

**Статистический анализ использования ЛГБТ-лексикона в англо- и русскоязычном интернет-дискурсе**

**Научный руководитель – Нефедова Лилия Амиряновна**

***Каримов Рауль Дамирович***

*Выпускник (магистр)*

Челябинский государственный университет, Челябинск, Россия

*E-mail: urecraefte@gmail.com*

В настоящей работе рассматривается вопрос использования различной лексики, используемой для обозначения членов ЛГБТ-сообщества, в русско- и англоязычном интернет-дискурсе, преимущественно в текстах онлайн-СМИ. Данная проблема является актуальной, так как ЛГБТ-сообщество (здесь и далее используем наиболее распространенную аббревиатуру: лесбиянки, геи, бисексуалы и транссексуалы) до сих пор остается одной из самых уязвимых социальных групп в мире: эти лица подвергаются уголовному преследованию в 76 странах, в России и Латвии приняты дискриминационные законы о т.н. «гей-пропаганде» [5]. В связи с этим представляет интерес коннотативная характеристика описания данного сообщества в языке СМИ. Материалом настоящего исследования послужили два крупных корпуса: Timestamped JSI web corpus 2014-2016 English объемом ок. 21,3 млрд слововхождений и Timestamped JSI web corpus 2014-2016 Russian объемом около 1,8 млрд слововхождений; оба корпуса сгенерированы веб-кроулерами из текстов СМИ и социальных сетей, имеют одинаковый временной охват; бесплатно доступны на платформе SketchEngine [3], которая использовалась в качестве системы поиска по корпусу. В целях поправки на разность в объемах корпусов все статистические данные нормализованы (даны либо в частях на миллион, либо весовым коэффициентом, рассчитанным скоринговой системой SketchEngine). В таблице 1 представлена исследованная лексика, в состав которой входят 20 русскоязычных лексем и 12 англоязычных. Разница обусловлена аналитическим характером английского языка, ввиду которого одно и то же слово может принадлежать разным частям речи и иметь по несколько эквивалентов в русском языке. В связи с этим, а также с синонимичностью таких слов, как gay и homosexual, разбивка по частям речи ни для русского, ни для английского языков не производилась.

Для сравнения частотностей некоторые слова были сгруппированы, их частотности в группе — суммированы, за счет чего удалось свести оба языка к 12 числовым значениям. Затем воспользовались языком программирования R [4] в целях расчета спаренного критерия Стьюдента, по результатам которого получили значение  $p = 0,06055$ . В то время как данное значение не соответствует общепринятому порогу достоверности ( $p < 0,05$ ) [2], вынуждая нас принять нуль-гипотезу, отметим, что оно находится достаточно близко к достоверному уровню. Отметим, что частотность некоторых лексем составила 0, что говорит о ничтожно малой употребимости того или иного слова относительно размеров корпуса. Теперь проанализируем лексическую сочетаемость слов «гей, гомосексуалист» в русском языке, «gay, homosexual» в английском; для чистоты эксперимента англоязычную выборку ограничим существительными. В русскоязычном корпусе найдено 61 типовое словосочетание с этими словами, из них 12 носит положительную окраску, 28 — нейтральную и 21 — отрицательную. В англоязычном корпусе найдено 172 словосочетания, из них 21 носит положительную окраску, 45 — нейтральную и 106 — отрицательную. Нормализованные (приведенные к проценту) значения составят: для русского языка — 19,67, 45,9, 34,42; для английского — 12,2, 26,16, 61,62. Парное сравнение этих показателей возвращает  $p =$

0,99, что говорит об отсутствии всяких достоверных различий, т.е. распределение словосочетаний по оценочному компоненту носит идентичный характер. Отметим также, что среди всех словосочетаний русского и английского языков может выделить 21 RU-EN пару идентичных семантически и стилистически фраз, в которых анализируемые лексемы сочетаются со следующими словами: болезнь, брак, быть, грех, дискриминация, запрещать, защищать, избиение, казнить, клуб, латентный, мужчина, ненависть, отец, открытый, пара, парад, преследовать, разрешать, скрытый, человек, штраф, убийство. Данные пары были собраны в таблицу 2 с указанием веса того или иного словосочетания в корпусе, качества оценки (нейтральное, отрицательное, положительное) в разрезе языка. Полученная статистическая информация проверена методом дисперсионного анализа [1] частотностей от языка, качества и сочетанной лексемы; полученные результаты показаны в таблице 2.

Как видно из таблицы 2, ни по какому из параметров не наблюдается статистически достоверных различий. Похожие результаты получены и в отношении других лексем из таблицы 1 (здесь не приводятся ввиду ограничений на объем тезисов). Совокупный анализ говорит о том, что характер употребления лексики в русско- и англоязычном интернет-дискурсах различается мало. Однако необходимо отметить существенные ограничения настоящего исследования, которое базируется на статистическом подсчете лексем без учета общего характера и оценочного суждения в источниках, включенных в тот или иной корпус. Так, говоря о «казни» геев, автор высказывания может выступать как за такую меру, так и против нее, или даже цитировать кого-либо еще. Данное обстоятельство, а также оправданность диахронического расширения исследуемого материала (в целях, например, сравнения характера русскоязычного медийного текста до и после 2013 года) говорит об актуальности дальнейшего исследования.

### Источники и литература

- 1) Johansson, C. Statistics for Cognitive Sciences. Bergen: University of Bergen. 2017. <http://www.sketchengine.co.uk> (SketchEngine).
- 2) Rowentree, D. Statistics Without Tears. London: Penguin Books. 2000.
- 3) <https://www.sketchengine.co.uk> (SketchEngine).
- 4) <https://cran.rstudio.com> (The Comprehensive R Archive Network).
- 5) <https://www.theguardian.com/global-development-professionals-network/2017/mar/01/where-are-the-most-difficult-places-in-the-world-to-be-gay-or-transgender-lgbt> (The Guardian: Where are the Most Difficult Places in the World to Be Gay or Transgender).

Русский язык	Частотность	Английский язык	Частотность
Гей	1,41	Gay	35,5
гомосексуалист	1,84	homosexual	4
гей-	5,87		
гомосексуал	0,03		
гомосексуальный	1,46		
гейский	0		
Гомосексуальность	0,69	homosexuality	3,7
содомия	0,16	sodomy	0,8
Педерастия	0,03	Pederasty	0,03
Содомит	0,1	Sodomite	0,02
однополый	5,88	Same-sex	12,9
Бисексуал	0,02	Bisexual	2,8
бисексуальный	0,05		
Бисексуальность	0,05	bisexuality	0,19
Лесбиянка	1,32	Lesbian, dyke	6,88
Лесбийский	0,23		
Транссексуал	0,13	Transsexual	0,44
трансгендер	0,15	transgender	11,7
транссексуальный	0		
трансгендерный	0,01		

**Таблица 1.** Перечень проанализированной лексики

	Df	Sum Sq	Mean Sq	F-value	Pr(>F)
Коннотация	2	18.79	9.395	1.776	0.182
Язык	1	0.60	0.605	0.114	0.737
Коннотация и язык	2	8.82	4.411	0.834	0.442
Остатки	40	211.56	5.289		

**Таблица 2.** Дисперсионный анализ словосочетаний