.,**

Cand. Philolog Sci., Researcher of the Regional Centre of Russian Language;

**LITVINOVA O. A.,**

Junior Researcher of the Corpus Idiolectology Laboratory,  
Voronezh State Pedagogical University

**ABSTRACT.** The article is devoted to the current problem of a social group speech portraying and to the possibilities of applying modern methods of corpus linguistics. The article conducts the comparative study of the corpus of texts of persons participating in the extremist forum “KavkazChat” and the comments of Internet users on the same topic, namely, the terrorist attack in the Moscow metro 2010. With the use of the WordSmith Tool Program, the features of the speech portrait of an extremist forum participant are identified.

**KEY WORDS:** speech portrait, Internet forum, extremist text, corpus of texts, corpus linguistics.

**В**ведение. В экстремистской и террористической деятельности интернет-ресурсы, в том числе блоги и форумы, активно используются для агитации и пропаганды соответствующей идеологии, причем есть все основания полагать, что этот процесс будет развиваться и далее [1]. В связи с этим особую актуальность приобретают исследования, направленные на лингвистический анализ и автоматическое распознавание в Сети подобного контента (см. обзор в работах [4; 5]).

Следует отметить, однако, что интерес исследователей в основном направлен на выявление опасного контента, а не экстремистки настроенных авторов. Практически не разработанной остается проблема построения коллективного речевого портрета участника экстремистского форума. Как отмечает О.В. Загоровская [2, с. 25], «в настоящее время работы в области речевого портретирования носят, в основном, аспектный характер, т.е. предполагают ориентацию на изучение и описание какого-либо отдельного или нескольких отдельных аспектов

языковой личности представителя той или иной социальной группы».

Целью нашего исследования является выявление особенностей речевого портрета участника экстремистского форума на основе частотного анализа лексики с использованием методов корпусной лингвистики. Материалом исследования является представительный корпус текстов, собранных с интернет-ресурса [www.kavkazchat.com](http://www.kavkazchat.com). В настоящее время данный интернет-ресурс находится в федеральном списке экстремистских материалов, опубликованном на официальном сайте Министерства юстиции РФ; доступ к данному ресурсу запрещен. Доступ к ресурсу производился через The Dark Web Portal Project [7].

**Речевое портретирование участника экстремистского форума.** Для целей текущего исследования нами был выбран раздел форума «Взрывы в Москве», посвященный обсуждению терактов в московском метро в марте 2010 года (<http://www.kavkazchat.com/showthread.php?t=34800>), совершенных двумя террористками-смертницами кавказского происхождения. В результате взрывов погибли 41 и ранены 88 человек. Ответственность за взрывы взял на себя лидер «Кавказского эмирата» Доку Умаров.

Временной интервал сообщений на форуме – 29.03.2010 (день совершения теракта) – 17 апреля

<sup>1</sup> Исследование выполнено при поддержке гранта Президента РФ для молодых ученых – кандидатов и докторов наук МК-5718.2018.6 «Речевой портрет экстремиста: корпусно-статистическое исследование (на материале экстремистского форума «Кавказчат»)».

© Литвинова Т.А., Литвинова О.А., 2019

Информация для связи с авторами: [centr\\_rus\\_yaz@mail.ru](mailto:centr_rus_yaz@mail.ru)

2010 (одно сообщение), обсуждения начались в день теракта (42 сообщения) 29, продолжились 30 марта (11 сообщений), далее последовал спад в связи с временным закрытием возможности регистрации пользователей в связи с их массовым наплывом и проблемами в работе сайта. Активное обсуждение возобновилось с 03.04 и шло по 08.04, после чего наблюдается уменьшение интереса к обсуждению (с 9 по 17 апреля – всего 4 сообщения). Число пользователей, участвующих в обсуждении темы, – 67. Число комментариев от пользователей варьировало от 1 до 46, общее число комментариев – 466. Назовем данный корпус текстов корпусом текстов участников экстремистского форума (КУЭФ).

Данный раздел выбран прежде всего из-за резонансности обсуждаемой темы в медиапространстве. Важность анализа экстремистской риторики непосредственно после совершения теракта обусловлена также тем, что после совершения резонансного террористического акта не только изменяется характер террористической пропаганды, но в этот период она начинает вестись с максимальной активностью [6, с. 27]. Отчасти это связано с тем, что общество в данный момент времени становится наиболее восприимчиво к материалам и информации, связанной с произошедшими террористическими актами.

В это время посещаемость террористических Интернет-ресурсов лицами, не относящимися к их постоянной аудитории, резко возрастает [6, с. 27]. В указанном источнике приведены данные о резком увеличении числа запросов «Кавказ Центр», выполненных в поисковой системе «Яндекс» в период после совершения террористического акта. Особенную обеспокоенность вызывает тенденция, согласно которой после такого рода обсуждений форум пополняется новыми пользователями, которые начинают разделять экстремистскую идеологию, а значит, цель пропаганды можно считать достигнутой [6, с. 27].

Участники форума в указанной теме открыто выражают одобрение терактов: «*Очень надеюсь, что это наши*». <> «*Судя по официальным ролям кафилов, так оно есть*». <> «*Дай Аллах чтобы была польза от таких спецопераций! И чтобы они (спецоперации) продолжались во благо Исламу и Мусульманам! Аллаху Акбар!*» <> «*Я рад что шахидки атаквали метро в Москве, ЭТО ЗА МОЕГО ВОШИ И ДАДУ!!!! За все оборвавшие жизни мусульман!*» Также высказываются угрозы, и делаются попытки перенести ответственность за случившееся на русский народ: «*Иниша Аллах в каждый их дом должен зайти страх и ужас и только тогда они поймут, что это такое*» <> «*Иниша Аллах то что было в Москве это еще игрушки, не придет и времени как вы получите снова с полна за свои преступления на Кавказе!*» <> «*Вас будут взрывать, убивать и использовать в своих целях, так как большего вы не заслуживаете пока не начнете уважать хотя бы самих себя и не оставите свою преступную власть от продолжения убийств на Кавказе*».

Нужно отметить, что на интернет-форуме активно навязывается мнение о том, что террористические акты допустимы в соответствии с религиозными канонами. Одновременно удаляются сообщения, в которых подвергаются сомнению псевдорелигиозные идеи, оправдывающие терроризм: «*Предупреждаем! Попытки ввести в заблуждение мусульман говоря, что эти действия [террористические акты] недопустимы с точки зрения Шариа, бу-*

*дут пресекаться! ...Мы даём возможность высказать своё мнение здесь, но вводить людей в заблуждение, давая фатвы от учёных-пораженцев мы здесь не позволим...*»

Указанное обсуждение привлекло внимание общественности и само стало предметом обсуждения в Сети, ср. типичные реакции интернет-пользователей (из обсуждений на новостном ресурсе the village): *у меня челюсть на пол упала. я в шоке; сами недолуды здесь* (на форуме Кавказчат – авт.) *все расскажут раньше; о, боги! что они там пишут, они одержимы* и т.д.

В качестве материала для сравнительного анализа текстов участников мы составили корпус текстов из двух источников. Первый – комментарии читателей интернет-издания the village к материалу на тему «В московском метро прогремели взрывы» <http://www.the-village.ru/village/people/people/88720-v-moskovskom-metro-proizoshel-vzryv#comments>. Комментарии размещались с 29 по 31 марта. Общее число комментариев – 816. Объем корпуса – 127045 знаков без пробелов, или 22641 слов. Число участников обсуждения – 62.

В качестве второго источника материала для сравнения мы использовали комментарии к материалу <https://www.svoboda.org/a/1996131.html>, всего 497 комментариев. Общий объем текстового материала – 134829 знаков без пробелов, или 20237 слов. Комментарии велись с 30.03 по 06.04.2010. Число участников обсуждения – 81.

Тексты из двух указанных источников были слиты в один и составили корпус для сравнения (далее – КдС), в текстах которого выражено отношение массовых пользователей к терактам 2010 года.

Материалы корпусов подвергались предобработке. Были удалены все метаданные (дата поста, имя автора и др.). Из самих текстов были удалены цитаты в том случае, если они представляли собой предыдущие сообщения пользователей, на которые отвечает данный пользователь. Цитаты из СМИ и др. не удалялись. В текстах были исправлены опечатки. Язык текстов корпуса – русский. Материалы экстремистского форума содержали вкрапления написанных на кириллице чеченских слов, такие вкрапления не удалялись. Все тексты были сохранены в формате txt.

Анализируемые корпуса текстов (КУЭС и КдС) являются сопоставимыми по объему, посвящены одной теме (обсуждению взрывов в метро), созданы непосредственно после обсуждаемого события, что делает их удобным материалом для сравнительного анализа с использованием корпусных средств. Корпусы различаются, однако, средней длиной постов (108 слов в КУЭФ vs. 32 в КдС), а также числом лиц, участвующих в комментариях. На наш взгляд, различия в длине текстов обусловлены особенностями самого экстремистского форума, пользователи которого активно ведут пропаганду, обосновывая совершенные смертницами теракты со ссылками на религиозную литературу и т.д. Различия в числе авторов могут быть вызваны трудностями с регистрацией на анализируемом форуме.

Нами была проведена серия экспериментов на материале корпуса «Взрывы в метро». В качестве основного средства анализа текстов нами был использован Oxford WordSmith Tools 4.0. Oxford WordSmith Tools – набор программных средств, пользующихся популярностью среди исследователей в области компьютерной лингвистики и став-

ший золотым стандартом в данной области. Однако среди российских лингвистов данное ПО менее известно. Мы использовали версию 4.0, так как она является бесплатной и в то же время ее функционал достаточен для достижения целей настоящего исследования.

Наиболее важными модулями программного обеспечения являются:

**1. Wordlist:** модуль позволяет пользователю создавать частотные списки слов на основе анализа одного или нескольких файлов. Также полученные списки могут быть использованы для вычисления так называемой меры статистической сочетаемости (меры ассоциации) языковых единиц 'mutual information score' (MI), которая показывает силу синтагматической связи между словами. Мера MI сравнивает зависимые контекстно-связанные частоты с независимыми, как если бы слова появлялись в тексте совершенно случайно [3, с. 78]. Если значение MI больше определенного значения (для русского языка часто называется значением 3 и более), тогда рассматриваемое сочетание слов можно считать статистически значимым.

**2. Concord:** позволяет пользователю создавать конкордансы для определенных слов и словосочетаний. При помощи этого модуля можно изучать информацию, касающуюся колокаций, кластеров слов и других лексических паттернов, характерных для анализируемого текста, а также изучать распределения тех или иных слов в тексте по его длине (word plots).

**3. Key words:** позволяет пользователю выделять ключевые слова текста, т.е. слова, встречающиеся относительно часто по сравнению с неким стандартом, т.е. корпусом общего употребления (им может быть как национальный корпус, так и отдельные, созданные исследователем корпуса – в зависимости от целей исследования).

В наших экспериментах мы использовали первые два модуля, Wordlist и Concord.

**Wordlist.** Мы построили два частотных списка, отдельно для каждого корпуса текстов. Лемматизация текста, а также исключение стоп-слов не проводились (для указанных целей в текущей версии программы необходимо сформировать список лемм и список стоп-слов). Затем мы провели сравнение данных списков, используя команду Comparing Wordlists. С использованием этой команды происходит сравнение слов в двух частотных списках, в результате чего формируется список слов, встречающихся чаще в одном из корпусов (статистически значимо), при этом можно установить минимальный порог вхождения слова в корпус. В качестве статистического теста мы использовали log-likelihood. Минимальная частота словоформы, заданная нами, составляла 10, т.е. мы анализировали только словоформы, которые встречаются более чем 10 раз как минимум в одном из корпусов.

На рис. 1 представлены частотности слов в корпусах текстов, их процент от общего числа слов корпуса, keyness (показатель, отражающий статистическую значимость ключевого слова в данном корпусе; чем он выше (по модулю), тем значимее его выделение в качестве частотного; на рис. словоформы приведены в порядке их keyness, т.е. чем выше словоформа в списке, тем более оно характерно для КУЭФ), а также уровень значимости статистического теста (все  $p < 0.000001$ ). Красным выделены слова, характерные для КдС, но не характерные для КУЭФ, их значение keyness отрицательно.

Анализ полученных данных, а также данных частотных списков (команда Consistency Analysis (Detailed)) позволяет сделать некоторые наблюдения.

Участники экстремистского форума используют в своих постах риторику противопоставления (*мы – вы, они; мусульмане, чеченцы – русские, русня, кафиры*), конструируя т.н. дискурс войны (частотные лексемы *война, воевать, враг, земля, территория, месть, оружие, против, убийство, убитые, шариат, смерть*) на религиозной основе (*аллах, ислам, религия*). Для КдС, в свою очередь, характерен дискурс трагедии и страха (*паника, жертва, погибшие, пострадавшие, страхи, страшно, трагедия*), а также обсуждение темы власти (*власти, ФСБ, Путин*) и причин происшедшего (*выгодно*).

Анализ текстов КУЭФ позволил сделать еще одно наблюдение. В них практически (так как мы работаем только словами с частотностью свыше 10) отсутствуют леммы *террор, теракт(терракт)* (лемматизация была проведена вручную), значительно меньше частотность леммы *террорист*. Для более детального анализа употребления словоформы «Террористы» мы воспользовались инструментом модуля Wordlist Compute Concordance (рис. 2-3).

Более детальный анализ конкордансов позволяет сделать вывод о том, что данная лексема воспринимается негативно участниками экстремистского форума: *«с точки зрения русских кафиры моджахеда – террористы», «тогда и вправду мы террористы», «начинают кричать, что террористы взорвали невинных», «смотрите, да они варвары, террористы»*. В КдС *террористы* – нейтральное слово, используемое для обозначения совершивших теракт: *«зачем террористы это делают», «мусульманские террористы», «террористы мечтают разжечь», «реально хотят напугать»* и т.д.

Проведенный нами с использованием инструмента модуля Wordlist Word Clusters кластерный анализ (cluster size = 3, min frequency = 5) показал (рис. 4), что среди самых частотных триграмм (то есть последовательностей из трех элементов) слов обоих корпусов является триграмма *«на самом деле»*.

При этом в КУЭФ преобладают триграммы, объединенные темой «Религия» (*благословит его аллах, на пути аллаха, жить по Корану, и его посланник, нет греха*), а в КдС – триграммы, обозначающие локацию (*в московском метро, на парке культуры – станция метро, где произошел взрыв, в нашей стране, ср. в КУЭФ – на своей земле*), а также триграмма *«кто это сделал»* и триграммы, связанные с обсуждением действий президента В. Путина (*Путин должен уйти – прецедентное выражение, аллюзия на одноименную общественную кампанию по сбору подписей под обращением к гражданам России с требованием отставки президента РФ В. Путина, начатую в Интернете 10/03/2010 активистами ряда оппозиционных общественных организаций, а также несколькими известными деятелями культуры России; мочить в сортире – выражение, публично использованное В. В. Путиным, на тот момент Председателем Правительства Российской Федерации, 24/09/1999 во время пресс-конференции в Астане, когда он комментировал события бомбардировки российской авиацией Грозного*).

Как мы видим, данные, полученные с использованием корпусной лингвистики, подтверждают наблюдения исследователей, согласно которым в период после совершения терактов общество особенно восприимчиво к критике действующей власти, и этим могут воспользоваться силы, заинтересованные в дестабилизации ситуации в стране [6].

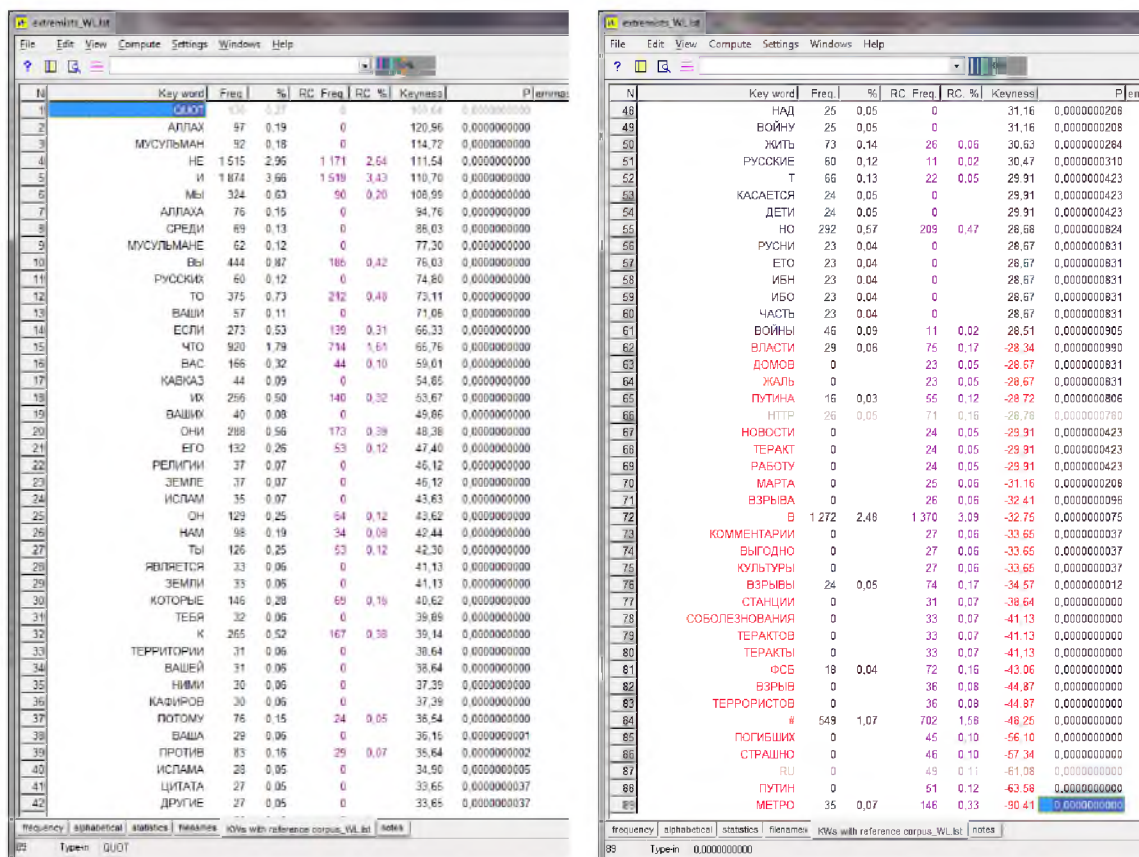


Рис. 1 – Частотные списки анализируемых корпусов текстов:  
а – КУЭФ, б – КдС

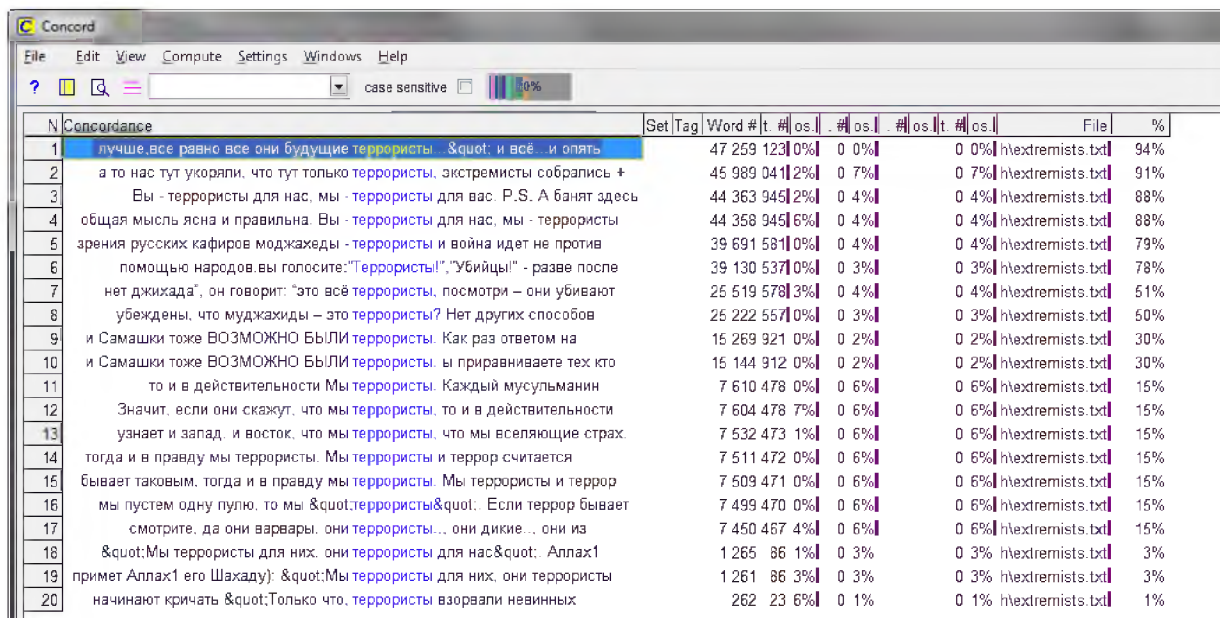


Рис. 2 – Конкорданс словоформы «Террористы» в КУЭФ



N	Concordance	Set	Tag	Word #	t. #	os. #	os. #	os. #	os. #	File	%
1	что это проло фанатики а не террористы. Что касемо МВД и прачик	41	585	155	0%	0	0%	0	0%	ence corpus.txt	97%
2	хаоса не заданя вопросом, почему террористы использовали такая малая	41	545	153	6%	0	0%	0	0%	ence corpus.txt	97%
3	Так что либо все давайте ночью пока террористы спят, мы свалим и	39	231	023	1%	0	4%	0	4%	ence corpus.txt	92%
4	этих уродов. И насчет версии что террористы выходцы с северного	35	073	804	9%	0	4%	0	4%	ence corpus.txt	82%
5	как раз-таки об этом и говорил... все террористы идноты и суки, погибшим и	34	619	790	4%	0	3%	0	3%	ence corpus.txt	80%
6	почему люди задают вопрос зачем террористы это делают? разве не	34	066	761	6%	0	2%	0	2%	ence corpus.txt	79%
7	кого-то потерял... в любом случае, террористы только и хотят, чтобы мы их	34	038	761	6%	0	2%	0	2%	ence corpus.txt	79%
8	Я вот думаю, что это никакие не террористы. Особенно радуют статейки.	31	754	642	0%	0	6%	0	6%	ence corpus.txt	74%
9	когда... Вы правда думаете, что это террористы? УЖАС!	31	613	633	0%	0	6%	0	6%	ence corpus.txt	74%
10	виноват не кто-то а мусульманские террористы... В страшное время живем,	28	895	493	0%	0	9%	0	9%	ence corpus.txt	67%
11	боевики, власти или "международные террористы" - они не должны добиться	27	955	446	3%	0	7%	0	7%	ence corpus.txt	65%
12	внешностью..... а ведь на все террористы... что творится??? кому от	27	690	427	1%	0	7%	0	7%	ence corpus.txt	65%
13	МЕТРО –MCS interfax.ru террористы забали Репортажи с места	27	327	400	2%	0	6%	0	6%	ence corpus.txt	64%
14	крестовые походы, я думаю, террористы мало понимают свою же	27	217	395	2%	0	5%	0	5%	ence corpus.txt	63%
15	что среди них живут такие вот боевики террористы. Они не виноваты, что в	27	167	391	0%	0	5%	0	5%	ence corpus.txt	63%
16	с разных сторон! Может например террористы сделали теракт а потом	22	757	126	9%	0	5%	0	5%	ence corpus.txt	53%
17	очевидцами, а что ж вы хотели, чтоб террористы в самый ожидаемый	22	356	102	0%	0	4%	0	4%	ence corpus.txt	52%
18	И эта практика «назначения в террористы» не просто вытеснила	20	221	959	8%	0	9%	0	9%	ence corpus.txt	47%
19	"террористов" — назначенные террористы. И эта практика	20	215	958	0%	0	9%	0	9%	ence corpus.txt	47%
20	на редкость гупые, или это не террористы, а именно те, кому на нас	20	187	955	3%	0	9%	0	9%	ence corpus.txt	47%
21	террористический режим. Так что или террористы на редкость гупые, или это	20	180	955	6%	0	9%	0	9%	ence corpus.txt	47%
22	лишают последнего. Если террористы реально хотят напугать	20	120	951	9%	0	8%	0	8%	ence corpus.txt	46%
23	подервался жених сестры. Не пойму, террористы хотят чего-то добиться от	20	073	946	0%	0	8%	0	8%	ence corpus.txt	46%
24	власти и повторят преступления. Террористы, кавказцы, какая разница?	19	259	875	0%	0	6%	0	6%	ence corpus.txt	44%
25	Интерпол. Очень страшно, если террористы почувствуют бессилие	19	252	874	1%	0	6%	0	6%	ence corpus.txt	44%
26	они подставляют Рамазана Кадырова. Террористы мечтают разжечь	18	985	852	9%	0	6%	0	6%	ence corpus.txt	44%
27	Господа Путин и Медведев заявили - Террористы будут пойманы и	18	901	844	3%	0	5%	0	5%	ence corpus.txt	44%
28	Не поймешь, когда взрывают террористы, а когда ФСБешники	18	489	807	5%	0	4%	0	4%	ence corpus.txt	43%
29	оп-kgb.org Товарищи террористы! А почему Вам не взорвать	17	431	711	0%	0	2%	0	2%	ence corpus.txt	40%

Рис. 3 – Конкорданс для словоформы «Террористы» в КдС

N	Word	Freq.	%
1	И Т Д	27	0.05
2	О ТОМ ЧТО	26	0.05
3	В ТОМ ЧТО	24	0.05
4	НА САМОМ ДЕЛЕ	14	0.03
5	ЕГО АЛЛАХ И	12	0.02
6	###	11	0.02
7	АЛЛАХ И ПРИВЕТСТВУЕТ	11	0.02
8	БЛАГОСЛОВИТ ЕГО АЛЛАХ	11	0.02
9	В ТОМ ЧИСЛЕ	11	0.02
10	В ТОМ ЧТОБЫ	11	0.02
11	ДА БЛАГОСЛОВИТ ЕГО	11	0.02
12	ДО ТЕХ ПОР	10	0.02
13	НА ПУТИ АЛЛАХА	10	0.02
14	НА ТО ЧТО	10	0.02
15	ПО ОТНОШЕНИЮ К	10	0.02
16	HTTP WWW.YOUTUBE	9	0.02
17	WWW.YOUTUBE.COM	9	0.02
18	НЕ МОЖЕТ БЫТЬ	9	0.02
19	В ПЕРВУЮ ОЧЕРЕДЬ	8	0.02
20	ДЛЯ ТОГО ЧТОБЫ	8	0.02
21	ЖЕНЩИН И ДЕТЕЙ	8	0.02
22	НА СВОЕЙ ЗЕМЛЕ	8	0.02
23	НЕТ ГРЕХА В	8	0.02
24	# МЛРД Т	7	0.01
25	YOUTUBE.COM.WATCH	7	0.01
26	АЛЛАХ И ЕГО	7	0.01
27	В МОСКОВСКОМ МЕТРО	7	0.01
28	ДО СИХ ПОР	7	0.01
29	ЖИТЬ ПО КОРАНУ	7	0.01
30	И ЕГО ПОСЛАННИК	7	0.01

N	Word	Freq.	%
1	###	84	0.19
2	В МОСКОВСКОМ МЕТРО	21	0.05
3	О ТОМ ЧТО	21	0.05
4	В #	15	0.03
5	НА ПАРКЕ КУЛЬТУРЫ	11	0.02
6	НА САМОМ ДЕЛЕ	11	0.02
7	У НАС В	10	0.02
8	В ЛЮБОМ СЛУЧАЕ	9	0.02
9	В МЕТРО НЕ	9	0.02
10	В ТОМ ЧИСЛЕ	9	0.02
11	И Т Д	9	0.02
12	В НАШЕЙ СТРАНЕ	8	0.02
13	В ТОМ ЧТО	8	0.02
14	В ЭТО ВРЕМЯ	8	0.02
15	МОЧИТЬ В СОРТИРЕ	8	0.02
16	ПО КРАЙНЕЙ МЕРЕ	8	0.02
17	ПУТИН ДОЛЖЕН УЙТИ	8	0.02
18	RU NEWS #	7	0.02
19	ДО СИХ ПОР	7	0.02
20	КОМУ ЭТО ВЫГОДНО	7	0.02
21	РОДНЫМ И БЛИЗКИМ	7	0.02
22	ТАК ЧТО НЕ	7	0.02
23	# В #	6	0.01
24	# MAPTA #	6	0.01
25	COM # HTML	6	0.01
26	LIVEJOURNAL.COM #	6	0.01
27	RU # #	6	0.01
28	КТО ЭТО СДЕЛАЛ	6	0.01
29	ТАК И НЕ	6	0.01
30	# # И	5	0.01

Рис. 4 – Кластеры слов

Еще одним интересным инструментом модуля Wordlist является такой инструмент, как Relationship, позволяющий выявлять коллокации при помощи показателя MI.

Суть этой процедуры в том, что она позволяет выявить совместную встречаемость не только последовательностей слов, но и слов, разделенных

другими слова (можно установить границы поиска, по умолчанию – 5). В этой функции мы использовали настройки по умолчанию, за исключением того, что минимальная частотность была увеличена нами с 5 до 10. На рис. 5 приведены коллокаты для леммы *Вы* в КУЭФ.

N	Word 1	Freq	Word 2	Freq	Texts	Gap	Joint	MI	Z	MI3	Log L
80	БыТЬ	78	ОНИ	288	1	1	6	3,76	0,74	8,93	20,67
81	ВА	5	#	549	1	1	5	6,53	6,09	11,17	45,30
82	ВАМ	145	ЗА	253	1	3	6	3,05	-0,47	8,22	15,14
83	ВАМ	145	БЫ	204	1	2	5	3,10	-0,35	7,74	12,88
84	ВАМ	145	БЫ	444	1	2	14	3,46	0,36	11,08	43,29
85	ВАМ	145	УБИВАТЬ	27	1	3	5	6,02	4,80	10,66	32,99
86	ВАМ	145	САМИМ	10	1	1	6	7,71	10,67	12,88	57,10
87	ВАМ	145	МЕШАЕТ	5	1	1	5	8,45	12,83	13,09	58,75
88	ВАС	166	НЕ	1515	1	1	46	3,21	-0,54	14,26	135,55
89	ВАС	166	ТАК	250	1	1	10	3,61	0,63	10,26	32,56
90	ВАС	166	САМИХ	9	1	1	5	7,41	8,65	12,05	45,03
91	ВАС	166	УЖЕ	126	1	1	7	4,08	1,41	9,70	27,09
92	ВАС	166	БУДУТ	78	1	3	6	4,55	2,15	9,72	27,03
93	ВАС	166	ЖЕ	207	1	2	6	3,15	-0,31	8,32	15,84
94	ВАС	166	БЫ	444	1	4	17	3,55	0,64	11,72	54,61
95	ВАС	166	ВАМ	145	1	3	5	3,40	0,11	8,04	14,76
96	ВАС	166	ВЗРЫВАТЬ	9	1	1	6	7,67	10,49	12,84	57,44
97	ВАША	29	В	1272	1	3	6	3,04	-0,48	8,21	15,85
98	ВАШЕГО	21	НАРОДА	42	1	1	12	9,43	28,36	16,60	145,45
99	ВАШЕЙ	31	СТОРОНЫ	20	1	1	6	8,94	16,81	14,11	65,62
100	ВАШИ	57	ОЮТ	136	1	1	5	5,03	2,81	9,68	25,78
101	ВАШИ	57	НА	675	1	2	8	3,40	0,15	9,40	24,26
102	ВАШИ	57	ТЕРРИТОРИИ	31	1	1	6	7,43	9,57	12,60	51,73
103	ВАШУ	17	ИДЕОЛОГИЮ	6	1	1	5	11,28	35,11	15,92	76,24

Рис. 6 – Коллокации для леммы «Вы» в КУЭФ

Как мы видим, среди коллокаций леммы *Вы* с наибольшим MI – *мешает, самим, взрывать, самих, убивать, будут*. Слово *Ваш* часто соседствует со словами *территория, идеология, сторона, народ*. Этот пример подтверждает, что коллокации можно рассматривать как отражение идеологии авторов [8, p. 337].

**Заключение.** Таким образом, проведенный нами анализ одной из веток экстремистского форума «КавказЧат», посвященных обсуждению теракта в московском метро, показал, что применение средств корпусного анализа текста позволяет выявить паттерны использования слов, характерные для речевого портрета его участников. В частности, с использованием таких приемов, как построение частотных списков, выделение кластеров слов, колло-

катов и построение конкордансов с помощью программы WordSmith tool 4.0, нами было выявлено, что участники форума используют в своей риторике противопоставление *мы vs. вы*, конструируют дискурс войны, избегая называть событие в метро терактом. Комментирующие указанное событие на новостных ресурсах конструируют дискурс трагедии, называя происшедшее терактом, однако значительную долю среди частотных слов занимает группа слов, номинирующих власть. Дальнейший анализ текстов форума с использованием полного набора функций программы WordSmith tool позволит получить новые данные о речевом портрете участников экстремистского форума.

#### СПИСОК ЛИТЕРАТУРЫ:

1. Жаворонкова, Т. В. Использование сети Интернет террористическими и экстремистскими организациями [Текст] / Т. В. Жаворонкова // Вестник ОГУ. – 2015. – №3 (178). – С. 30-36.
2. Загоровская, О. В. Речевое портретирование социальной группы как задача современной лексикографии [Текст] / О. В. Загоровская // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – 2013. – №. 2. – С. 174-177.
3. Захаров, В. П. Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке [Текст] / В. П. Захаров, М. В. Хохлова // Компьютерная лингвистика и интеллектуальные технологии. – 2010. – №. 9. – С. 137-143.
4. Литвинова, Т.А. Исследование «языка вражды» с использованием средств автоматической обработки текстов [Электронный ресурс] / Т.А. Литвинова, О.А. Литвинова // Форум молодых ученых. – 2019. – №1(29). – URL: [https://forum-nauka.ru/domains\\_data/files/29/Litvinova.pdf](https://forum-nauka.ru/domains_data/files/29/Litvinova.pdf) (дата обращения: 23.02.2019).
5. Литвинова, Т.А. Лингвистические методы выявления в Сети экстремистского контента и лиц, склонных к экстремизму [Текст] / Т.А. Литвинова, О. В. Загоровская // Современное право. – 2016. – №3. – С. 107-113.
6. Степин, Д.С. Особенности осуществления террористической агитации и пропаганды с использованием Интернет-ресурсов (на примере форума «Кавказ-Чат») [Текст] / Д.С. Степин // Проблемы теории и практики борьбы с экстремизмом и терроризмом: материалы Всерос. науч.-практ. конф. – Ставрополь: Северо-Кавказский фед. ун-т, 2015. – С. 25-32.
7. Qin, J. The Dark Web Portal Project: Collecting and Analyzing the Presence of Terrorist Groups on the Web / J. Qin, Y. Zhou, G. Lai, E. Reid, M. Sageman, H. Chen // Lecture Notes in Computer Science. – 2005. – vol. 3495. – P. 623-624.
8. Salama, A. Ideological collocation and the recontextualization of Wahhabi-Saudi Islam post-9/11: A synergy of corpus linguistics and critical discourse analysis / A. Salama // Discourse & Society. – 2011. – № 22(3). – С. 315-342.