

УДК 81-139: 7.017.412: 316.613

ОБОСНОВАНИЕ ВЫБОРКИ МАССОВОГО ЭКСПЕРИМЕНТАЛЬНОГО ОБСЛЕДОВАНИЯ НОСИТЕЛЕЙ РУССКОГО ЯЗЫКА, НАПРАВЛЕННОГО НА ДИАГНОСТИКУ РАЗВИТИЯ СИСТЕМЫ ЦВЕТООБОЗНАЧЕНИЙ

©Грибер Ю. А., SPIN-код: 8214-8269, ORCID: 0000-0002-2603-5928, д-р культурологии,
Смоленский государственный университет, г. Смоленск, Россия, julia_griber@mail.ru
©Двойнев В. В., SPIN-код: 9806-4316, ORCID: 0000-0002-8931-3382, канд. социол. наук
Смоленский государственный университет, г. Смоленск, Россия, silidvoinev@yandex.ru

SAMPLE VALIDATION FOR THE EXPERIMENTAL MASS SURVEY OF NATIVE RUSSIAN SPEAKERS FOCUSSED ON THE DIAGNOSTICS OF COLOR NAMING

©Griber Yu., SPIN-code: 8214-8269, ORCID: 0000-0002-2603-5928, Dr. habil.,
Smolensk State University, Smolensk, Russia, julia_griber@mail.ru
©Dvoinev V., SPIN-code: 9806-4316, ORCID: 0000-0002-8931-3382, Ph.D.,
Smolensk State University, Smolensk, Russia, vassilidvoinev@yandex.ru

Аннотация. Массовые обследования носителей языка представляют собой сложную практику социолингвистических исследований, поскольку сопряжены с комплексом методических, финансовых и организационных трудностей. Одной из наиболее заметных проблем массового обследования носителей языка является отбор информантов, изучение которых дает возможность распространить полученные выводы на весь изучаемый социум. Целью настоящей статьи является обоснование выборочной совокупности массового экспериментального обследования носителей русского языка, направленного на определение современной структуры системы цветоименований: (1) определение критериев отбора респондентов из генеральной совокупности; (2) расчет объемов выборки; (3) обсуждение ее репрезентативности и достаточности. Приводится описание инструментария онлайн эксперимента, позволяющего сопоставлять цветовые термины, которые используются носителями разных языков, с цветовыми шкалами RGB и Манселла. Определяются лингвистический и социальный объекты исследования. Приводится численное выражение генеральной совокупности. Описывается стратегия формирования выборочной совокупности, приводятся формулы и выполненные расчеты. С опорой на произведенные расчеты внутри генеральной совокупности и в соответствии с полученными процентными долями рассчитывается количество респондентов, которых необходимо опросить для реализации выборочной совокупности. Предлагаемая схема формирования выборочной совокупности в настоящее время используется при проведении массового социолингвистического обследования носителей русского языка, направленного на диагностику развития системы цветообозначений, и может использоваться в качестве модели для анализа других языковых подсистем.

Abstract. Mass surveys of native speakers are an intricate research practice in the field of sociolinguistics, connected with many methodological, financial, and organizational difficulties. One of the most noticeable problems of a native speakers' mass survey is the selection of

respondents, who can provide information based on which conclusions can be made about the entire social object to be studied. The objective of this article is to validate the sample calculated for an experimental mass survey of native speakers of Russian. This survey is focused on understanding the actual structure of the colour-naming system, which means to: (1) determine the selecting criteria for respondents; (2) calculate the sample and (3) discuss its representativeness and sufficiency. The paper presents a description of the research toolkit (the RGB and Munsell colour scales) for the online experiment that allows to compare the terms used by native speakers of different languages for naming colours. The linguistic and social objects of the conducted research are also defined in the article. The article provides quantitative data about the population of Russia. Further, the authors represent the strategy of sampling. The calculations based on special formulas made possible to form the representative body of those to be surveyed from the entire population. The proposed way of sampling is being used for current sociolinguistic mass surveys of native speakers of Russian with the aim to test, to which extend their system of colour designation is elaborated. The presented sampling model can also be implemented in the analyses of other linguistic subsystems.

Ключевые слова: массовое обследование, русский язык, цветообозначения, обоснование выборки.

Keywords: mass survey, Russian, colour names, sample validation.

Введение

Массовые обследования носителей языка представляют собой достаточно сложную практику социолингвистических исследований, поскольку сопряжены с целым комплексом методических, финансовых и организационных трудностей (см., напр.: [1–2]). Обычно они проводятся для того, чтобы определить специфику функционирования в обществе отдельных подсистем языка, выводы о которых нельзя сделать, опираясь на единичные или случайные наблюдения. Поскольку в центре внимания при этом находится использование языка социальными группами, а не отдельными индивидами, одной из наиболее заметных проблем массового обследования носителей языка становится отбор информантов, изучение которых позволит распространить выводы на весь изучаемый социум.

Целью настоящей статьи является обоснование выборочной совокупности массового экспериментального обследования носителей русского языка, направленного на определение современной структуры системы цветоименований: (1) определение критериев отбора респондентов из генеральной совокупности; (2) расчет объемов выборки; (3) обсуждение ее репрезентативности и достаточности.

Материал и методы исследования

Инструментарий онлайн эксперимента [3] позволяет сопоставлять цветовые термины, которые используются носителями разных языков, с цветовыми шкалами RGB и Манселла [4]. Процедура эксперимента включает несколько этапов, на главном из которых респонденту последовательно предъявляются равномерно окрашенные прямоугольные карточки (цветовые образцы), случайно отобранные компьютером из палитры эксперимента, и предлагается назвать каждый цвет, используя наиболее подходящее цветообозначение (простые и составные слова, словосочетания, предложения). Вместе с названиями, которые участники дают появляющимся на экране цветовым образцам, записывается скорость каждого ответа, которая рассчитывается как временной интервал между предъявлением

цветового стимула и введением первого знака в строке ответа. Кроме этого, регистрируются условия наблюдения и собирается информация биографического характера, которая касается места рождения и постоянного проживания участников, их национальности, уровня владения языком, образования, возраста, пола и опыта работы с цветом.

Результаты и обсуждение

В случае массового экспериментального изучения носителей русского языка с целью выяснения современной структуры системы цветоименований определение объема выборки представляет собой многоступенчатую процедуру.

Сначала определяются лингвистический и социальный объекты исследования: совокупность лексических единиц, которые используются носителями современного русского языка для обозначения цвета и его оттенков, и русскоговорящие жители Российской Федерации (носители русского языка) в возрасте 16 лет и старше. Соответственно, в генеральную совокупность должны войти все респонденты старше 16 лет, которые говорят на русском языке, родились и постоянно проживают на территории РФ.

Далее проводится пилотажное исследование на случайной выборке, в ходе которого уточняются методические детали и определяются социальные характеристики членов изучаемого социума, релевантные для изучаемого объекта [5–7].

Только после этого определяются критерии отбора респондентов из генеральной совокупности, производится расчет объемов выборки и обоснование ее репрезентативности и достаточности.

На основе имеющихся данных государственной статистики о численности населения на 1 января 2017 года [8–9] в численном выражении генеральная совокупность (N) составляет 121 255 699 человек.

Расчет выборочной совокупности производится с применением формулы для случайного бесповторного отбора респондентов [10, с. 44]:

$$n = \frac{N \cdot t^2 \cdot p \cdot q}{N \cdot \Delta^2 + t^2 \cdot p \cdot q}$$

где n — объем выборочной совокупности; N — объем генеральной совокупности, Δ — предельная ошибка репрезентативности, p и q — выборочные доли, t — коэффициент доверия.

Предельная ошибка репрезентативности Δ составляет для данного исследования 0,05, следовательно, значение доверительной вероятности равно 0,95. Это позволяет говорить о верности выводов по результатам исследования относительно параметров генеральной совокупности в 95% случаев из возможных 100%. Соответствующий выбранному значению предельной ошибки репрезентативности Δ коэффициент доверия t составляет 1,96 [10, с. 44]. Для удобства расчетов его значение может быть увеличено до 2 (в этом случае значение доверительной вероятности составит 0,954) [11, с. 60]. Предельная ошибка выборки, таким образом, не превысит 5%, что допускается в исследованиях, проводимых в области общественных наук [11, с. 63]. Значение выборочных долей p и q принимается за 0,5 для каждой [10, с. 44].

$$n = \frac{121255699 \cdot 4 \cdot 0,5 \cdot 0,5}{121255699 \cdot 0,0025 + 4 \cdot 0,5 \cdot 0,5} = \frac{121255699}{75785,811875} = 1599,9$$

Таким образом, репрезентативная выборочная совокупность для исследования при предельной ошибке выборки, не превышающей 5%, составит 1600 человек. В целях повышения достоверности результатов исследования целесообразно предусмотреть запас в размере 400 человек. С учетом этой корректировки итоговая выборочная совокупность составит 2000 человек.

Поскольку носители языка представляют собой объект исследования, крайне неоднородный в возрастном, территориальном, социальном и других отношениях, критерии отбора респондентов должны определяться в строгом соответствии с рабочей гипотезой о корреляции структуры системы цветообозначений русского языка с полом, возрастом и местом проживания носителей.

Квоты выборочной совокупности (n=2000) рассчитываются в соответствии с составом генеральной совокупности.

В первую очередь, в генеральной совокупности определяются процентные доли мужского и женского населения старше 15 лет (по возрастным группам) в городах и сельской местности (Таблица 1).

Таблица 1.

СОСТАВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ [8, с. 21; 9, с. 92]

возраст	городское население				сельское население			
	мужчины		женщины		мужчины		женщины	
	N	%	N	%	N	%	N	%
15–19	2454068	6	2369727	5	964073	7	902121	6
20–24	2937833	7	2908866	6	1055540	7	925492	6
25–29	4608721	11	4653103	9	1425577	10	1191029	7
30–34	4798310	12	4952620	10	1471851	10	1314667	8
35–39	4208161	11	4449282	9	1297025	9	1240011	8
40–44	3771836	9	4130368	8	1235914	9	1242920	8
45–49	3246864	8	3627995	7	1191824	8	1212787	7
50–54	3205367	8	3887340	8	1339644	9	1402543	9
55–59	3451046	9	4579719	9	1496065	10	1628295	10
60–64	2807763	7	4226204	9	1155755	8	1420110	9
65–69	2151124	5	3590868	7	791372	6	1103859	7
70–74	898256	2	1703481	3	295709	2	518658	3
75–79	1087541	3	2624347	5	426763	3	962777	6
80–84	482622	1	1358257	3	208999	1	544253	3
85 +	344886	1	1182461	2	129702	1	461328	3
всего	40454398	100	50244638	100	14485813	100	16070850	100

Следующим шагом является выявление соотношения мужчин и женщин в генеральной совокупности, а также соотношения городского и сельского населения (Таблицы 2–3).

Таблица 2.

СОСТАВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ
 (население РФ в возрасте 15 и старше; мужчины и женщины)

	<i>N</i>	%
Мужчины	54940211	45
Женщины	66315488	55
Всего	121255699	100

Таблица 3.

СОСТАВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ
 (городское и сельское население в возрасте 15 и старше)

	<i>N</i>	%
Городское население	90699036	75
Сельское население	30556663	25
Всего	121255699	100

После этого в генеральной совокупности вычисляются процентные доли мужчин и женщин старше 15 лет в городском и сельском населении (Таблицы 4–5).

Таблица 4.

СОСТАВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ
 (городское население в возрасте 15 и старше; мужчины и женщины)

	<i>N</i>	%
Мужчины	40454398	45
Женщины	50244638	55
Всего	90699036	100

Таблица 5.

СОСТАВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ
 (сельское население в возрасте 15 и старше; мужчины и женщины)

	<i>N</i>	%
Мужчины	14485813	47
Женщины	16070850	53
Всего	30556663	100

Дальнейшие расчеты проводятся с целью дифференциации в генеральной совокупности процентных долей мужского населения старше 15 лет в городах и сельской местности и женского населения старше 15 лет в городах и сельской местности (Таблицы 6–7).

Таблица 6.

СОСТАВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ
 (городское и сельское население; мужчины в возрасте 15 лет и старше)

	<i>N</i>	%
Городское население	40454398	74
Сельское население	14485813	26
Всего	54940211	100

Таблица 7.

СОСТАВ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ
 (городское и сельское население; женщины в возрасте 15 лет и старше)

	<i>N</i>	%
Городское население	50244638	76
Сельское население	16070850	24
Всего	66315488	100

С опорой на произведенные расчеты внутри генеральной совокупности и в соответствии с полученными процентными долями рассчитывается количество респондентов, которых необходимо опросить для реализации выборочной совокупности (Таблицы 8–13).

Таблица 8.

СОСТАВ ВЫБОРОЧНОЙ СОВОКУПНОСТИ
 (n=2000; мужчины и женщины в возрасте 15 и старше)

	<i>N</i>	%
Мужчины	900	45
Женщины	1100	55
Всего	2000	100

Таблица 9.

СОСТАВ ВЫБОРОЧНОЙ СОВОКУПНОСТИ
 (n=2000; городское и сельское население в возрасте 15 и старше)

	<i>N</i>	%
Городское население	1502	75,1
Сельское население	498	25,0
Всего	2000	100

Таблица 10.

СОСТАВ ВЫБОРОЧНОЙ СОВОКУПНОСТИ
 (n=1502; городское население в возрасте 15 и старше; мужчины и женщины)

	<i>N</i>	%
Мужчины	676	45
Женщины	826	55
Всего	1502	100

Таблица 11.

СОСТАВ ВЫБОРОЧНОЙ СОВОКУПНОСТИ
 (n=498; сельское население в возрасте 15 и старше; мужчины и женщины)

	N	%
Мужчины	234	47
Женщины	264	53
Всего	498	100

Таблица 12.

СОСТАВ ВЫБОРОЧНОЙ СОВОКУПНОСТИ
 (городское и сельское население; мужчины в возрасте 15 лет и старше)

	N	%
Городское население	666	74
Сельское население	234	26
Всего	900	100

Таблица 13.

СОСТАВ ВЫБОРОЧНОЙ СОВОКУПНОСТИ
 (городское и сельское население; женщины в возрасте 15 лет и старше)

	N	%
Городское население	836	76
Сельское население	264	24
Всего	1100	100

В заключение рассчитанные квоты распределяются между возрастными группами в городской и сельской местности в полном соответствии с долями генеральной совокупности (ср. Таблицы 1 и Таблицы 14).

Таблица 14.

СОСТАВ ВЫБОРОЧНОЙ СОВОКУПНОСТИ (n=2000)

возраст	городское население				сельское население				
	мужчины		женщины		мужчины		женщины		
	N	%	N	%	N	%	N	%	
15–19	40	6	42	5	16	7	16	6	
20–24	47	7	49	6	16	7	16	6	
25–29	74	11	74	9	23	10	18	7	
30–34	81	12	83	10	23	10	21	8	
35–39	74	11	74	9	21	9	21	8	
40–44	61	9	66	8	21	9	21	8	
45–49	54	8	58	7	19	8	18	7	
50–54	54	8	66	8	21	9	24	9	
55–59	61	9	74	9	23	10	27	10	
60–64	47	7	74	9	19	8	24	9	
65–69	34	5	58	7	14	6	18	7	
70–74	13	2	25	3	5	2	8	3	
75–79	20	3	42	5	7	3	16	6	
80–84	8	1	25	3	3	1	8	3	
85 +	8	1	16	2	3	1	8	3	
всего	676	100	826	100	234	100	264	100	
ИТОГО					1502				498

Выборочная совокупность формируется с помощью комбинированной стратегии. На первом этапе ($n < 1000$) используется простая случайная бесповторная выборка. Далее применяется стратегия типической выборки, стратифицированной по полу, возрасту и месту проживания носителей русского языка, и осуществляется направленный рекрутинг участников исследования до выполнения всех необходимых условий репрезентативности выборки.

Заключение

Предлагаемая схема формирования выборочной совокупности в настоящее время используется при проведении массового социолингвистического обследования носителей русского языка, направленного на диагностику развития системы цветообозначений. Данные собираются в ходе онлайн эксперимента [3]. Предполагается, что в ходе исследования будет получено не менее 40 000 валидных цветообозначений от по меньшей мере 2 000 участников с различными социальными характеристиками (полом, возрастом, профессией, местом проживания). Собранные данные пройдут многоступенчатую «очистку», включающую проверку написания и орфографии, которая будет проводиться с использованием компьютерных программ и вручную, и представлены в форматах OpenOffice и Excel, удобных для последующего количественного анализа данных.

Публикация подготовлена в рамках поддержанного РФФИ научного проекта № 17-29-09145 «Картография цвета: диагностика развития цветоименований русского языка с использованием естественно-научных, историографических, социологических и психологических методов».

Источники:

- (1). Colour naming experiment. Режим доступа: <http://colornaming.net> (дата обращения 19.10.2018).
- (2). Демографический ежегодник России. 2017: статистический сборник / Росстат. М.: Федеральная служба государственной статистики, 2017. 263 с.
- (3). Российский статистический ежегодник. 2017: статистический сборник / Росстат. М.: Федеральная служба государственной статистики, 2017. 686 с.

Список литературы:

1. Крысин Л. П. Русский язык по данным массового опроса. М.: Проспект, 1968. 114 с.
2. Патрушева Г. М. Шорцы сегодня: современные этнические процессы. Новосибирск: Наука, 1996. 120 с.
3. Mylonas D., MacDonald L. Online Colour Naming Experiment Using Munsell Samples // Proceedings of the 5th European Conference on Colour in Graphics, Imaging and Vision (CGIV). Joensuu: IS&T, 2010. P. 27-32.
4. Paramei G. V., Griber Y. A., Mylonas D. An online color naming experiment in Russian using Munsell color samples // Color Research and Application. 2018. №43. P. 358-374. DOI: 10.1002/col.22190.
5. Грибер Ю. А., Милонас Д. Картография цвета: эмпирический анализ цветоименований русского языка // Человек и культура. 2015. №6. С. 64-94. DOI: 10.7256/2409-8744.2015.6.16636.

6. Griber Y. A., Paramei G. V., Mylonas D. Gender Differences in Russian Colour Naming // Being Color with health. AIC Color 2017 Proceedings / ed. by Y. J. Lee, J. Hwang, H. J. Suk, Y. K. Park. Jeju: KSCS, 2017. OS05-4.

7. Ротман Д. Г. Оперативные социологические исследования в комсомольской работе. Минск: Университетское, 1988. 113 с.

8. Могильчак Е. Л. Методика социологического исследования. Выборочный метод. М.: Юрайт, 2018. 117 с.

References:

1. Krysin, L. P. (1968). Russkii yazyk po dannym massovogo oprosa. Moscow, Prospekt, 114. (in Russian).

2. Patrusheva, G. M. (1996). Shortsyy segodnya: sovremennyye etnicheskiye protsessy. Novosibirsk, Nauka, 120. (in Russian).

3. Mylonas, D., & MacDonald, L. (2010). Online Colour Naming Experiment Using Munsell Samples. *Proceedings of the 5th European Conference on Colour in Graphics, Imaging and Vision (CGIV), Joensuu, IS&T*, 27-32.

4. Paramei, G. V., Griber, Y. A., & Mylonas, D. (2018). An online color naming experiment in Russian using Munsell color samples. *Color Research and Application*, (43), 358-374. doi:10.1002/col.22190.

5. Griber, Y. A., & Milonas, D. (2015). Kartografiya tsveta: empiricheskii analiz tsvetonaimenovaniy russkogo yazyka. *Chelovek i kul'tura*, (6), 64-94. doi:10.7256/2409-8744.2015.6.16636. (in Russian).

6. Griber, Y. A., Paramei, G. V., & Mylonas, D. (2017). Gender Differences in Russian Colour Naming. *Being Color with health. AIC Color 2017 Proceedings / ed. by Y. J. Lee, J. Hwang, H. J. Suk, Y. K. Park. Jeju, KSCS, OS05-4*.

7. Rotman, D. G. (1988). Operativnyye sotsiologicheskiye issledovaniya v komsomol'skoi rabote. Minsk, Universitetskoe, 113. (in Russian)

8. Mogilchak, E. L. (2018). Metodika sotsiologicheskogo issledovaniya. Vyborochnyy metod. Moscow, Yurait, 117. (in Russian)

*Работа поступила
в редакцию 11.10.2018 г.*

*Принята к публикации
16.10.2018 г.*

Ссылка для цитирования:

Грибер Ю. А., Двойнев В. В. Обоснование выборки массового экспериментального обследования носителей русского языка, направленного на диагностику развития системы цветообозначений // Бюллетень науки и практики. 2018. Т. 4. №11. С. 527-536. Режим доступа: <http://www.bulletennauki.com/griber-dvoinev> (дата обращения 15.11.2018).

Cite as (APA):

Griber, Yu., & Dvoinev, V., (2018). Sample validation for the experimental mass survey of native Russian speakers focussed on the diagnostics of color naming. *Bulletin of Science and Practice*, 4(11), 528-536. (in Russian).