NATIONAL RESEARCH UNIVERSITY

Quantitative Analysis of Narratives

НУГ Seminar 10
Online 21 May 2020

Dr. Caroline Schlaufer, cschlaufer@hse.ru

Today's objectives

- Present our first results
- Understand how we analyze narratives statistically. Get to know some basic tests of descriptive statistics.
  - Distribution of data: chi-squared test ($\chi 2$), t-tests
  - Correlation of data: Cramer's V
  - Inter-Coder Reliability

- Discuss next steps

Inter-coder reliability

- We randomly selected 20% of all texts to be coded by a second person
- I analyzed two aspects:
  - Percentage agreement: in how many cases coders agreed (%)
  - Krippendorff's alpha: a statistical measure for coder agreement that also considers agreement by chance
    - Higher than 0.8: good
    - Between 0.67 and 0.8: satisfactory
    - Below 0.67: data is not reliable, cannot be used

Inter-coder reliability

	Percentage agreement	Krippendorff's alpha
hero	82.5 %	0.71
villain	91.2 %	0.84
victim	92.1 %	0.73
beneficiary	86.8 %	0.71
plot	86.8 %	0.74
benefit	90.4 %	0.83
cost	93.0 %	0.80
cause	89.5 %	0.80

Inter-coder reliability

- Different coders did not always code in the same way but in the 80-90% of cases
- Inter-coder reliability is not perfect, but sufficient to continue the analysis

Reminder: research question

Our research question was:

- What policy narratives and narrative strategies are used on both sides of the debate?

Reminder: hypotheses

The government side will:

1. Strategically use heroes and beneficiaries
2. Use control stories
3. Contain conflict: Diffuse benefits / concentrate costs

The other side will:

1. Strategically use villains and victims
2. Use decline or illusion stories
3. Expand conflict: concentrate benefits / diffusing costs

Sides of the debate

	Government	Opposed
Renovation	Governmental websites	Facebook posts of Москвичи против сноса
Transport	Governmental websites and media Experts: strelka, avtostat	NGOs: Gorodskie proiekti, TAMA Bloggers: Varlamov, Gershman, Katz Politicians: Navalny, Besedina
Waste	Governmental actors, websites, media and RT invest	NGOs: Greenpeace, rsbor, Experts: Anna Garkuscha Politicians: Green party, Gudkov, Navalny, Gennady Zuyganov, Lyaskin Bloggers: Varlamov, Gershman

Results: character use

- I calculated how many characters are used on average per text = mean use of characters.
- mean = mathematical average

	Governmental side Mean use of characters	Opposing side Mean use of characters
Heroes	1.35	0.61
	t = 14,203; p < 0.001	
Villains	0.06	1.41
	t = -29,837; p < 0.001	
Victims	0.09	1.30
	t = -24,034; p < 0.001	
Beneficiaries	1.01	0.42
	t = 13,181; p < 0.001	

T-test and p-value

- I tested whether the means are significantly different: t-test.
- The t-test determines if the means of two sets of data are significantly different from each other
- p-value (for all statistical tests) needs to be close to 0. If it is close to 0, your test is statistically significant.

Results: character use

- 1st hypotheses can be confirmed: the government side uses statistically significantly more heroes and beneficiaries, while the opposing side uses statistically significantly more villains and victims.

Heroes

Governmental side			Opposing side		
Government	289	58.62%	NGOs	52	22.22%
Sobyanin	110	22.31%	citizens	50	21.37%
Business	35	7.10%	Opposition politician	38	16.24%
Citizens	31	6.29%	Government	31	13.25%
NGOs	11	2.23%	Business	15	6.41%
experts	5	1.01%	Bloggers	11	4.70%
media	3	0.61%	Sobyanin	9	3.85%
opposition politician	3	0.61%	experts	7	2.99%
activists	2	0.41%	activists	6	2.56%
others	4	0.81%	media	5	2.14%
	493	100.00%	urban environm	4	1.71%
			foreign countries	4	1.71%
			others	2	0.85%
				234	100.00%

Villains

Governmental side				Opposing side		
citizens	8	32.00%		government	278	50.18%
business	5	20.00%		business	108	19.49%
others	5	20.00%		Sobyanin	83	14.98%
government	3	12.00%		police	21	3.79%
opposition politician	2	8.00%		garbage mafia	18	3.25%
media / journalists	1	4.00%		citizens	15	2.71%
activists	1	4.00%		media / journalists	9	1.62%
	25	100.00%		opposition politician	8	1.44%
				experts	5	0.90%
				others	5	0.90%
				bloggers	3	0.54%
				urban environment	1	0.18%
					554	100.00%

Victims

Governmental side				Opposing side		
environment	14	41.18%		citizens	254	50.20%
citizens	13	38.24%		urban environment	75	14.82%
opposition politician	2	5.88%		environment	72	14.23%
urban environment	1	2.94%		opposition politician	26	5.14%
Sobyanin	1	2.94%		NGO	19	3.75%
government	1	2.94%		buiness	17	3.36%
others	1	2.94%		media	14	2.77%
	34	100.00%		government	12	2.37%
				activist	6	1.19%
				others	5	0.99%
				bloggers	2	0.40%
				experts	2	0.40%
				Sobyanin	2	0.40%
					506	100.00%

Beneficiaries

Governmental side				Opposing side		
citizens	280	75.07%		citizens	45	27.95%
environment	38	10.19%		Business	45	27.95%
urban environment	38	10.19%		Government	29	18.01%
Business	6	1.61%		urban environment	13	8.07%
opposition politician	6	1.61%		media/journalists	8	4.97%
Government	3	0.80%		garbage mafia	7	4.35%
media/journalists	1	0.27%		opposition politician	7	4.35%
NGO	1	0.27%		environment	4	2.48%
	373	100.00%		Sobyanin	2	1.24%
				NGO	1	0.62%
					161	100.00%

Plots

	Governmental side	Opposing side
story of decline	0 (0%)	165 (46.0%)
story of control	360 (99,7%)	101 (28,1%)
Illustion stories	0 (0%)	68 (18,9%)
stymied progress	1 (0,3%)	25 (7,0%)
Total	361 (100%)	359 (100%)

$\chi 2 = 400{,}663$, $p < 0.001$

Cramer-V = 0.746, $p < 0.001$

44 texts were coded as having no plot, they were excluded from the analysis

Chi-square test

- If we'd like to know if 2 categorical variables are associated or not, our first step is to do a chi-square ($\chi 2$) test.
- A chi-squared test is used to determine whether there is a statistically significant difference between the expected and observed frequencies in one or more categories of a contingency table.

Chi-square test

- The chi-square test only tells us that there is an association, but it does not say whether the variables are strongly associated. For this, we calculate a correlation among categorial variables ➡ Cramer's V.

Cramer V

- Cramer's V measures how strongly two categorical variables are associated.
- V = 0: no association, V = 1: perfect association
- Usually, if Cramer's V > 0.5, the association of the two variables is considered as strong.

Results: Plots

- 2nd hypotheses can be partly confirmed: the government side uses statistically significantly more control stories, while the opposing side uses statistically significantly more decline and illusion stories, but also control stories.

Costs

	Governmental side	Opposing side
the narrative does not discuss any costs	352 (94,9%)	213 (54,2%)
the narrative says that only a few have to pay	14 (3,8%)	24 (6,1%)
the narrative says that many need to pay	5 (1,3%)	156 (39,7%)
	371 (100%)	393 (100%)

$\chi 2 = 177{,}963;\ p < 0.001$

Cramer V = 0.483; p < 0.001

Benefits

	Governmental side	Opposing side
the narrative does not discuss any benefits	58 (15,6%)	209 (53,2%)
the narrative says that only a few benefit from the policy	43 (11,6%)	110 (28,0%)
the narrative says that many benefit from the policy	270 (72,8%)	74 (18,8%)
	371 (100,0%)	393 (100,0%)

$\chi 2 = 225{,}965;\ p < 0.001$

Cramer V = 0.544; p < 0.001

Causes

	Governmental side	Opposing side
Intentional	5 (12,2%)	248 (73,4%)
Inadvertent	12 (29,3%)	55 (16,3%)
Accidental	8 (19,5%)	0 (0,0%)
Mechanical	16 (39,0%)	35 (10,4%)
	41 (100,0%)	338 (100,0%)

$\chi 2 = 112{,}279;\ p < 0.001$

Cramer V = 0.544; p < 0.001

The statistical tests

- Krippendorff Alpha: to measure inter-coder reliability
- T-tests: to test if the difference of the means of two groups are statistically significant
- Chi-square: to test if the differences of allocations of categorial varaibles among groups are statistically significant
- Cramer's V: to measure how strong two categorial variables are associated

Next steps for our NUG

- Continue data analysis
  - Problems and policy solutions, devil-angel shift
  - Analysis of each case, comparison between the cases
  - Add qualitative examples
- Write (minimum) 4 Articles → work in groups under the leadership of first authors
- Define further research questions and analyses → qualitative methods

Case	Authors	Conferences	Submission
Transport	Artem, Tatiana G, Dmitri, Mariia	ECPR 26-28 August	European Policy Analysis special issue on NPF September
Waste 1 (Waste 2)	Caroline, Tatiana Kh, Marina, Viktoria, Sanjay (?) Marina, Tatiana Kh, Viktoria, Caroline	(ECPR 26-28 August) APSA РАПН	European Policy Analysis special issue on NPF September
Renovation	Victor, Ekatarina G, Svetlana, Aleksaner	ECPR 26-28 August	Policy&Politics or Critical Policy Studies
Comparison	Caroline, Dilyara, Ekatarina Zh, Marina	ECPR 26-28 August (EPGA 2-4 September)	Policy Studies Journal or Review of Policy Research

Next seminars

- Qualitative narrative analysis → Victor
- July-August, online: train for ECPR → all groups make a presentation, everybody reads and comments the articles (1-3 seminars)
- September, hopefully in Moscow → a brainstorm-seminar on further research questions and analysis

April conference

- This Saturday, 23 May at 1 pm. Register on the link I sent you.

- Thanks to everybody for the excellent work you did!
- I hope you all continue in the NUG – if not let me know as soon as possible