

УДК 316.728

О.Н. Мусина

СОВРЕМЕННОЕ АМЕРИКАНСКОЕ ОБЩЕСТВО: АНАЛИТИЧЕСКИЙ ОБЗОР ТЕМЫ СЫРОВ В ОБЩЕСТВЕННОМ СОЗНАНИИ ЗАРУБЕЖНОГО ПОЛЬЗОВАТЕЛЯ ГЛОБАЛЬНОЙ СЕТИ

Сделана попытка провести первоначальный статистический анализ по демографическому, гендерному, географическому и семантическому признакам повседневных разговоров жителей США в сети Интернет на темы, касающиеся сыров. Через призму «сырных тем» получена картина современного американского общества. Выявленные закономерности могут быть полезны при продвижении продукции сыродельной отрасли на западный рынок.

Анализ, разговоры, Интернет, США, сыр, статистика.

Введение

Проведение любой научно-исследовательской работы на должном уровне предполагает предварительное ознакомление с состоянием вопроса в изучаемой области, выявление тенденций развития, узких мест и проблем, иными словами, анализ, на основании которого выкристаллизовывается формулировка цели и задач НИР. Качественное выполнение этого этапа работы является залогом получения конкурентоспособной продукции. В то же время проведение поиска информации о состоянии вопроса всегда сопряжено с существенными затратами времени и интеллектуальных ресурсов. Особенно это актуально для информации на иностранных языках. Вместе с тем даже предварительный частотный анализ ключевых слов в изучаемой области весьма информативен.

Целью данной статьи является ответ на вопрос: какую роль в повседневных сетевых разговорах жителей США играют «сырные темы», а также последующий качественный и количественный анализ полученной информации, характеризующей срез современного американского общества.

Материалы и методы

Объект исследования – ежедневные повседневные разговоры пользователя американского сектора Интернета на темы, связанные с сырами. Обработаны данные за один месяц, анализ проведен по полу, возрасту и географии. Для получения первичного массива данных использован ресурс *Lexicalist* [1].

Lexicalist позиционирует себя как «демографический словарь современного американского варианта английского языка». *Lexicalist* использует искусственный интеллект, чтобы, проанализировав web и отсортировав данные «кто говорит и о чем», получить в результате демографическую картину в фактическом современном преломлении. Средства *Lexicalist* анализируют «миллионы слов разговоров в Интернете», здесь под разговорами подразумеваются повседневные беседы, зачастую о незначительных вещах. *Lexicalist* работает, анализируя обширные источники информации онлайн, включая сообщения в блогах, комментариях новостей, и социальные сети,

такие как *Twitter* и другие социальные медиасервисы. С помощью *Lexicalist* мы получаем текущие тенденции по частоте обсуждения ключевого слова, а также информацию о том, кто стоит за этим обсуждением. Также *Lexicalist* находит слова и фразы, которые семантически подобны оригинальному ключевому слову (*related words*) – встречаются в предложениях очень близко к оригиналу или обнаруживаются в подобных контекстах (например, «апельсины» и «яблоки» семантически подобны, потому что они оба появляются в компании таких слов, как «дерево» и «едят»).

Результаты и их обсуждение

При отборе информации средствами *Lexicalist* мы сталкиваемся с впервые сформулированным автором данной статьи «парадоксом ключевых слов» [2]. Какие ключевые слова, помимо собственно *cheese*, вводить в поисковую строку? Чтобы рационально сформулировать запрос, т.е. не получить завышенное число ссылок или отсутствие ссылок по запросу, необходимо в деталях представлять себе состояние вопроса в изучаемой области. Но ведь именно детальное представление и является целью поиска.

Выход из создавшегося положения мы нашли, используя «алгоритм предварительного библиографического поиска» [2], состоящий из трех крупных этапов: 1) контент-анализ; 2) библиометрический анализ; 3) содержательный анализ. Здесь необходимы некоторые пояснения. Контент-анализ – количественный анализ текстов с целью последующей интерпретации выявленных числовых закономерностей. Основа контент-анализа – это подсчет встречаемости слов в анализируемом информационном массиве. В результате выполнения первого этапа получаем ключевые слова по теме исследования. На втором этапе проводят библиометрический анализ, состоящий в нашем случае в поиске информации средствами *Lexicalist* по уже выявленным ключевым словам по теме «Сыры в повседневной коммуникации жителей США». Далее можно приступать к содержательному анализу выборки.

Для контент-анализа текстов существует многочисленное программное обеспечение, на этом во-

просе останавливаться не будем. Через такую программу были пропущены тексты, полученные с ресурса [3]. Предпосылкой к выбору ресурса послужили следующие причины. *Scirus* – наиболее специализированная для науки поисковая машина в сети. Посредством современных технологий поиска *Scirus* анализирует более 410 млн строго научных web-страниц, включая серверы препринт, цифровые архивы, базы данных журналов и патентов. *Scirus* работает только с источниками, имеющими научное содержание, отфильтровывая непригодные результаты поиска, позволяя быстро и точно отыскать академические, технические и медицинские данные в глобальной сети; найти новейшие сообщения, тезисы, статьи, патенты, препринт-издания и журналы, которые пропускают другие поисковые машины.

Для получения англоязычной совокупности терминов, относящихся к сырам, запрос был сформулирован следующим образом: искать слово «cheese» среди ключевых слов в публикациях от 1900 до 2012 г. среди любых научных ресурсов (<http://www.scirus.com/srsapp/advanced/index.jsp?q1=>). Для абсолютно всех слов публикаций, касающихся сыра и полученных с *Scirus*, включая союзы, глаголы и т.д., была определена частота встречаемости в тексте. Из полученного частотного ряда отбросили служебные слова, объединили лексико-семантические гнезда, т.е. один термин, находящийся в разных словоформах; в итоге получили представленную в виде таблицы парадигму ключевых слов анализируемого текста (табл. 1).

Поскольку совокупность ключевых слов была получена при анализе строго научных публикаций, закономерно, что не все из этих слов входят в активный вокабуляр жителя США и находят отражение в повседневных коммуникациях. Пропустив ключевые слова через *Lexicalist*, обнаружено, что в обыденных разговорах американцев фигурирует из этой совокупности парадигма следующих ключевых слов: *lactose, swiss cheese, science, cheddar cheese, microbiology, milk, cheese, goat cheese, dairy, mozzarella, cottage cheese, food, whey, yogurt, ice cream, ricotta, feta, cream cheese, frozen yogurt, butter, buttermilk, sour cream, parmesan*.

Отмечено (рис. 1), что суммарно по всей парадигме слов, касающихся сыра, несколько большую активность в повседневных коммуникациях жителей США проявляют женщины – около 56,7 %.

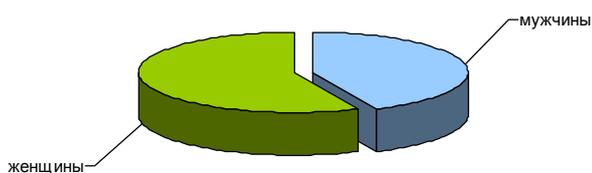


Рис. 1. Гендерная диаграмма распределения по парадигме ключевых слов

Далее для парадигмы слов была проведена статистическая обработка (табл. 2). Установлена единич-

ная встречаемость ключевого слова на каждые $v \cdot 10^6$ слов бесед, коэффициент v приведен в табл. 2. Так, например, слово *cheese* в повседневных коммуникациях американцев встречается в среднем единожды на каждые 15,5 тыс. слов, произнесенных в сети.

Выявлены тенденции в частоте употребления единичных ключевых слов: говорят в сентябре жители США о нем больше, чем месяц назад, и на сколько процентов, меньше или столь же часто. Хотя колебания в частоте обсуждения терминов могут быть вызваны сиюминутными причинами, стоит отметить, что согласно данным табл. 2 сыр как таковой, сыр чеддер, молоко, пицца, масло, сметана и сыр пармезан относятся к «вечным» темам, интерес к которым непреходящ и примерно остается на одном уровне.

Таблица 1

Парадигма ключевых английских слов по теме «Сыр»

Термин	Перевод
antimicrobial	антибактериальный препарат
antimicrobial activity	антибактериальная активность
bioactive peptides	Биологически активные пептиды
bioprocessing	биотехнология
butter	масло
buttermilk	пахта
casein	казеин
caseinate	казеинат
cheddar cheese	чеддер сыр
cheese	сыр
cheese manufacture	изготовление сыра
cheese ripening	созревание сыра
cheese whey	подсырная сыворотка
cheese yield	выход сыра
chromatography	хроматография
cottage cheese	домашний сыр
cream cheese	сливочный сыр
dairy	молочная отрасль
emulsifying salt	соль-эмульгатор
enterococcus	энтерококки
ewe milk	овечьё молоко
fermentation	брожение, ферментация
fermented milk	ферментированное молоко
feta	сыр фета
food	пицца
food chemistry	химия пищи
food engineering	пищевая инженерия
food science	пищевая наука
frozen yogurt	замороженный десерт из йогурта, более кислый, чем мороженое, и менее жирный
functional foods	функциональные продукты
goat cheese	козий сыр
hydrocolloids	гидроколлоиды
ice cream	мороженое
imitation cheese	заменители сыра
inactivation	инактивация
lactic acid bacteria	молочнокислые бактерии
lactobacillus	лактобацилла
lactococcus	лактококки
lactose	лактоза
listeria	листерия
microbiology	микробиология
milk	молоко
mozzarella	сыр моцарелла
parmesan	сыр пармезан
pasteurized milk	пастеризованное молоко
peptides	пептиды

probiotic	пробиотический
process cheese	плавленый сыр
processed cheese	переработанный сыр
proteolysis	протеолиз
raw milk	сырое молоко
rheology	реология
ricotta	сыр рикотта
ripened cheese	созревший сыр
ripening period	срок созревания
sheep and goats	овцы и козы
skim milk	обезжиренное молоко
sour cream	сметана
staphylococcus	стафилококк
swiss cheese	швейцарский сыр
water buffalo	индийский буйвол
whey	сухая сыворотка
yogurt	йогурт

Таблица 2

Статистический анализ парадигмы ключевых слов

Термин	Коэффициент ν	Тенденция за сентябрь	Количественное выражение тенденции, %
lactose	8,874	↓	15
swiss cheese	27,714	↑	3,5
science	0,183	↑	30
cheddar cheese	3,269	→	0
microbiology	15,664	↑	6,3
milk	0,209	→	0
cheese	0,155	→	0
goat cheese	15,138	↓	19
dairy	1,818	↓	27
mozzarella	10,852	↓	29
cottage cheese	14,527	↓	37
food	0,032	→	0
whey	8,619	↓	26
yogurt	1,142	↓	21
ice cream	0,194	↓	27
ricotta	0,031	↑	3,1
feta	13,473	↓	35
cream cheese	3,511	↑	22
frozen yogurt	7,383	↓	34
butter	0,382	→	0
buttermilk	24,676	↓	33
sour cream	5,930	→	0
parmesan	0,975	→	0

Для большей наглядности данных мы ввели коэффициент k (частота встречаемости единичного ключевого слова): $k = \frac{1}{\nu}$. Полученные коэффициенты

k показаны на рис. 2. Прежде всего привлекает интерес большая обсуждаемость в глобальной сети жителями США итальянского сыра рикотта, изготавливаемого из молочной сыворотки. Частота встречаемости ключевого слова «рикотта» сопоставима с «пища». «Сыр» как таковой упоминается в беседах почти так же часто, как «мороженое» и «молоко», ненамного от них отстает частота встречаемости слова «сливочное масло». Следует отметить, что

«сыр» встречается в речи американцев чаще, чем «наука» или, например, «микробиология».

Проранжировав отдельные виды сыров по встречаемости в повседневных коммуникациях, получен следующий частотный ряд (в порядке убывания): рикотта, пармезан, чеддер, сливочный сыр, моцарелла, фета, домашний сыр, козий сыр, швейцарский сыр. «Мороженое» приблизительно в пять раз обогнало по популярности «йогурт». Такой продукт, как сметана, не столь популярен в США, как, например, в России, и частота встречаемости этого слова в речи американцев сравнима с молочным десертом «замороженный йогурт».

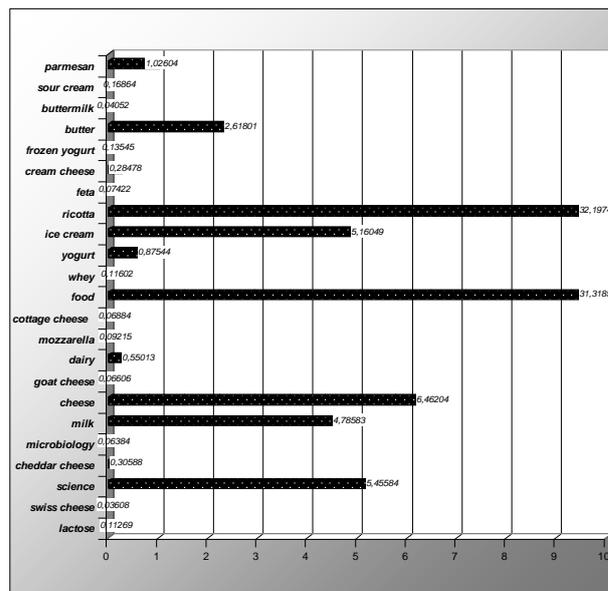


Рис. 2. Частота встречаемости отдельных ключевых слов в повседневной коммуникации жителей США

Информативен анализ семантически подобных слов-компаньонов (Related Words), вблизи которых и встречаются ключевые слова изучаемой парадигмы. Так, *science* упоминается в контексте слов: история, физика, математика, биология, химия, т.е. в таких разговорах речь идет не столько о молочном деле, сколько об естественно-научных областях знаний, по существу *science* выпадает из парадигмы. Все остальные термины находятся в компании слов, так или иначе связанных с ежедневным приемом пищи: *food* (обед, ленч, работа, напитки); *cheddar cheese* (домашний сыр, брокколи, сэндвич, крекеры, салат-латук, жареный); *cheese* (бекон, сэндвич, цыпленок, мороженое, завтрак, стейк); *ice cream* (пирожное, шоколад, печенье, суши, завтрак); *yogurt* (виноград, жареный цыпленок, китайская еда, мороженое, попкорн, арбуз); *whey* (порошок, снеки, взболтать, брикет, диета); *butter* (арахисовое масло, желе, печенье, сэндвич, пирог); *sour cream* (брокколи, домашний сыр, жареный, помидоры, чеснок); *feta* (домашний сыр, жареный, брокколи, крекеры, сэндвичи, овечий); *parmesan* (брокколи, жареный, вафли, полоски, ленты, корзинка, маленькие кусочки); *milk* (мороженое, кофе, пирожное, шоколад, алкоголь, «Starbucks» (сеть кофеен)).

руемых нами ключевых слов, проявили жители западных штатов – Колорадо и Орегон, немного отстают от них Мичиган, Вашингтон, Массачусетс, Калифорния, Вермонт, Иллинойс, Пенсильвания, Северная Каролина, Теннесси, штат Нью-Йорк, Миннесота, Огайо, Нью-Мексико, Невада, Алабама, Висконсин и Мэриленд (перечислены в порядке убывания активности). Средней, но все еще существенной активностью в повседневных коммуникациях характеризуются жители штатов Оклахома, Нью-Гэмпшир, Кентукки, Луизиана, Южная Каролина, Гавайи, Техас, Вирджиния, Флорида, Нью-Джерси, Аризона, Коннектикут, Миссури, Джорджия и Небраска. Примерно в 2–3 раза реже встречаются ключевые слова изучаемой парадигмы в речах жителей штатов Аляска, Юта, Айова, Айдахо, Мэн, а также самого маленького по площади штата Род-Айленд и штатов Индиана, Канзас, Миссисипи, Монтана, Арканзас, Делавэр, Южная и Северная Дакота, Западная Вирджиния. Минимальна активность в высокогорном штате на западе США – Вайоминге; вероятнее всего, это связано с тем, что это регион с самой низкой в стране плотностью населения.

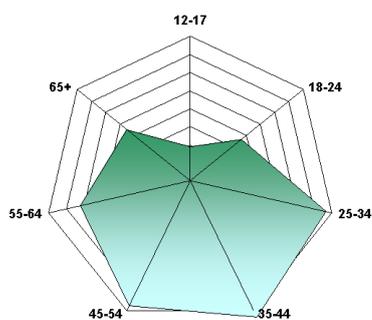


Рис. 6. Вклад различных возрастных групп в парадигму ключевых слов

Теперь, когда нам известно распределение по географическому признаку активности в использовании всех ключевых слов парадигмы, может быть небезынтересна информация по каждому штату о том, какое именно из ключевых слов в речах жителей фигурирует чаще всего. Итак, на Аляске и в Миссисипи наиболее часто говорят о пармезане; в Луизиане, Мичигане, Миннесоте, Оклахоме, Пенсильвании, Иллинойсе, Флориде, Джорджии, Колорадо, Нью-Джерси, Северной Каролине и Мэриленде – о швейцарском сыре; штат Мэн – единственный, в котором из всех ключевых слов наиболее часто фигурирует в повседневных коммуникациях козий сыр; в Массачусетсе, Вермонте, Нью-Гэмпшире и Миссури говорят о рикотте; в Делавэре и Айове – о моцарелле; в Айдахо и Род-Айленде – о сыре фета; в Индиане, Небраске, Висконсине и Вирджинии – о домашнем сыре; Вайоминг, являющийся наименее активным штатом в использовании ключевых слов, вошедших в парадигму, если и говорит о каком-то из молочных продуктов, то это сыр чеддер. В жарких местах – на Гавайях, в Калифорнии, Неваде, а также Канзасе и Орегоне чаще всего говорят о десерте – «замороженном йогурте». О молочных продуктах в целом гово-

рят в Западной Вирджинии, Северной и Южной Дакоте. Неожиданно, но в северо-западном штате Монтана из всех анализируемых нами ключевых слов чаще всего фигурирует сметана, по плотности населения штат занимает 44-е место в США. О побочных продуктах переработки молока, пахте и сыворотке, часто упоминают в сети люди из штатов Арканзас, Коннектикут, Кентукки, Огайо, Нью-Йорк, Южная Каролина, Юта, Вашингтон. В штатах Алабама, Аризона, Нью-Мексико, Теннесси и Техас чаще всего говорят о микробиологии, причем, вероятно, эти разговоры не связаны напрямую с микробиологией молока. Интересно, что нет какого-либо отдельного штата (в том числе южного), в котором из всей парадигмы ключевых слов было бы отдано предпочтение мороженому.

Автором была проанализирована детальная картина для каждого ключевого слова парадигмы: в каких из штатов США оно чаще всего встречается в ходе повседневных коммуникаций, происходящих в сети Интернет. Для этого число упоминаний в разговорах единичного ключевого слова в целом по всем штатам было принято за 100 % и рассчитано распределение этой величины по каждому из штатов США.

Рамки статьи не позволяют полностью привести результирующую таблицу, тем не менее анализ данных оказался весьма информативен. Так, с помощью результирующей таблицы может быть оценена заинтересованность в конкретном молочном продукте в целом по США. Например, рассмотрим популярность мороженого. Как посылает анализ данных, максимальное число упоминаний этого десерта приходится на штат Род-Айленд и представляет собой небольшую цифру в 3,1 %, что косвенно свидетельствует о равномерном распределении интереса потребителей к мороженому по всей стране. Аналогичная ситуация с йогуртом, молоком, сливочным маслом и сыром как таковым, т.е. можно обоснованно предположить, что эти традиционные молочные продукты не сдают своих позиций в рационе современных американцев и, вероятно, будут и в дальнейшем пользоваться стабильным спросом. Всплеск же интереса к отдельному молочному продукту в отдельном штате может быть продиктован модой или иными преходящими причинами. Например, почти треть всех упоминаний сыра рикотта в сети приходится на один-единственный небольшой штат на северо-востоке США – Вермонт. Аляска, хотя в целом и не особенно активная в сетевом обсуждении молочных продуктов, тем не менее является лидером по упоминанию в повседневных беседах сыра пармезан, сметаны и сливочного сыра. Отмечен интерес к моцарелле в Делавэре (11,1 %), причем все остальные штаты существенно отстают от этой цифры. Интересна ситуация со швейцарским сыром: четко определена группа штатов, в беседах жителей которых этот сыр фигурирует очень часто (Колорадо, Мичиган, Оклахома, Флорида, Миннесота, Мэриленд, Луизиана и Пенсильвания), в то же время есть группа штатов, где о швейцарском сыре за анализируемый период времени не было упомянуто вовсе (Алабама, Аляска, Аризона, Канзас, Монтана, Теннесси, Юта и др.). Такие общие темы, как «наука»,

«пицца», «молочные продукты», обсуждаются равномерно по всей стране. Аналогичные данные легко получить на основании анализа данных по каждому интересующему ключевому слову анализируемой парадигмы.

Выводы

Через призму «сырных тем» получена картина современного американского общества. Изучены данные о роли в повседневных сетевых разговорах жителей США этих тем, информация проанализирована по демографическому, гендерному, географическому и семантическому признакам.

Сыр как таковой, сыр чеддер, молоко, пицца, масло, сметана и сыр пармезан относятся к «вечным» темам, интерес к которым непреходящ и в беседах примерно остается на одном уровне.

Отмечена популярность среди пользователей американского сектора глобальной сети итальянского сыра рикотта, изготавливаемого из молочной сыворотки, частота встречаемости ключевого слова «рикотта» сопоставима с «пицца».

Мужчины больше говорят о швейцарском сыре, моцарелле и домашнем сыре, а также сыворотке, в

их разговорах сравнительно часто упоминается чеддер, сыр как таковой и сливочное масло. Среди женщин более популярен козий сыр, сливочный сыр, сыр фета, йогурт, пармезан, замороженный йогурт, сметана и мороженое, а среди всех упоминаний сыра рикотта в американском секторе глобальной сети 75 % принадлежит женщинам.

О швейцарском сыре говорят преимущественно мужчины старшего возраста (55–65 лет и старше), интерес к чеддеру появляется с 35 лет и чаще у женщин, козий сыр преимущественно фигурирует в речах молодых американок (25–44 года), пик интереса к моцарелле приходится на 45–54-летних мужчин, интерес к домашнему сыру постепенно нарастает с возрастом и достигает у американцев своего максимума после 65 лет, о рикотте, сливочном сыре и фете говорят молодые женщины 25–44 лет, а пармезан интересует главным образом взрослых американок в возрасте 35–64 лет.

Выявленные закономерности могут быть полезны при продвижении продукции молокоперерабатывающей отрасли на западный рынок.

Список литературы

1. www.lexicalist.com
2. Мусина, О.Н. Нетрадиционный алгоритм изучения состояния вопроса в пищевой отрасли / О.Н. Мусина // Техника и технология пищевых производств. – 2010. – № 1. – С. 78.
3. www.scirus.com

ГНУ «Сибирский научно-исследовательский институт сыроделия Россельхозакадемии»,
656016, Россия, г. Барнаул, ул. Советской Армии, 66.
Тел.: (3852) 56-46-12
e-mail: sibniis.altai@mail.ru

SUMMARY

O.N. Musina

MODERN AMERICAN SOCIETY: REVIEW OF THE «CHEESE TOPICS» IN THE CONSCIOUSNESS OF A FOREIGN USER OF THE GLOBAL NETWORK

The attempt to undertake an initial statistical analysis on the demographic, gender, geographic and semantic grounds of ordinary conversations of the United States inhabitants in the network of Internet on the topics concerning cheese has been done. Through the prism of the «cheese topics», a picture of modern American society has been obtained. The revealed regularities can be useful when promoting the products of the cheese-making industry to the Western market.

Analysis, conversations, the USA, cheese, statistics.

Siberian research Institute for cheese-making
Russian Academy of agricultural Sciences
656016, Russia, Barnaul, Sovetskaya Army, 66
Phone: (3852) 56-46-12
e-mail: sibniis.altai@mail.ru

