

Google и лингвистические инструменты копирайтера

Челнокова Анна Алексеевна

Бакалавр в области европейских исследований,

Студентка второго курса магистратуры

Тверского Государственного Университета

по направлению Лингвистика – РФ, г.Тверь

Google and linguistic tools for copywriters

Chelnokova Anna

Bachelor of European studies,

Second-year student of Magistracy in Tver State University,

Department of Linguistics – Russia, Tver

Аннотация. Рекламный дискурс оказывает значительное влияние на лингвокультурное развитие человечества, его ценности и формирование лексикона, представляет собой значительный пласт коммуникации. Размышлять о творчестве, маркетинге, рекламе более объективно нам помогают независимые научные данные. В статье рассматриваются инструменты корпусной лингвистики Google, используемые для составления рекламных текстов, их концепции и принципы работы.

Abstract. Advertising discourse has a significant impact on the linguistic and cultural development of mankind, its values and the formation of a lexicon, it represents a significant layer of communication. Independent scientific data helps us to think about creativity, marketing and advertising more objectively. The article describes Google corpus linguistics tools used for the compilation of advertising texts, their concepts and principles of work.

Ключевые слова: корпусная лингвистика; корпус текстов; копирайтинг; нейминг; рекламный текст; Google Trends; Google Ngram Viewer; диахронические исследования; графики встречаемости слов.

Keywords: corpus linguistics; text corpus; copywriting; naming; advertising text; Google Trends; Google Ngram Viewer; diachronic studies; charts of word frequency.

Писательство — это искусство, маркетинг требует навыка. Слова могут нас рассмешить, заставить плакать и даже вдохновить на великие дела, подобрать правильные слова бывает сложно, но и для этого существует наука. Данные помогают нам размышлять о творчестве, маркетинге, рекламе более объективно.

Основная цель дизайнеров, копирайтеров, неймеров заключается в том, чтобы удостовериться, что каждое используемое слово осмысленно. Одно неверно выбранное выражение может навредить опыту потребителя или пользователя, если речь идёт о дизайне интерфейсов и программном обеспечении [11].

Неясная подпись на кнопке программы или незнакомый термин на упаковке могут запросто разочаровать аудиторию. Чтобы убедиться в правильности выбора слов, профессионалы используют различные технологичные методы, помогающие оценить лингвистическую и маркетинговую эффективность текстов и названий.

Одним из условий успеха в современном маркетинге является необходимость максимально быстро и корректно отследить социальные тренды, в том числе и языковые, чтобы также максимально корректно построить коммуникацию с потребителем. С точки зрения лингвистики, наиболее эффективным инструментом отслеживания языковых трендов являются корпуса текстов, так как они предоставляют в принципе неограниченные возможности для проведения всевозможных видов лингвистического структурного анализа. К тому же современные технологии привели нас к возможности создания веб-корпусов, что означает, во-первых, расширение ресурсов для создания корпуса, во-вторых, ускорение процесса обработки информации и увеличение размера веб-корпуса по отношению к традиционной модели [9, с.131].

Тем не менее, и это уже не ново, ведь самый крупный интернациональный корпус текстов появился с момента создания глобальных поисковых систем в сети Интернет и продолжает расширяться до сих пор силами как их разработчиков, так и самих пользователей. Созданные исключительно как инструмент поиска, предоставляющий список ссылок на источники информации, в порядке релевантности задаваемой в виде текстовой фразы-запроса, поисковые системы – такие как Google, Yahoo, MSN и даже отечественный Yandex – к 2017 году сами по себе превратились в мощные лингвистические базы данных, главным образом потому, что содержат на своих серверах хронологически систематизированную информацию о каждом из сотен миллиардов введённых пользователями текстовых запросов. Причём при анализе запросов результаты возможно отсеивать не только по хронологии с точностью до часа, но и по географии запроса или по отдельно взятому языку [1].

Google является крупнейшей в мире поисковой системой, и в совокупности с набором своих сервисов представляет собой также и крупнейшую глобальную текстовую базу. Медиагигант Google предоставляет открытый доступ к своей статистике запросов [8].

Только тщательное предметное исследование может помочь выяснить, на чём стоит сосредоточиться, к примеру, в выборе названия для линейки продукции или новой функции программного обеспечения. Стандартным инструментом исследования в маркетинге является опрос целевой аудитории.

Современные интерактивные технологии предлагают множество инструментов для проведения опросов и тестирований, такие как UserTesting, SurveyMonkey, Google Consumer Surveys и другие. Плюсом данного способа исследования является возможность установить обратную связь с потенциальными потребителями, однако проведение и распространение тестов в среде целевой аудитории требует времени и участия дополнительных агентов [11], а также не всегда в полной мере позволяет раскрыть

психосемантическую сущность и восприятие рекламного текста в сознании реципиента, так как большинство ответов, получаемых при исследовании на фокус-группах или с помощью сплошных опросов, являются осознанными, а значит недостоверными. Поскольку время в рекламе – решающий фактор, а точность информации несёт в себе определяющую роль, подобный корпус текстов является действительно надёжным и «живым» статистическим источником для копирайтера и рекламодателя при отслеживании тенденций, позволяющим в значительной степени экономить усилия и ресурсы во время проведения маркетинговых исследований [5, с.304].

В 2007 году Google подарил миру сервис, анализирующий поисковые запросы, записывающий и хранящий эти данные. Изначально сервис был доступен только в США, но со были добавлены другие страны, в 2013 году появилась возможность посмотреть статистику поисковых запросов Google и из России. Разберёмся в возможностях инструмента, чтобы понять его важность и полезность. С точки зрения лингвистики, данный сервис может быть полезен как корпусный инструмент проведения синхронических исследований.

Функционал Google Trends может показаться довольно простым и примитивным, показаться не очень-то информативным, но это далеко не так. Сервис дает возможность просмотреть статистику поисковых запросов Google в любой момент прошлого, узнать, что происходит сейчас со статистикой и формами запросов, проведя анализ полученных данных, сделать прогноз на будущее. Знакомым элементом сервиса является поисковая строка Google, именно сюда необходимо вводить интересующие пользователя запросы [8]. При возникновении спорных случаев, как например при вводе слова «калина» (непонятно, что имел в виду пользователь, бренд-нейм Лада «Калина» или же растение) предусмотрена функция выбора тематики, с целью отсеивания нецелевой аудитории. Ниже в окне сервиса располагается фильтр, в котором предусмотрены более широкие функции выбора:

1. Выбор региональности запроса. Можно указать как страну и регион, так и город.

2. Хронологический период, чтобы просмотреть статистику за определенный год или месяц или за последние 7 дней (так же по часам), что может быть весьма полезным, если копирайтер работает на информационном портале, где постоянно размещаются свежие новости, либо если маркетологам и владельцам бизнеса необходимо оперативно проследить динамику спроса в определённом регионе, чтобы сконцентрировать внимание на данном рынке, как например, это происходит со сбытом одежды, бренд-неймы производителей которой мы рассматриваем в данной работе. Это сезонный вид деятельности, ориентирующийся на фактор региональности, основной чертой которого является необходимость быстро разрабатывать привлекательные эстетически рекламные кампании и, соответственно, давать имена моделям и коллекциям, здесь невозможно обойтись без отслеживания тенденций, ведь иногда спрос на всю марку может зависеть заинтересованности пользователей в определённой тематике.

3. Выбор категории поиска. Google Trends предлагает свои варианты тематик, но можно выбрать и самостоятельно из полного списка.

4. Направление поиска помогает понять, где пользователь пытался отыскать информацию: в картинках, в видео, новостях или с помощью обычного поиска в Интернет [7].

Следует рассмотреть реальный опыт использования Google Trends при разработке продукта. К примеру, разрешение вопроса выбора наиболее подходящей формулировки копирайтерами сервиса облачного хранения информации Dropbox. Проблема состояла в выборе выражения, которую потенциальный пользователь предпочтёт видеть при входе на сайт: Log in, Log on, Sign in, Sign on. По результатам исследования копирайтеров Dropbox "Sign in" является явным лидером [11]. Это означает, что большинство склонно

использовать его при обращении к данному действию. Таким образом именно эту надпись можно наблюдать при входе на англоязычную версию программы, выбор был сделан в пользу комфорта использования программы для широкой аудитории пользователей. Постепенно, большинство онлайн сервисов и ПО, требующее регистрации, пришли к тому, что для входа на сайт используется именно кнопка «Sign in». Опыт работы с сервисом Google Trends представлен в Приложении.

Вторым инструментом статистического корпусного поиска Google является Google Ngram Viewer или Google books Ngram Viewer – полное название системы. В отличие от Google Trends, Google books Ngram Viewer признан в среде лингвистов как серьёзный и проверенный ресурс для проведения исторических и сравнительных исторических исследований. Система Google books Ngram Viewer является в настоящее время наиболее мощным инструментом диахронических исследований. Доступ к ней открыт примерно с 2009 года. Google books Ngram Viewer представляет собой информационную систему, содержащую несколько корпусов размеченных текстов книг на 9 языках [12].

Каждый корпус, исключая итальянский, имеет две версии – 2009 и 2012 года. Система также содержит отдельно корпусы британского и американского английского языка, корпус всех вариантов английского языка, корпус художественной литературы на английском языке и так называемый «гугловский» миллион - книги на английском языке, изданные с 1500 по 2008. Для каждого года издания отобрано не более 6000 книг [2]. Это значит, что для ранних лет издания присутствуют все отсканированные книги, в то время как книги, опубликованные позднее, представлены выборочно. Отображается распределение по тематике, характерное для заданного года.

Самые поздние публикации, доступные пользователям в настоящее время, относятся к 2008 году. На конец 2012 г. база данных насчитывала более 8 миллионов книг, текстов, что составляет около 6% всех когда-либо

опубликованных в мире печатных книг. Корпус книг на русском языке на 2012 г. содержал 591 310 книжных текстов, а именно 67 137 666 353 словоупотреблений [3, с. 309].

Основой работы системы является N-грамма, специальная лексическая единица – последовательность от одной до пяти словоформ. Причем, чтобы быть учтенной и обработанной, N-грамма должна встречаться в корпусе не менее 40 раз. Для каждой заданной лексической единицы и, соответственно, для заданного временного интервала система строит график, по вертикальной оси которого она откладывает относительную частоту встречаемости данной N-граммы в корпусе в заданном году, выражая результат в процентах. На горизонтальной оси показываются годы, входящие в данный временной интервал. Каждая кривая графика маркируется отдельным цветом, в конце каждой кривой указывается, какому конкретно слову или словосочетанию (N-грамме) она соответствует [12]. При построении графиков изменения частоты употребления лексической единицы используется «сглаживание» (smoothing), а при нулевом сглаживании в графике учитывается относительная частота встречаемости N-граммы за каждый год. Однако по-настоящему тенденция в динамике встречаемости слов прослеживается более отчетливо при скользящем усреднении данных, когда частота встречаемости N-граммы за какой-то год вычисляется как средняя величина заданного интервала сглаживания [3, с. 321]. Кроме построения графиков, система предоставляет ссылки ко всем найденным текстам, где встретились заданные лексические единицы. Как правило, это библиографические описания книг и фрагменты текстов с выделенными найденными N-граммами. В некоторых случаях полный текст книги доступен в графическом формате.

Каким образом пользоваться сервисом Ngram Viewer? В первую очередь необходимо зайти на сайт Google Ngram Viewer, затем в текстовое поле "Graph these case-sensitive comma-separated phrases" ввести искомое слово (фразу), а в текстовое поле "between" задать год начала цикла отслеживания лексической

единицы, а в текстовое поле and внести год конца цикла. В выпадающем списке "from the corpus" необходимо выбрать язык и запустить поиск с помощью кнопки "Search lots of books". Таким образом, будет построен график частоты использования заданной единицы, а под графиком в разделе "Search in Google Books" и будут приведены все ссылки с результатами в Google Books [6].

Аналитические ресурсы, которые открывает для лингвистов Google books Ngram Viewer, практически безграничны, так как в системе существуют возможности сложного поиска, работы с кривыми графика, тегирование контекста и прочие детальные функции. В практическом же смысле и в контексте лингвомаркетинга система помогает в значительной степени сэкономить время и силы при решении такой ключевой проблемы как сочетаемость слов [10]. Сочетаемость слов можно проследить с помощью словарей или национальных корпусов языков, однако среднестатистический копирайтер и неймер, в недостаточной степени владеющий специальными знаниями в данной области, скорее отдаст предпочтение общедоступному и проверенному сервису Google.

Как же работает Google books Ngram Viewer в маркетинговой реальности? К примеру, сервис также помог копирайтерам Dropbox избежать серьезной речевой ошибки при запуске мобильного приложения, оснащённом графическим инструментом ввода личной клиентской подписи. Изначально окошко должно было быть оснащено надписью "Sign your signature", разработчикам такое употребление показалось некорректным. За правильным вариантом употребления они обратились к Google Ngram Viewer. Выяснилось, что данный вариант не употребляется вообще, и следует указать «Sign your name» [11]. Также имеется возможность задавать распознавание заглавных и строчных букв при формулировке условий поиска или вовсе игнорировать различие между ними. В системе отсутствует грамматическая нормализация лексических единиц, иначе говоря, поиск лексической единицы и построение

графиков частоты ее встречаемости осуществляется исключительно для заданной словоформы.

Как же само по себе диахроническое исследование и корпус текстов могут помочь в позиционировании бренда? Составить корректный слоган для бренда, проверить тенденцию употребляемости названия бренда, если в нейме сделан акцент на историзм, проверить наличие и популярность исторической реалии на определённом промежутке времени, наличие и отсутствие определённых коннотаций и так далее. В настоящее время сервис Ngram Viewer поддерживает следующие языки: китайский, английский, французский, немецкий, иврит, русский [3, с. 319-327].

В современном обществе рекламный дискурс и его лингвофункциональная семантика выступают одной из ментальных единиц языковой картины мира, аккумулируют различные ассоциации и образы, понятия, представления, ценностные установки, системы оценок, которые реализуются в каждом из языков. Семантико-понятийное наполнение имени бренда строится на основе выявления и анализа способов образования имен брендов [5, с. 304].

Дискурс бренд-коммуникации – суггестивная сила, внушающая публике чувство привязанности к определенной марке, тем важнее предельно осознанно подходить к выбору слов. Ведь каждый элемент рекламного дискурса: наименование бренда, слоган, рекламный текст, подпись на кнопке компьютерной программы или приложения – прочно входят в наш обиход, восходят в раздел эпонимов, обогащают лексикон и лингвокультурное наследие человечества [9, с. 131]. При создании корректного рекламного текста важно учитывать уже сложившиеся элементы этого духовного наследия, традиции, лексические особенности культур и так далее, а помогают в этом современным специалистам глобальные коммуникационные системы, такие как Google, и созданные им инструменты исследования языка.

Список литературы:

1. Захаров В.П. Веб-пространство как языковой корпус. [Электронный ресурс]. – Режим доступа: <http://www.dialog-21.ru/media/2438/zakharovv.pdf> (дата обращения: 12.05.2017)
2. Захаров В.П. Лингвистика больших корпусов. [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/281150286_Lingvistika_bolsih_korpusov_Linguistics_of_Big_Corpora (дата обращения: 15.05.2017)
3. Захаров В.П., Масевич А.Ц. Диахронические исследования на основе корпуса русских текстов Google Books Ngram Viewer/Diachronic investigations on the base of Russian corpus of Google Books Ngram Viewer. Структурная и прикладная лингвистика. – СПб, 2014. №10. –303–327 с.
4. Как отслеживать частоту применения слов с помощью Ngram Viewer? [Электронный ресурс]. – Режим доступа: <http://netler.ru/articles/ngram-viewer.htm>. (дата обращения: 13.05.2017)
5. Ягодкина, М.В. Реклама в коммуникационном процессе. – СПб.: Питер, 2014. – 304 с.
6. Google books Ngram Viewer [Электронный ресурс]. – Режим доступа: <http://books.google.com/ngrams> (дата обращения: 15.05.2017)
7. Google Trends [Электронный ресурс]. – Режим доступа: <https://trends.google.ru/trends> (дата обращения: 14.05.2017)
8. Google Trends. Definition. [Электронный ресурс]. – Режим доступа: <http://whatis.techtarget.com/definition/Google-Trends> (дата обращения: 14.05.2017)
9. Czinkota M.R., Ronkainen I.A. International marketing, 7-th edition USA: international student edition. New York, 2004. – 131 с.
10. Michel J.B. Quantitative Analysis of Culture Using Millions of Digitized Books science. [Электронный ресурс]. – Режим доступа:

<http://www.sciencemag.org/content/331/6014/176.full.html> (дата обращения: 16.05.2017)

11. Sign on: копирайтер Dropbox о том, как выбирать слова для интерфейсов. [Электронный ресурс]. – Режим доступа: <https://vc.ru/p/design-words-with-data> (дата обращения: 10.05.2017)

12. Yuri Lin, Jean-Baptiste Michel, Erez Lieberman Aiden, Jon Orwant, Will Brockman and Slav Petrov. Syntactic Annotations for the Google Books Ngram Corpus. [Электронный ресурс]. – Режим доступа: <http://aclweb.org/anthology/P/P12/P12-3029.pdf> (дата обращения: 14.05.2017)