The Prevalence and Impact of the “QWERTY Effect” on the Russian Language

By John Bucher Technical Mentor: Dr. Armstrong

I. Problem Description

Is the “QWERTY Effect” a prevalent force in the Russian language and does it follow the same trends that have been observed in other languages that use Roman characters?

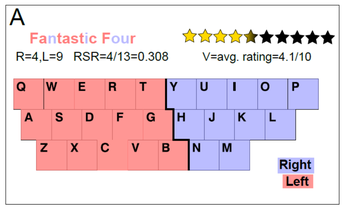
II. Literature Review

The “QWERTY Effect” is defined as:

- Right Side Ratios can be used to predict “Valence”
- Valence: the attractiveness or averseness of a word or sentence
- A rising usage of words with higher RSRs, with particular spikes around the 1990s (introduction of the internet)

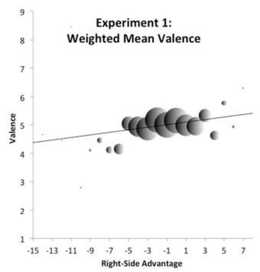
Effect observable in many Roman-character languages

- English, Spanish, Dutch, German, etc.



Left: The division of right-hand characters versus left-hand characters (Garcia and Strohmaier) [1]

Right:
Regression model on combined English, Spanish, and Dutch data
[b = .044, Wald χ2(1) = 5.34, p = .02] [3]



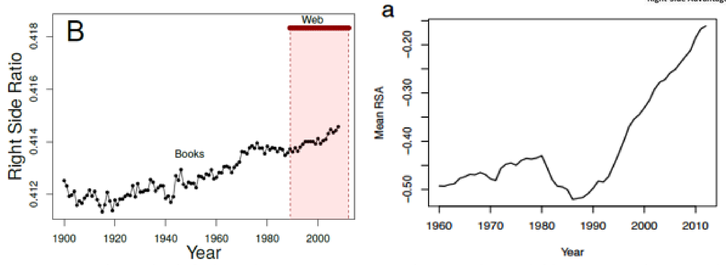


Figure 2: Graph on the left describes average RSR in the Google Book library over the period of approximately 100 years (Garcia and Strohmaier) and (Lin, Michel, Aiden, Orwant, Brockman, and Petrov). Graph on the right describes mean RSA of names given to at least 100 children per year from 1960–2012 (Cassanto, Jasmine, Brookshire, and Gijssels). [1 and 2]

The Russian language’s designated keyboard layout: JCUKEN

Left Side = Й, Ф, Я, Ц, Ы, Ч, У, В, С, К, А, М, Е, П, И Н, Р, Т --> 18 letters
Right Side = Г, О, Ь, Ш, Л, Б, Щ, Д, Ю, З, Ж, Х, Э, Ъ --> 14 letters

III. Primary Objective

To observe the existence, prevalence, and history of the “QWERTY effect” in the previously unexplored language of Russian, with its associated keyboard JCUKEN, using product reviews and name frequencies (Limitation: 120 person-hours over 10 weeks).

IV. Solution Description

Tools:

- C# and Visual Studio

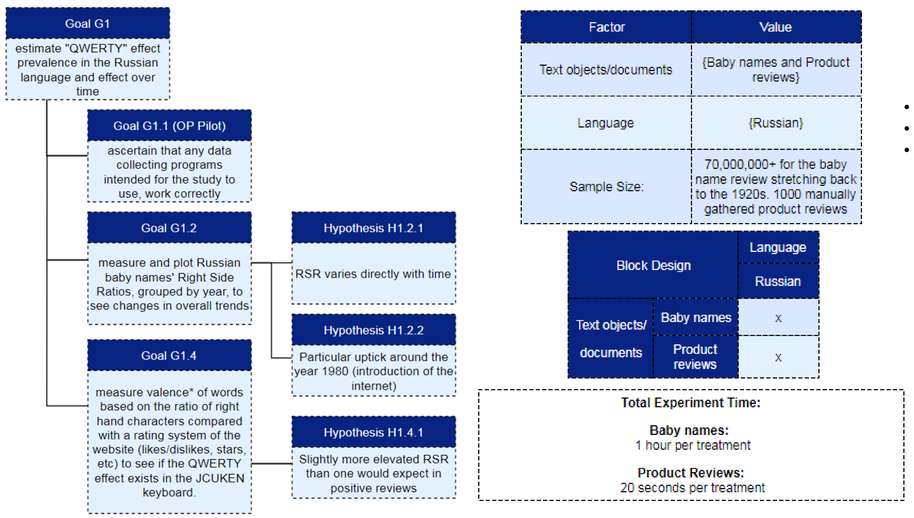
Gathering data

- 70,000,000 names using the Vkontakte API (social media site). Grouping names by their birth year [4].
- Gather 1,000 product reviews manually and group product reviews by their star rating.

Developing data programs

- Compute the average Right Side Ratio for each year or star rating.
- Plot the calculated data points with a line of best fit and observe trends in the data.

V. Goal Tree



VI. Experiment Design

Factor	Value
Text objects/documents	{Baby names and Product reviews}
Language	{Russian}
Sample Size:	70,000,000+ for the baby name review stretching back to the 1920s. 1000 manually gathered product reviews

Block Design		Language
		Russian
Text objects/ documents	Baby names	x
	Product reviews	x

Total Experiment Time:

Baby names:
1 hour per treatment

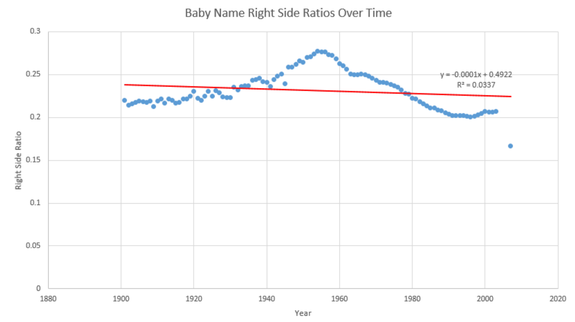
Product Reviews:
20 seconds per treatment

VII. Results and Conclusions

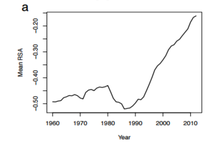
Name Frequency Study:

Example of data:

1	Павел	Дуров	durov	10.10.1984	2	1
14	Андрей	Городецкий	thz	29.6.1982	1719	49
2	Александра	Владимирова	alexandra	-	0	0
28	DELETED		id28	-	-	-
39	David	Mirelli	id39	31.3	3296	8



Previous study [2]:



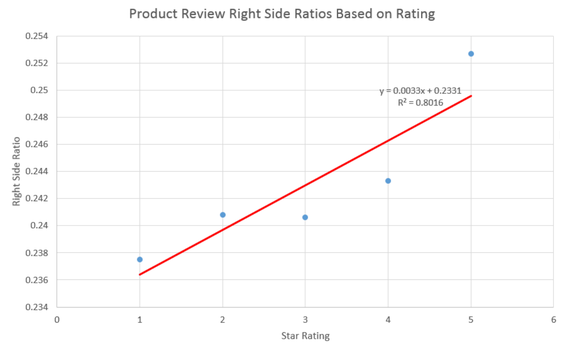
Mean RSR of names given to at least 100 children per year from 1960–2012 (Cassanto, Jasmine, Brookshire, and Gijssels).

- 74,868,169 names were grouped together by their birth year, and an average RSR for each birth year was calculated
- Over time the slope of the line of best fit nearly equals 0, showing no trend upwards in Right Side Ratios overtime for the Russian language. This contrasts with Roman-character languages which observed a general trend upward in RSRs over time.
- However the data points do follow a general pattern and are not just randomly distributed above and below the line of best fit

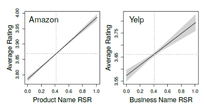
Product Review Study:

Example of data:

	A	B
1	Очень хорошие и твердые обувь! я люблю их. они оснащены очень хорошо, и они ищут Предположить. хорошая цена, доставка была быстрой, и продавец также очень хорошие и доброжелательные! мы в нидерландах мы скажем ЕЕЛ ТОРРЕРТЛЕЛЛ Б	5
2	Быстрая доставка! купон в подарок, упаковано хорошо. доволен очень. кеды без запаха!	5
3	Кедосы супер!!!нитки не торчат,запаха резины нет вообще!на ноге сидят просто шикарно,очень удобные.В подарок продавец в подарок положил носки и купон на 25.Доставка всего 16 дней до ЮФО.Почта России как всегда отличилась.Пакет был над	5
4	Отличная обувь. Для моего мужа на 2 сезона самое оно. Доставка супер быстрая. На 41 размер самое оно. Спасибо	5
5	Спасибо продавец топ совершенные обувь любовь супер спасибо. быстрая доставка Я рекомендую продавца...	5



Previous study [1]:



- 1000 product reviews were manually gathered and grouped by their associated star rating
- An average Right Side Ratio was then calculated for each star rating group and those averages were plotted.
- According to the calculated values, as the average star rating increases so does the average Right Side Ratio with a high R-squared value of 0.8016.

IX. Directions for Future Work

I was unable to find further academic sources for Russian name frequency such as census data

- If data were to obtained, then further analysis could explain the peaks and valleys observed in the Baby Name experiment

Web scrapping could be used to gather a larger sample size of product reviews to further analyze the relationship between RSR and star rating (Valence) in the Russian language.

Sources:
[1] Garcia, David and Strohmaier, Markus. “The QWERTY Effect on the Web: How Typing Shapes the Meaning of Words in Online Human-Computer Interaction.” WWW ’16, 25th International Conference on World Wide Web, 2016, pp. 661-670.
[2] Cassanto, Daniel and Jasmin, Kyle and Brookshire, Geoffrey and Gijssels, Tom. “The QWERTY Effect: How typing shapes word meanings and baby names.” Proceedings of the 36th Annual Conference of the Cognitive Science Society, 2014.
[3] Jasmin, Kyla and Casasanto, Daniel. “The QWERTY Effect: How typing shapes the meanings of words.” Psychonomic Bulletin and Review, 2012, pp. 499-504.
[4] olBaa. (2017, August). Data set from VK API (Users.get method). Unpublished raw data.