

**ИССЛЕДОВАНИЕ ЛЕКСИКИ ТЕКСТОВ ЖАНРА 2.0
МЕТОДАМИ КВАНТИТАТИВНОЙ И КОРПУСНОЙ ЛИНГВИСТИКИ
(НА ПРИМЕРЕ ТЕКСТОВ ИНСТАГРАМ)**

О. В. Донина

Воронежский государственный университет

**USING THE CORPUS LINGUISTICS'
AND QUANTITATIVE LINGUISTICS' METHODS IN THE STUDY
OF THE GENRE 2.0 LEXICON
(ON THE EXAMPLE OF INSTARAM TEXTS)**

O. V. Donina

Voronezh State University

Аннотация: в рамках данной статьи рассмотрены возможности использования инструментария квантитативной и корпусной лингвистики для анализа текстов жанра 2.0 на примере социальной сети Инстаграм. Объем исследовательского корпуса составил 43 000 словоупотреблений. Все рассматриваемые тексты были разделены на три группы по количеству подписчиков у их авторов (100 тыс. – 400 тыс.; 500 тыс. – 900 тыс.; 1 млн – ∞). Целью исследования была апробация средств компьютерной лингвистики для анализа текстового материала жанра 2.0. Среди решаемых в рамках статьи задач стоит указать: 1) предварительную обработку данных (лемматизация, удаление стоп-слов); 2) выявление ключевых слов при помощи корпусного менеджера AntConc; 3) визуализацию данных при помощи Voyant Tools; 4) проведение кластеризации ресурсами языка R; 5) сопоставление полученных показателей по авторам и по указанным выше группам. Согласно гипотезе исследования, выделенные группы инстаграм-блогеров должны кластеризоваться по авторам в зависимости от количества подписчиков. Подтверждение гипотезы позволило бы в дальнейшем разработать автоматический классификатор инстаграм-текстов. Самыми частотными словами для всего исследовательского корпуса оказались: свой; очень; самый; большой; год. Были сопоставлены частоты использования этих слов по группам (самое существенное отклонение при этом составило 0,26 %) и по авторам внутри групп (где указанная величина варьировалась от 0,5 до 0,75 %). Множественный коэффициент корреляции также показал, что сходство частотного распределения слов выше между группами (45 %), чем между авторами внутри одной группы (варьируется от 15 до 35 %). Далее были сопоставлены топ-20 самых частотных слов каждой группы и указаны частеречные предпочтения: в первой группе половина слов представлена прилагательными, в то время как в третьей группе 45 % составляют существительные. Далее был рассчитан процент уникальной и совпадающей лексики по группам (уникальная лексика составила 74,9 %) и по авторам (уникальная лексика составила 70,6 %). На последнем этапе проверки гипотезы было решено выяснить, существуют ли уникальные для каждой группы стилевые характеристики; для чего был применен метод стилеметрии, который также не выявил зависящих от количества подписчиков групп. Подводя итоги проведенному анализу, стоит отметить, что, несмотря на то, что гипотеза исследования не подтвердилась и никаких статистически значимых отличительных особенностей групп инстаграм-аккаунтов в зависимости от количества подписчиков выявить не удалось, в рамках данной статьи был предложен комплексный инструментарий для квантитативного анализа текстов, в том числе жанра 2.0. Учитывая полученные в рамках текущей работы результаты, показавшие, что индивидуальные особенности инстаграм-блогеров являются более существенными, чем предполагаемые внутригрупповые сходства, в качестве перспектив



исследования стоит указать возможность изучения индивидуальных характеристик авторов в Инстаграм и поиск возможных закономерностей, в том числе повлиявших на результаты стилиметрии.

Ключевые слова: Instagram, количественный анализ, AntConc, стилиметрия, R, Voyant Tools, визуализация данных, жанр 2.0.

Abstract: the article discusses possibilities of using the corpus linguistics' and quantitative linguistics' methods in the study of the genre 2.0 texts (on the example of Instagram texts). The volume of the research corpus was 43,000 words. All texts were divided into three groups depending on the number of subscribers of their authors (100 thousand – 400 thousand; 500 thousand – 900 thousand; 1 million – ∞). Among the tasks solved within the framework of the article were: 1) preliminary data processing (lemmatization and the removal of stop words); 2) identification of keywords using AntConc; 3) data visualization with the help of Voyant Tools; 4) clustering using R language; 5) comparison of the obtained indicators by authors and by the above-mentioned groups. According to the research hypothesis, the selected groups of Instagram bloggers should be clustered by authors depending on the number of their subscribers. If the hypothesis were confirmed, it would allow developing an automatic classifier of Instagram texts. The most frequent words for the entire research corpus were found (svoy (own); ochen (very); samyi (most); bolshoi (big); god (year)). The frequencies of using these words were compared by groups (the most significant deviation was 0.26 %) and by authors within groups (where the indicated value varied from 0.5 % to 0.75 %). The multiple correlation coefficient also showed that the similarity of the frequency distribution of words was higher between groups (45 %) than between authors within the groups (varied from 15 % to 35 %). Then the top 20 most frequent words of each group were compared and POS preferences were indicated: half of the words in the first group were represented by adjectives, while in the third group 45 % were nouns. Then the percentage of unique vocabulary was calculated by group (unique vocabulary was 74.9 %) and by authors (unique vocabulary was 70.6 %). The last step included the usage of the method of stylemetry, which also did not reveal groups depending on the number of subscribers. Despite the fact that the research hypothesis was not confirmed and no statistically significant distinctive features of the mentioned groups (depending on the number of subscribers) was identified, a comprehensive toolkit for quantitative text analysis was proposed.

Key words: Instagram, quantitative analysis, AntConc, stylometry, R, Voyant Tools, visualization, genre 2.0.

Введение

В связи с усовершенствованием технических возможностей интернет-платформ, а также развитием культуры соучастия (participative culture), происходит постоянная трансформация онлайн-жанров (которые также называют цифровыми жанрами (Л. Ю. Щипицина), интернет-жанрами (Е. А. Жигалина, Е. И. Горошко), электронными жанрами (Л. А. Капаназе, Е. Н. Вавилова), сетевыми жанрами (Г. Н. Трофимова), виртуальными жанрами (О. В. Лутовинова, Л. Ф. Компанцева), дигитальными жанрами (Р. К. Потапова) [1]), изучением, описанием и структурированием которых занимается виртуальное жанроведение [2; 3]. Несмотря на разнообразие разработанных классификаций цифровых жанров, наиболее перспективным кажется подход, при котором учитывается параметр развития веб-технологий: веб 1.0 (статичные сайты), веб 2.0 (динамическое создание содержания), веб 3.0 (концепция развития Интернета). Сейчас мы живем в эпоху веб 2.0, ключевой особенностью которой является возможность пользователей самостоятельно наполнять сайты и общаться между собой, что отражается в таких формах онлайн-взаимодействия, как блоги, wiki-проекты, социальные сети. Появление подобных форматов способствовало возникновению цифрового жанра социальных медиа, называемого жанром 2.0, а также направления языкознания, занимающегося анализом этого жанра – лингвистики 2.0 [4]. Среди отличительных

особенностей жанра 2.0 стоит отметить креализованность, гипертекстуальность, интерактивность и оперативность [5].

Материалы и методы исследования

В рамках данной статьи будут рассмотрены возможности использования инструментария количественной и корпусной лингвистики для анализа текстов жанра 2.0 на примере социальной сети Инстаграм. Согласно мировой статистике Инстаграм в 2020 г. ежедневно посещает 1 млрд пользователей, при этом 71 % пользователей моложе 25 лет, а гендерное распределение довольно равномерно: 52 % женщин и 48 % мужчин. В России сервисом пользуется каждый четвертый человек – 37 млн пользователей, 67 % из этого количества составляют пользователи в возрасте 18–34 года. Инстаграм рассматривается не только как социальная сеть, но и как площадка для продвижения товаров и услуг (причем как при помощи создания рекламных постов прямого характера (когда бренд рекламирует свой продукт), так и косвенного характера (когда медийные личности продвигают продукцию брендов) [6]): каждый месяц 2 млн пользователей размещают рекламу в рассматриваемом сервисе, а потенциальный рекламный охват составляет 849,3 млн человек. Согласно опросу компании ResearchMe (партнер Mail.Ru Group по проведению маркетинговых исследований), 58 % пользователей заходят в Инстаграм, чтобы следить за жизнью других людей, при этом 24 % ищут от сети вдохновения.

Таким образом, на платформе Инстаграм продаются не только различные продукты, но и образ жизни; рассматриваемая социальная сеть используется не только для рекламы, но и для саморекламы. Именно поэтому было решено проанализировать lifestyle-блогеров, подписчики которых посещают их страницы в поисках вдохновения.

В связи с тем, что целью ведения популярного блога часто является получение прибыли, которая напрямую зависит от количества подписчиков, было выдвинуто предположение, что посты инстаграм-авторов могут отличаться в зависимости от охвата аудитории. Поэтому было принято решение в рамках данной статьи условно разделить аккаунты инстаграм-блогеров на три группы по количеству подписчиков и в каждой группе рассмотреть три аккаунта инстаграм-авторов:

1-я группа: 100 тыс. – 400 тыс. подписчиков (блогеры @assol13, @elenagomashinaa и @ginevskaya);

2-я группа: 500 тыс. – 900 тыс. подписчиков (блогеры @dariachebanov, @kristabelkrayt и @mezenova);

3-я группа: от 1 млн подписчиков (блогеры @buzova86, @gudovanata и @karina_paletskikh).

Объем исследовательского корпуса составил 43 000 словоупотреблений. Целью исследования была апробация средств компьютерной лингвистики для анализа текстового материала жанра 2.0. Среди решаемых в рамках статьи задач стоит указать: 1) выявление ключевых слов при помощи корпусного менеджера AntConc; 2) визуализация данных при помощи Voyant Tools; 3) проведение кластеризации ресурсами языка R; 4) сопоставление полученных показателей по авторам и по указанным выше группам. Согласно гипотезе исследования, выделенные группы инстаграм-блогеров должны кластеризоваться по авторам в зависимости от количества подписчиков, так как лингвистические особенности рассматриваемых текстов (в том числе такие, как стилевые характеристики, ключевые слова и процент уникальной лексики) могут быть обусловлены популярностью и «читаемостью» инстаграм-постов. Подтверждение гипотезы позволит более подробно изучить особенности каждой группы с целью последующей разработки автоматического классификатора инстаграм-текстов, который можно будет использовать, например, при решении задач атрибуции текстов.

Результаты исследования и их обсуждение

Анализ работ по классификации жанров интернет-коммуникации [7–12] позволил выделить следующие ключевые жанрообразующие критерии: коммуникативная цель (Н. Г. Асмус, J. M. Swales, J. Yates, W. J. Orlikowski, C. Ihlstrom, Л. Ф. Компанцева), модель автора и адресата (Н. Г. Асмус, J. Yates, W. J. Orlikowski, Л. Ф. Компанцева), вербальное воплощение жанра (Н. Г. Асмус, J. M. Swales, Л. Ф. Компанцева), визуально-слуховое воплощение жанра

(J. M. Swales, Л. Ф. Компанцева), темпоральные параметры существования жанра (Н. Г. Асмус, J. Yates, W. J. Orlikowski, Л. Ф. Компанцева), локальные параметры существования жанра (J. Yates, W. J. Orlikowski), содержание и последовательность смысловых элементов (J. Yates, W. J. Orlikowski, C. Ihlstrom, Л. Ф. Компанцева), гипертекстовость (I. Askehave, A. Nielsen, Л. Ф. Компанцева), форма сообщения (J. Yates, W. J. Orlikowski, Н. Г. Асмус, C. Ihlstrom), структура функционирования (J. M. Swales, C. Ihlstrom, Н. Г. Асмус, Л. Ф. Компанцева), субъективность/оценочность/нестандартизованность (Н. Г. Асмус), коммуникативный регистр дискурса (Л. Ф. Компанцева), возможность глобального интерактивного развития жанра (Л. Ф. Компанцева).

Говоря о жанрообразующих признаках интернет-коммуникации, стоит обратиться к наиболее комплексной, на наш взгляд, модели Л. Ю. Щипициной [13], содержащей следующие параметры: 1) медийные (оформление, мультимедийность, гипертекстуальность, интерактивность, синхронность, количество коммуникантов); 2) прагматические (коммуникативная цель, тип автора, тип адресата, хронотоп); 3) структурно-семантические (содержание текста, структурные элементы текста); 4) языковые (конкретные языковые средства).

На основе указанных выше признаков Е. И. Горюшко [14] выделяет жанрообразующие параметры инстаграм-жанра: 1) структурное оформление домашней страницы пользователя; 2) цель; 3) концепция автора; 4) концепция адресата; 5) тематика сообщения; 6) время; 7) место; 8) языковые/неязыковые средства оформления публикаций. Как отмечает Ю. В. Щурина, инстаграм-пост – это «вторичный комплексный (комбинированный) жанр, гипержанр, включающий совокупность речевых жанров, организованных нелинейно (подпись к фотографии, сообщение о месте, сообщение об отмеченных на фотографии лицах)» [6, с. 158], характеризующийся креолизованным характером сообщений, гипертекстовостью, интерактивностью, синхронностью и наличием хэштегов и объединяющий элементы межличностной, групповой и массовой коммуникации.

Несмотря на наличие выделенных жанрообразующих признаков инстаграм-постов, количественный анализ текстов в Инстаграм проводился отечественными лингвистами только на материале английского языка для определения целей публикаций (анализировалось количество продвигающих постов, постов развлекательного характера и информационно-образовательных постов в выборке) [15]. Наличие этой лакуны в изучении рассматриваемого жанра во многом обусловило выбор темы данной статьи.

На первом этапе сопоставление наиболее частотных слов проводилось при помощи визуализации, предоставляемой ресурсом Voyant Tools (<https://voyant-tools.org/>).

Первым шагом стало создание облака слов для всего исследовательского корпуса (без разбиения по группам); предварительно тексты были лемматизированы и очищены от стоп-слов. Визуализация при помощи облака слов предполагает, что частота встречаемости слова в тексте и кегль, используемый при его написании на изображении, находятся в прямой зависимости. Таким образом, было выявлено, что для всего исследовательского корпуса (рис. 1) наиболее частотными словами выступают: *свой*; *очень*; *самый*; *большой*; *год*.



Рис. 1. Облако слов всего исследовательского корпуса

Далее (также при помощи программы Voyant Tools) было произведено сравнение частот использования пяти выделенных на предыдущем этапе слов по трем интересующим нас группам текстов инстаграм-пользователей с разным количеством подписчиков. На рис. 2 показано, что слово *свой* используется во всех группах с одинаковой относительной частотой (высчитывающейся как отношение количества словоупотреблений слова к общему количеству слов в подкорпусе) (0,58 % во всех трех группах); *самый* имеет одинаковый показатель встречаемости в группах 1 и 2 (0,4 %) и минимальное отличие в группе 3 (0,37 %); лексемы *очень* и *год* наиболее представлены во второй группе (0,49 и 0,4 % соответственно), в то время как в группах 1 и 3 их частота примерно равна (*очень* встречается в первой группе 0,4 %, а в третьей – 0,37 %; а показатели для *года* составляют 0,27 и 0,24 % соответственно); слово *большой* является наиболее часто встречающимся в первой группе (0,5 %) (а в группах 2 и 3 представлено в равной степени (0,24 %). Таким образом, слово *большой* представляет самое существенное отклонение в рамках рассматриваемых слов, которое составляет всего 0,26 %.

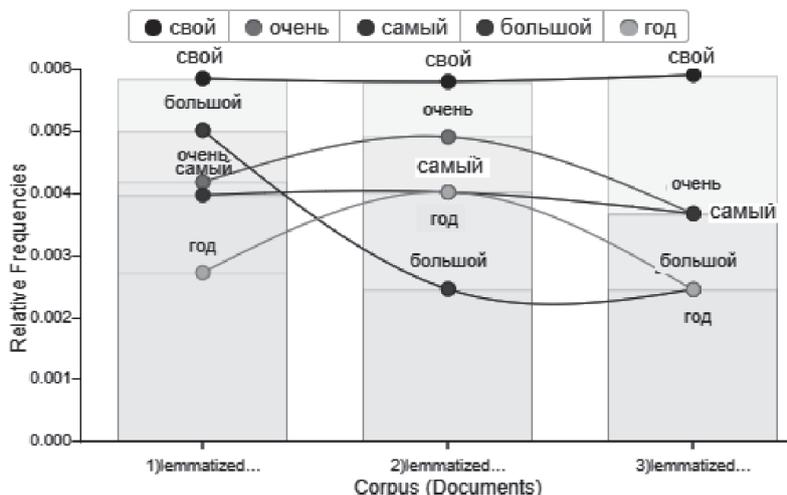


Рис. 2. Сравнение частоты использования слов *свой*, *очень*, *самый*, *большой*, *год* в трех группах текстов инстаграм-блогеров с разным количеством подписчиков

На следующем этапе были выявлены ключевые слова для трех указанных выше групп lifestyle-блогеров. Для работы использовался корпусный менеджер AntConc (<http://www.laurenceanthony.net/software/antconc/>) [16], с помощью которого было проведено ранжирование слов по частоте (предварительно из текстов были удалены стоп-слова и проведена леммати-

зация). В таблице представлены двадцать наиболее часто употребляемых слов отдельно для каждой группы.

Из приведенной таблицы видно, что в топ-20 слов каждой группы наблюдаются частеречные предпочтения: в первой группе половина слов представлена прилагательными, в то время как в третьей группе 45 % составляют существительные.

Таблица 1

Ключевые слова трех групп инстаграм-блогеров (слова ранжированы по частоте встречаемости)

Ранг	100 тыс. – 400 тыс. подписчиков	500 тыс. – 900 тыс. подписчиков	От 1 млн подписчиков
1	свой	свой	свой
2	большой	самый	простой
3	волосы	год	кожа
4	самый	сделать	самый
5	каждый	фото	день
6	новый	знать	большой
7	любить	простой	крутой
8	какой	большой	сделать
9	простой	всегда	время
10	красивый	любить	остров
11	продукт	первый	знать
12	идеальный	скидка	лицо
13	подарок	кожа	всегда
14	сейчас	новый	карта
15	знать	красивый	любить
16	цвет	начать	ребенок
17	много	поездка	результат
18	новогодний	сейчас	важный
19	год	цвет	какой
20	день	какой	мир

Чтобы определить значимость указанных выше частеречных отличий между группами, было принято решение проверить, насколько тексты внутри групп однородны. Для этого в Voyant Tools были сопостав-

лены относительные частоты встречаемости ключевых слов в текстах внутри каждой из трех групп (рис. 3–5).

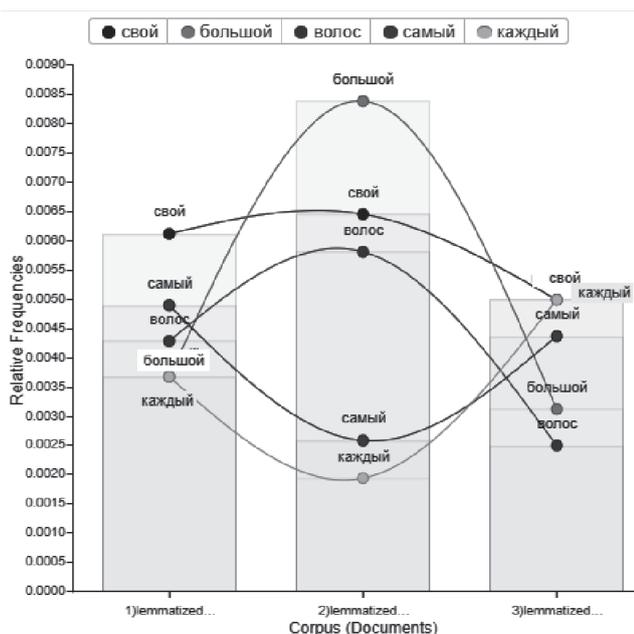


Рис. 3. Сравнение частоты использования ключевых слов в трех текстах первой группы (100 тыс. – 400 тыс. подписчиков)

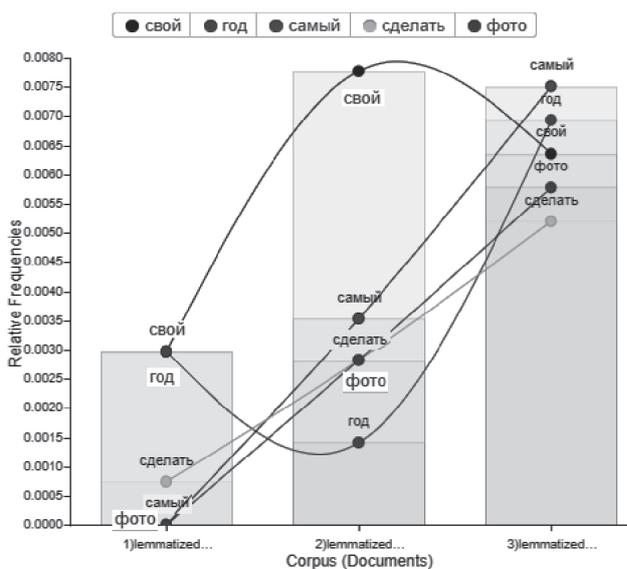


Рис. 4. Сравнение частоты использования ключевых слов в трех текстах второй группы (500 тыс. – 900 тыс. подписчиков)

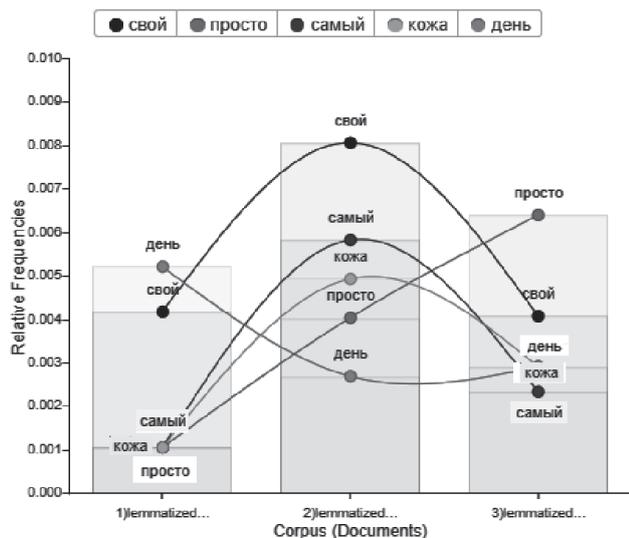


Рис. 5. Сравнение частоты использования ключевых слов в трех текстах третьей группы (от 1 млн подписчиков)

Как видно из рис. 3–5, частота употреблений ключевых слов авторами внутри групп не отличается однородностью: если при сопоставлении, производимом между группами, наибольшая разница составляла 0,26 %, то при сопоставлении авторов внутри групп эта величина варьируется от 0,5 до 0,75 %.

Множественный коэффициент корреляции [17] подтвердил этот вывод: сходство частотного распределения слов оказалось выше между группами (45 %), чем между авторами внутри одной группы (варьируется от 15 до 35 %). Полученные результаты показывают, что статистически значимые отличия между группами отсутствуют.

После получения этих результатов было решено проверить, сколько лемм совпадает у всех рассматриваемых девяти авторов. Для этого при помощи программы AntConc было рассчитано количество повторяющихся слов (с учетом лемматизации и с удалением стоп-слов).

Процент совпадающих лексических единиц (по авторам)

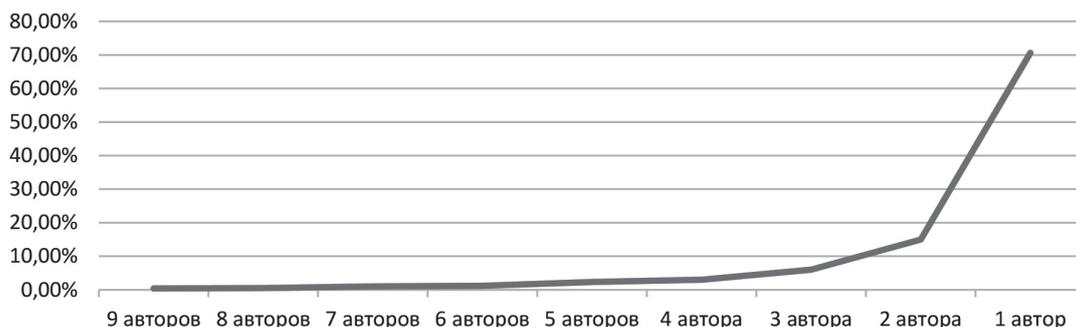


Рис. 6. Сравнение совпадающих лемм в девяти анализируемых аккаунтах

На рис. 6 на оси абсцисс указано количество авторов, у которых была обнаружена совпадающая лексика, а на оси ординат – процент совпадающих лемм: так, ближе к началу координат мы видим, что количество лемм, которые были обнаружены во всех 9 аккаунтах, близко к 0 % (а именно 0,4 %), при этом превалирует уникальная лексика, встречающаяся только у одного автора (резкий пик в правой части

графика) – 70,6 %; лемм, которые повторяются в двух (из девяти) аккаунтах, было найдено 15 %; лемм, которые встречаются сразу у трех блогеров, было найдено только 6 %; а остальные показатели не превышают 3 %.

Аналогичным образом был рассчитан процент уникальной и совпадающей лексики внутри рассматриваемых нами групп (рис. 7).

Уникальная лексика (по группам и между группами)

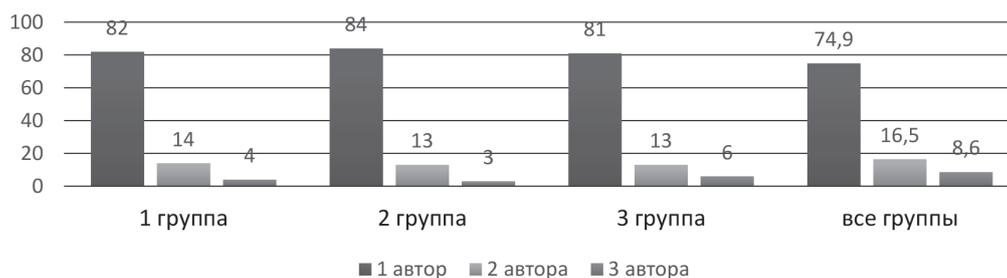


Рис. 7. Сравнение совпадающих лемм (с разбиением по группам и между группами)

Сравнение авторов внутри каждой из выделенных трех групп дало следующие результаты: в первой группе только 4 % лемм совпадают во всех трех рассматриваемых нами аккаунтах, 14 % лемм совпадают в двух аккаунтах, 82 % – уникальная лексика, встречающаяся только у одного автора; во второй группе 3 % лемм являются аналогичными у всех трех авторов, 13 % повторяются в двух блогах, 84 % – оригинальная лексика; в третьей группе 6 % лемм совпадают у трех инстаграм-пользователей и 13 % лемм аналогичны у двух авторов. Мы также сравнили, насколько группы схожи между собой (рис. 7), и получили следующие показатели: совпадение лемм у всех трех групп составило 8,6 %, а только у двух групп – 16,5 %, процент уникальной лексики составил, соответственно, 74,9 %.

Таким образом, используемый нами количественный инструмент показал, что инстаграм-блогеры, вне зависимости от количества подписчиков, при написании постов прибегают к уникальной лексике, т. е. отличий между тремя рассматриваемыми группами выявить не удалось.

На последнем этапе проверки гипотезы было решено выяснить, существуют ли уникальные для каждой группы стилиевые характеристики; для чего был применен метод стилеметрии (при помощи пакета *stylo* языка R). Данный инструмент применяется в том числе для атрибуции текстов, но для нас важна предоставляемая им возможность автоматической кластеризации текстов на основе их стилиевых характеристик.

На рис. 8 отражены результаты кластеризации в виде древовидного графа, где в качестве вершин выступают рассматриваемые инстаграм-аккаунты (при этом указанные на рисунке имена состоят из

цифры (1–3), обозначающей номер выделенной нами группы, и непосредственно названия самого аккаунта, как оно прописано в Инстаграм). Таким образом, кластерный анализ также не выявил зависящих от количества подписчиков групп (единственной парой аккаунтов из одной группы, объединенных в кластер, является пара 1_ginevskaya и 1_elenaromashinaa). В качестве перспектив исследования стоит отметить необходимость выявления стилиевых особенностей рассматриваемых текстов, влияющих на результат стилеметрии.

Заключение

Подводя итоги проведенному анализу, стоит отметить, что, несмотря на то, что гипотеза исследования не подтвердилась и никаких статистически значимых отличительных особенностей групп инстаграм-аккаунтов в зависимости от количества подписчиков выявить не удалось, в рамках данной статьи был предложен комплексный инструмент для количественного анализа текстов, в том числе жанра 2.0 (т. е. текстов социальных медиа). Учитывая полученные в рамках текущей работы результаты, показавшие, что индивидуальные особенности инстаграм-блогеров являются более существенными, чем предполагаемые внутригрупповые сходства, в качестве перспектив исследования стоит указать возможность изучения индивидуальных характеристик авторов в Инстаграм и поиск возможных закономерностей.

ЛИТЕРАТУРА

1. Кукуева Г. В. Интернет-жанр или коммуникативная практика? (к вопросу о терминологическом определении текстов, функционирующих в виртуальном пространстве) // Экология языка и коммуникативная практика. 2018. № 1. С. 44–48.
2. Горошко Е. И. Лингвистика Интернета : формирование дисциплинарной парадигмы // Жанры и типы текста в научном и медийном дискурсе. Орел, 2007. Вып. 5. С. 223–237.
3. Горошко Е. И., Жигалина Е. А. Виртуальное жанроведение : устоявшееся и спорное // Вопросы психолингвистики. 2010. № 12. С. 105–124.
4. Горошко Е. И. «Чирикающий» жанр 2.0 Твиттер или что нового появилось в виртуальном жанроведении? // Вестник Твер. гос. ун-та. 2011. № 3. С. 11–20.
5. Горошко Е. И., Полякова Т. Л. К построению типологии жанров социальных медий // Жанры речи. 2015. № 2 (12). С. 119–127.
6. Щурина Ю. В. Жанровое своеобразие социальной сети Instagram // Жанры речи. 2016. № 1 (13). С. 156–168.
7. Асмус Н. Г. Лингвистические особенности виртуального коммуникативного пространства : дис. ... канд. филол. наук. Челябинск, 2005. 266 с.
8. Swales J. M. Genre Analysis English in academic and research settings. Cambridge : Cambridge University Press, 1990. 260 p.

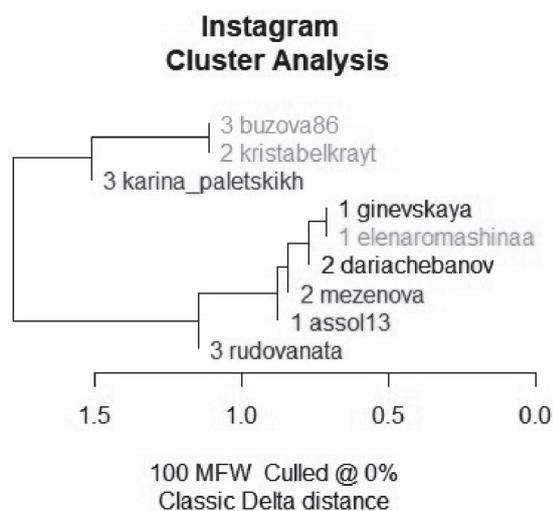


Рис. 8. Автоматическая кластеризация текстов девяти рассматриваемых нами авторов

9. Askehave I., Nielsen A. What are the Characteristics of Digital Genres? – Genre Theory from a Multimodal Perspective // *Proceedings of the 38th Hawaii International Conference on System Sciences*. 2005. Pp. 120–141.
10. Yates J., Orlikowski W. J. *Genre Systems : Structuring Interaction through Communicative Norms*. Cambridge, Massachusetts, 1998. URL: <http://ccs.mit.edu/papers/CCSWP205/>
11. Ihlstrom C. *The Evolution of a New(s) Genre* : Doctoral Dissertation. School of Information Science, Computer and Electronic Engineering, Halmstad University; Department of Informatics, Goteborg University. Gothenburg Studies in Informatics. 2004. 164 p.
12. Компанцева Л. Ф. Интернет-лингвистика : когнитивно-прагматический и лингвокультурологический подходы. Луганск : Знание, 2008. 528 с.
13. Ципицина Л. Ю. Жанры компьютерно-опосредованной коммуникации : монография. Архангельск : Помор. ун-т, 2009. 238 с.
14. Горошко Е. И., Полякова Т. Л. Инстаграм как жанр 2.0 (на примере политической коммуникации) // *Жанры речи*. 2019. № 4 (24). С. 300–313.
15. Мамонова Н. В. Классификация постов в англоязычной социальной сети Инстаграм (лингвосинергетический аспект) // *Вестник ЧелГУ*. 2019. № 4 (426). С. 137–142.
16. Anthony L. *AntConc (Version 3.5.8)* [Computer Software]. Tokyo, Japan: Waseda University. URL: <https://www.laurenceanthony.net/software>
17. Дони́на О. В. Способы визуализации результатов криптоклассного исследования // *Вестник Воронеж. гос. ун-та. Сер.: Лингвистика и межкультурная коммуникация*. 2015. № 3. С. 105–112.
- science?]. In *Vestnik Tverskogo gos. un-ta*. 2011. No. 3. Pp. 11–20.
5. Goroshko E. I., Polyakova T. L. K postroeniyu tipologii zhanrov sotsial'nykh medij [Toward the construction of a typology of social media genres]. In *Zhanry rechi*. 2015. No. 2 (12). Pp. 119–127.
6. Shchurina Y. V. Zhanrovoe svoebrazie sotsial'noj seti Instagram [Genre originality of the social network Instagram]. In *Zhanry rechi*. 2016. No. 1 (13). Pp. 156–168.
7. Asmus N. G. *Lingvisticheskie osobennosti virtual'no-govornogo kommunikativnogo prostranstva* [Linguistic features of the virtual communicative space]: PhD Dissertation. Chelyabinsk, 2005. 266 p.
8. Swales J. M. *Genre Analysis English in academic and research settings*. Cambridge : Cambridge University Press, 1990. 260 p.
9. Askehave I., Nielsen A. What are the Characteristics of Digital Genres? – Genre Theory from a Multimodal Perspective. *Proceedings of the 38th Hawaii International Conference on System Sciences*. 2005. Pp. 120–141.
10. Yates J., Orlikowski W. J. *Genre Systems: Structuring Interaction through Communicative Norms*. Cambridge, Massachusetts, 1998. Available at: <http://ccs.mit.edu/papers/CCSWP205/>
11. Ihlstrom C. *The Evolution of a New(s) Genre*: Doctoral Dissertation. School of Information Science, Computer and Electronic Engineering, Halmstad University; Department of Informatics, Goteborg University. Gothenburg Studies in Informatics. 2004. 164 p.
12. Компанцева Л. Ф. *Интернет-лингвистика когнитивно-прагматический и лингвокультурологический подходы* [Internet linguistics cognitive-pragmatic and linguocultural approaches]. Lugansk : Znanie, 2008. 528 p.
13. Shchipsina L. Y. *Zhanry komp'yuterno-oposredovannoj kommunikatsii* [Genres of computer-mediated communication]. Arkhangel'sk: Pomor. un-t, 2009. 238 p.
14. Goroshko E. I., Polyakova T. L. Instagram kak zhanr 2.0 (na primere politicheskoy kommunikatsii) [Instagram as a genre 2.0 (on the example of public communications)]. In *Zhanry rechi*. 2019. No. 4 (24). Pp. 300–313.
15. Mamonova N.V. *Klassifikatsiya postov v angloyazychnoj sotsial'noj seti Instagram (lingvosinergeticheskij aspekt)* [Classification of posts in the Russian-speaking social network Instagram (linguosynergetic aspect)]. In *Vestnik ChelGU*. 2019. No. 4 (426). Pp. 137–142.
16. Anthony L. (2019). *AntConc (Version 3.5.8)* [Computer Software]. Tokyo, Japan: Waseda University. Available at: <https://www.laurenceanthony.net/software>
17. Donina O. V. *Sposoby vizualizatsii rezul'tatov kriptoklassnogo issledovaniya* [Methods of visualizing the results of cryptoclass studies]. In *Vestnik Voronezhskogo gosudarstvennogo universiteta. Ser. Lingvistika i mezhkul'turnaya kommunikatsiya*. 2015. No. 3. Pp. 105–112.

REFERENCES

1. Kukueva G. V. *Internet-zhanr ili kommunikativnaya praktika? (k voprosu o terminologicheskom opredelenii tekstov, funkcioniruyushchikh v virtual'nom prostranstve)* [Internet genre or communicative practice? (on the issue of terminologically accessible texts functioning in the virtual space)]. In *Ekologiya jazyka i kommunikativnaya praktika*. 2018. No. 1. Pp. 44–48.
2. Goroshko E. I. *Lingvistika Interneta: formirovanie distsiplinarnoj paradigmy* [Linguistics of the Internet: scientific discipline and paradigm]. In *Zhanry i tipy teksta v nauchnom i medijnom diskurse*. Вып. 5. Orel: Kartush, 2007. Pp. 223–237.
3. Goroshko E. I., Zhigalina E. A. *Virtual'noe zhanrovedenie : ustoyavsheesya i spornoe* [Virtual genre studies: well-established and controversial]. In *Voprosy psikholingvistiki*. 2010. No. 12. Pp. 105–124.
4. Goroshko E. I. «Chirikayushhij» zhanr 2.0 Twitter ili chto novogo poyavilos' v virtual'nom zhanrovedenii? [The “Tweet” genre 2.0 Twitter or what’s new in virtual genre

Воронежский государственный университет

Донина О. В., кандидат филологических наук, преподаватель кафедры теоретической и прикладной лингвистики

E-mail: olga-donina@mail.ru

Поступила в редакцию 10 марта 2020 г.

Принята к публикации 15 июня 2020 г.

Для цитирования:

Донина О. В. Исследование лексики текстов жанра 2.0 методами квантитативной и корпусной лингвистики (на примере текстов Инстаграм) // Вестник Воронежского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2020. № 3. С. 25–33. DOI: <https://doi.org/10.17308/lic.2020.3/2928>

Voronezh State University

Donina O. V., Candidate of Philology, Lecturer of the Theoretical and Applied Linguistics Department
E-mail: olga-donina@mail.ru

Received: 10 March 2020

Accepted: 15 June 2020

For citation:

Donina O. V. Using the corpus linguistics' and quantitative linguistics' methods in the study of the genre 2.0 lexicon (on the example of Instaram texts). *Proceedings of Voronezh State University. Series: Linguistics and Intercultural Communication*. 2020. No. 3. Pp. 25–33. DOI: <https://doi.org/10.17308/lic.2020.3/2928>