

# LEXICAL RESEARCH IN RUSSIAN: ARE MODERN CORPORA FLEXIBLE ENOUGH?

Natalia Y. Lukashevich ([natalukashevich@mail.ru](mailto:natalukashevich@mail.ru)), Moscow State University

Eduard S. Klyshinsky ([klyshinsky@mail.ru](mailto:klyshinsky@mail.ru)), Keldysh IAM RAS

Irina M. Kobozeva ([kobozeva@list.ru](mailto:kobozeva@list.ru)), Moscow State University

# Outline

- Search options in Russian language corpora
- *Gordy* in news and newspaper texts
- Comparison of results

# General issues

- Register variation (Belikov et al, 2013):
  - allows to account for heterogeneity of language data
  - is relevant not only for sociolinguistic studies, but also for a whole lot of linguistic tasks including comparative studies of texts from different genres (Belikov et al, 2012)
- Genre (and style) granularity in a corpus?

## Flexibility in a corpus

- The more a corpus allows tailoring the initial data and search in accordance with the needs of a particular study, the more varied tasks can be solved with its help.
- Until quite recently the existing corpora for Russian did not offer a way to create a subcorpus of your choice.
- At the moment several corpora provide this opportunity to a researcher: GICR, RNC.

# General Internet Corpus of Russian (GICR)

- GICR ([www.webcorpora.ru](http://www.webcorpora.ru)) was developed as a resource that allows to apply the method of segmental statistics in a wide range of linguistic tasks.
- Its interface makes it possible to select not only the corpus segment, but also the type of text sources according to the list of segment-specific attributes (such as the author's year of birth, place of birth, gender for blogs, the name of the source for news, etc).

## Russian National Corpus (RNC)

- has previously only allowed to select text sources from the main subcorpus on the basis of certain features (year of publication, etc.);
- added a similar possibility in other subcorpora (including its newspaper subcorpus) quite recently.

How is *gordy* 'proud' used in news texts?

- (1000 contexts of RNC's main subcorpus)
- RNC's newspaper subcorpus
- GICR news segment
- a self-designed collection of texts (CoSyCo news subcorpus)

# RNC's newspaper subcorpus (as of 12/12/2015)

<b>Total</b>	<b>173.5 mln (words)</b>	<b>100%</b>
Izvestia	9,282.250	5.35%
Komsomolskaya Pravda	44,867.100	25.86%
Novy region	25,174.850	14.51%
RBK Daily	26,424.050	15.23%
RIA Novosti	15,545.600	8.96%
Sovetsky sport	12,977.800	7.48%
Trud	39,228.350	22.61%



# GICR: news segment

(as of 03/02/2016)

<b>Source</b>	<b>851 mln (words)</b>	<b>100%</b>
lenta	110,418.290	12.97%
regnum	218,025.910	25.60%
ria	337,669.976	39.65%
rosbalt	185,456.412	21.78%

# CoSyCo news subcorpus: news

<b>Total</b>	<b>982.2 mln (words)</b>	<b>100%</b>
lenta.ru (lenta)	71,300.115	7.26%
RBC (rbc)	61,933.721	6.31%
RIA Novosti (ria)	409,971.920	41.74%
Nezavisimaya gazeta (ng)	48,923.879	4.98%
Vzglyad (vz)	72,370.767	7.37%
Rossiyskaya gazeta(rg)	71,467.194	7.28%
Commersant (commersant)	140,585.843	14.31%
Polit.ru (polit)	49,697.364	5.06%
Utro.ru(utro)	45,770.623	4.66%
Ibusiness.ru (ibusiness)	10,131.894	1.03%

# CoSyCo news subcorpus: IT news

<b>Total</b>	<b>113 mln (words)</b>	<b>100%</b>
Membrana (membrana)	7,391.018	6.40%
CNews (cnews)	35,830.813	31.04%
Computerra (computerra)	28,068.619	24.32%
Compulenta (compulenta)	16,204.248	14.04%
PCWeek (pcweek)	27,924.230	24.19%

# What is CoSyCo?

- A set of texts tagged by the *Crosslator* dictionary
- A software tool based regular expressions tackling morphological ambiguity
- A COrpus of SYntac CO-occurences extracted from the mentioned set of texts

# What is CoSyCo?

Существительные:

Сортировать по [Алфавиту](#)

[Частотности](#) | Частота от

Формы существительного

Для прилагательных

показывать

[Начальную форму](#) [Форму слова](#)

Прилагательные:

Сортировать по [Алфавиту](#)

[Частотности](#) | Частота от

АДМИНИСТРАЦИЯ  
17121

АДМИНИСТРАТОР 469

АДМИНИСТРАТОРША  
322

АДМИРАЛ 293

АДМИРАЛЬША 10

АДМИНИСТРАЦИЕЙ 34

АДМИНИСТРАЦИЕЮ 3

АДМИНИСТРАЦИЮ 6863

АДМИНИСТРАЦИЯ 9857

АДМИНИСТРАЦИЯМИ 364

МЕСТНЫЙ 1726

ГОРОДСКОЙ 1666

АМЕРИКАНСКИЙ 1278

НОВЫЙ 967

ОБЛАСТНОЙ 731

ПРЕЗИДЕНТСКИЙ 635

Выбрать корпус:

Уже в первом месте служенья, в селе Кривополянье, он приобрел репутацию человека беспокойного и неуживчивого, вследствие трений с местною администрацией.

[Источник](#)

Незамерзаемость Екатерининской гавани в этой губе засвидетельствована местною администрацией и командами судов, которые почти ежегодно в ней остаются на зимовку.

[Источник](#)

# Why CoSyCo?

- A set of subcorpora in hand
- Convenient software tool in hand
- Lucky possibility to test the developed toolset (yes, it is still under construction)

## Gordy in newspaper texts: queries used

- (RNC's main subcorpus: 1000 entries with *gordy* 'proud')
- RNC's newspaper subcorpus:  
gordy + (in)animate noun
- GICR news segment:  
gordy + (in)animate noun  
(with automatic deduplication turned on)
- CoCyCo news subcorpus:  
(ГОРДЫЙ;adj; case+, gender+, number+)(;adj; case+, gender+, number+)\*(;noun; case+, gender+, number+)

## *5 meanings of gordy* ‘proud’:

1) *gordy X*  $\approx$  a person X whose behavior shows that (s)he has a sense of dignity and self-respect:

- (1) *(Katya) suddenly felt doubly happy: her beloved was not an ordinary man, no, he was tough, proud and pure.* (RNC, E.Kazakevich, Zvezda)
- (Катя) вдруг почувствовала себя вдвойне счастливой: ее любимый был не обычный человек, нет, он суровый, гордый и чистый. (Э. Г. Казакевич. Звезда (1946))



## *5 meanings of gordy* ‘proud’:

2) *gordy X* ≈ a person X who is feeling pleased with the fact that smth that (s)he (or someone associated with him/her) owns or smth (s)he (or the associates) achieved should make other people think better of him/her or rank him/her higher:

- (2) *Alevtina is proud that she earns her living herself and does not depend on anybody...[\[1\]](#)*  
(RNC, V.Makanin, Otdushina)

Алевтина горда тем, что зарабатывает на жизнь сама и ни от кого не зависит .... (В. Маканин, Отдушина (1977))

## *5 meanings of gordy* 'proud':

- 3) *gordy X*  $\approx$  a person X who thinks of oneself as being better than other people and treats them with contempt because of that:
- (3) *You had better go and stay with the guests, or they will think you are too proud.* (RNC, A.Chekhov, *V rodnom uglu*)
  - Ты бы посидела с гостями, а то подумают, что ты гордая. (А. П. Чехов. В родном углу (1897))

## *5 meanings of gordy 'proud':*

- 4) (figurative) majestic, stately
- (4) *High, proud celestial mountain peaks were glimmering golden in the sunset sky.* (RNC, V.Skripkin, Tinga)
  - Высокие, гордые вершины небесных гор румянились в закатном небе. (В. Скрипкин. Тинга // «Октябрь», 2002)
- 5) (figurative) sublime, lofty, elevated:
- (5) *Fyodor thought...with proud, joyous energy, with passionate impatience, he was already looking for the creation of something ..* (RNC, V.Nabokov. The Gift (M. Scammel, V.Nabokov, 1962))
  - Федор Константинович ...с какой-то радостной, гордой энергией... уже искал создания чего-то нового... (В. В. Набоков, Дар (1935-1937))

## *Semantic classes of co-occurring nouns*

- 1. Nouns denoting persons: (>>> *gordy1/3*)
- - a person in general or a male/female person:
  - *gordaya devushka* 'a proud girl'
- - a person according to family status
  - *gordy otets* 'a proud father'
- - a person according to nationality/ethnicity:
  - *gordy amerikanets* 'a proud American'
- - a person according to social status
  - *gordy korol'* 'a proud king'
- - a big group of people (community)
  - *gordy narod* 'proud people'
- etc.

# *Semantic classes of co-occurring nouns*

2. Nouns related to the situation of a person being proud (in sense 2): (>>> *gordy2*)

- nouns denoting a person as a possessor of smth or an agent of some deed:

*gordy pobeditel'* 'a proud winner'

- nouns denoting emotions experienced by a person who feels proud (in the sense 2):

*gordaya radost'* 'proud joy'

3. Nouns referring to an object which is similar to a proud person in some respect: (>>> *gordy4, metaphoric transfer*)

- an inanimate object:

*gordaya bashnya* 'a proud tower'

- an animal, a bird:

*gordy oryol* 'a proud eagle'

## *Semantic classes of co-occurring nouns*

### 4. Nouns denoting features of a person or results of human activity:

(>>> *gordy5*, transfer by metonymy)

- general characteristics of a person (as a subject of mental and social activity):

*gordiy um* 'a proud mind'

- emotional states, feelings, personality traits:

*gordoe spokoistvie* 'proud tranquility'

*gordoe prezrenie* 'proud contempt'

- features and qualities of appearance and movement:

*gordaya osanka* 'proud demeanor'

- psychological states, processes, and 'products' of mental work

*gordaya mechta* 'a proud dream'

- representational objects (linguistic units and expressions)

*gordoye imya* 'a proud name'

*gordaya nadpis'* 'a proud inscription'

etc.

Names of representational objects with *gordy* in news and newspaper texts (17 nouns):

	<i>imya</i> 'name'			total		
	a	b,%	c,%	a	b,%	c,%
RNC 1000	14	3,91		31	8,7	
<b>RNC news</b>	<b>60</b>	<b>5,89</b>		<b>149</b>	<b>14,7</b>	
<b>GICR news</b>	<b>39</b>	<b>4,63</b>		<b>126</b>	<b>15,0</b>	
lenta*	1	1,18	0,12	15	17,65	1,76
rian	13	4,71	1,53	39	14,13	4,58
regnum	18	6,72	2,11	37	13,81	4,34
rosbalt	7	3,14	0,82	35	15,70	4,11

# CoSyCo news

	<i>imya</i> 'name'			total		
	a	b,%	c,%	a	b,%	c,%
commerciant	14	5,49	0,70	47	18,4	2,35
rian	28	4,03	1,40	73	10,5	3,64
rg	14	6,06	0,70	34	14,7	1,70
utro	3	1,96	0,15	26	17,0	1,30
vz	5	2,75	0,25	33	18,1	1,65
nezavisimaya	14	6,36	0,70	29	13,5	1,45
lenta**	1	1,89	0,05	2	3,8	0,10
polit	1	1,28	0,05	4	5,2	0,20
rbc	2	2,13	0,10	18	20,2	0,90
ibusiness	3	6,82	0,15	9	22,5	0,45
<b>AVG CoSyCo NEWS</b>	<b>85</b>		<b>4,24</b>	<b>275</b>		<b>13,72</b>



# CoSyCo IT news

	<i>imya</i> 'name'			total		
	a	b,%	c,%	a	b,%	c,%
cnews	2	9,09	0,48	2	9,09	0,48
compulenta	6	13,95	1,45	11	23,26	2,66
computerra	39	14,18	9,42	87	26,55	21,01
membrana	3	9,68	0,72	7	16,14	1,69
PC week	2	4,65	0,48	6	13,96	1,45
<b>AVG IT NEWS</b>	<b>52</b>		<b>12,56</b>	<b>113</b>		<b>27,29</b>

Names of representational objects with *gordy* in news and newspaper texts (17 nouns):

	<i>imya</i> 'name'			total		
	a	b,%	c,%	a	b,%	c,%
RNC 1000	14	3,91		31	8,7	
<b>RNC news</b>	<b>60</b>	<b>5,89</b>		<b>149</b>	<b>14,7</b>	
<b>GICR news</b>	<b>39</b>	<b>4,63</b>		<b>126</b>	<b>15,0</b>	
<b>lenta*</b>	<b>1</b>	<b>1,18</b>	<b>0,12</b>	<b>15</b>	<b>17,65</b>	<b>1,76</b>
rian	13	4,71	1,53	39	14,13	4,58
regnum	18	6,72	2,11	37	13,81	4,34
rosbalt	7	3,14	0,82	35	15,70	4,11

# CoSyCo news

	<i>imya</i> 'name'			total		
	a	b,%	c,%	a	b,%	c,%
commerciant	14	5,49	0,70	47	18,4	2,35
rian	28	4,03	1,40	73	10,5	3,64
rg	14	6,06	0,70	34	14,7	1,70
utro	3	1,96	0,15	26	17,0	1,30
vz	5	2,75	0,25	33	18,1	1,65
nezavisimaya	14	6,36	0,70	29	13,5	1,45
<b>lenta**</b>	<b>1</b>	<b>1,89</b>	<b>0,05</b>	<b>2</b>	<b>3,8</b>	<b>0,10</b>
polit	1	1,28	0,05	4	5,2	0,20
rbc	2	2,13	0,10	18	20,2	0,90
ibusiness	3	6,82	0,15	9	22,5	0,45
<b>AVG CoSyCo NEWS</b>	<b>85</b>		<b>4,24</b>	<b>275</b>		<b>13,72</b>

# CoSyCo IT news

	<i>imya</i> 'name'			total		
	a	b,%	c,%	a	b,%	c,%
cnews	2	9,09	0,48	2	9,09	0,48
compulenta	6	13,95	1,45	11	23,26	2,66
<b>computerra</b>	<b>39</b>	<b>14,18</b>	<b>9,42</b>	<b>87</b>	<b>26,55</b>	<b>21,01</b>
membrana	3	9,68	0,72	7	16,14	1,69
PC week	2	4,65	0,48	6	13,96	1,45
<b>AVG IT NEWS</b>	<b>52</b>		<b>12,56</b>	<b>113</b>		<b>27,29</b>

Names of representational objects with *gordy* in news and newspaper texts:

	<b><i>zvanie</i> 'rank'</b>			<b>total</b>		
	a	b,%	c,%	a	b,%	c,%
RNC 1000	5	1,40		31	8,7	
<b>RNC news</b>	<b>42</b>	<b>4,12</b>		<b>149</b>	<b>14,7</b>	
<b>GICR news</b>	<b>47</b>	<b>5,58</b>		<b>126</b>	<b>15,0</b>	
lenta*	4	4,71	0,47	15	17,65	1,76
rian	13	4,71	1,53	39	14,13	4,58
regnum	12	4,48	1,41	37	13,81	4,34
rosbalt	18	8,07	2,11	35	15,70	4,11

# CoSyCo news

	<b><i>zvanie</i> 'rank'</b>			<b>total</b>		
	a	b,%	c,%	a	b,%	c,%
commerciant	19	7,45	0,95	47	18,4	2,35
rian	18	2,59	0,90	73	10,5	3,64
rg	11	4,76	0,55	34	14,7	1,70
utro	9	5,88	0,45	26	17,0	1,30
vz	14	7,69	0,70	33	18,1	1,65
nezavisimaya	8	3,64	0,40	29	13,5	1,45
lenta**	-	-	-	2	3,8	0,10
polit	-	-	-	4	5,2	0,20
rbc	6	6,38	0,30	18	20,2	0,90
ibusiness	1	2,27	0,05	9	22,5	0,45
<b>AVG CoSyCoNews</b>	<b>86</b>		<b>4,29</b>	<b>275</b>		<b>13,72</b>

# CoSyCo IT news

	<i><b>zvanie 'rank'</b></i>			total		
	a	b,%	c,%	a	b,%	c,%
cnews	-	-	-	2	9,09	0,48
compulenta	1	2,33	0,24	11	23,26	2,66
computerra	13	4,73	3,14	87	26,55	21,01
membrana	1	3,23	0,24	7	16,14	1,69
PC week	1	2,33	0,24	6	13,96	1,45
<b>AVG IT NEWS</b>	<b>16</b>		<b>3,86</b>	<b>113</b>		<b>27,29</b>

Names of representational objects with *gordy* in news and newspaper texts:

	<i>nazvanie</i> 'name'			total		
	a	b,%	c,%	a	b,%	c,%
RNC 1000	10	2,79		31	8,7	
<b>RNC news</b>	<b>32</b>	<b>3,14</b>		<b>149</b>	<b>14,7</b>	
<b>GICR news</b>	<b>27</b>	<b>3,21</b>		<b>126</b>	<b>15,0</b>	
lenta*	5	5,88	0,59	15	17,65	1,76
rian	11	3,99	1,29	39	14,13	4,58
regnum	3	1,12	0,35	37	13,81	4,34
rosbalt	8	3,59	0,94	35	15,70	4,11



# CoSyCo news

	<i><b>nazvanie 'name'</b></i>			<b>total</b>		
	a	b,%	c,%	a	b,%	c,%
commerciant	9	3,53	0,45	47	18,4	2,35
rian	21	3,03	1,05	73	10,5	3,64
rg	8	3,46	0,40	34	14,7	1,70
utro	13	8,50	0,65	26	17,0	1,30
vz	10	5,49	0,50	33	18,1	1,65
nezavisimaya	6	2,73	0,30	29	13,5	1,45
lenta**	1	1,89	0,05	2	3,8	0,10
polit	2	2,56	0,10	4	5,2	0,20
rbc	10	10,64	0,50	18	20,2	0,90
ibusiness	5	11,36	0,25	9	22,5	0,45
<b>AVG CoSyCoNews</b>	<b>85</b>		<b>4,24</b>	<b>275</b>		<b>13,72</b>

# CoSyCo IT news

	<i>nazvanie</i> 'name'			total		
	a	b,%	c,%	a	b,%	c,%
cnews	-	-	-	2	9,09	0,48
compulenta	3	6,98	0,72	11	23,26	2,66
computerra	21	7,64	5,07	87	26,55	21,01
membrana	1	3,23	0,24	7	16,14	1,69
PC week	3	6,98	0,72	6	13,96	1,45
<b>AVG IT NEWS</b>	<b>28</b>		<b>6,76</b>	<b>113</b>		<b>27,29</b>

Names of states/feelings with *gordy* in news and newspaper texts:

	<b><i>odinochestvo</i></b> <b>'loneliness'</b>			<b>total</b>		
	a	b,%	c,%	a	b,%	c,%
RNC 1000	32	8,94		51	14,2	
<b>RNC news</b>	<b>174</b>	<b>17,08</b>		<b>195</b>	<b>19,1</b>	
<b>GICR news</b>	<b>137</b>	<b>16,27</b>		<b>168</b>	<b>20,0</b>	
<b>lenta*</b>	<b>12</b>	<b>14,12</b>	<b>1,41</b>	<b>14</b>	<b>16,47</b>	<b>1,64</b>
rian	52	18,84	6,10	57	20,65	6,69
regnum	42	15,67	4,93	59	22,01	6,92
rosbalt	32	14,35	3,76	38	17,04	4,46

# CoSyCo news

	<b>odinochestvo 'loneliness'</b>			<b>total</b>		
	a	b,%	c,%	a	b,%	c,%
commerciant	35	13,73	1,75	41	16,1	2,05
rian	73	10,52	3,64	81	11,7	4,04
rg	43	18,61	2,15	56	24,2	2,79
utro	38	24,84	1,90	40	26,1	2,00
vz	28	15,38	1,40	30	16,5	1,50
nezavisimaya	34	15,45	1,70	44	20,5	2,20
<b>lenta**</b>	-	-	-	-	-	-
polit	8	10,26	0,40	11	14,1	0,55
rbc	24	25,53	1,20	24	25,5	1,20
ibusiness	3	6,82	0,15	<b>3</b>	<b>7,5</b>	<b>0,15</b>
<b>AVG CoSyCoNews</b>	<b>286</b>		<b>14,27</b>	<b>330</b>		<b>16,47</b>

# CoSyCo IT news

	<b><i>odinochestvo</i></b> <b>'loneliness'</b>			<b>total</b>		
	a	b,%	c,%	a	b,%	c,%
cnews	2	9,09	0,48	3	13,6	0,72
compulenta	14	32,56	3,38	14	32,6	3,38
computerra	29	10,55	7,00	35	12,7	8,45
membrana	7	22,58	1,69	7	22,6	1,69
PC week	13	30,23	3,14	14	32,6	3,38
<b>AVG IT NEWS</b>	<b>65</b>		<b>15,70</b>	<b>73</b>		<b>17,63</b>

Names of big groups of people (communities) with *gordy* in news and newspaper texts (24 nouns):

	<i>narod</i> 'people'			total		
	a	b,%	c,%	a	b,%	c,%
RNC 1000	5	1,40		9	2,5	
<b>RNC news</b>	<b>15</b>	<b>1,47</b>		<b>57</b>	<b>5,6</b>	
<b>GICR news</b>	<b>61</b>	<b>7,24</b>		<b>127</b>	<b>15,1</b>	
lenta*	3	3,53	0,35	9	10,59	1,06
rian	17	6,16	2,00	26	9,42	3,05
regnum	26	9,70	3,05	59	22,01	6,92
rosbalt	15	6,73	1,76	33	14,80	3,87

# CoSyCo news

	<i>narod</i> 'people'			total		
	a	b,%	c,%	a	b,%	c,%
commerciant	5	1,96	0,25	23	9,1	1,15
rian	25	3,60	1,25	48	6,9	2,40
rg	6	2,60	0,30	13	5,6	0,65
utro	10	6,54	0,50	18	11,8	0,90
vz	2	1,10	0,10	9	4,9	0,45
nezavisimaya	5	2,27	0,25	13	6,0	0,65
lenta**	1	1,89	0,05	2	3,8	0,10
polit	5	6,41	0,25	12	15,4	0,60
rbc	1	1,06	0,05	5	5,3	0,25
ibusiness	-	-	-	2	5,0	0,10
<b>AVG CoSyCoNews</b>	<b>60</b>		<b>2,99</b>	<b>145</b>		<b>7,24</b>

# CoSyCo IT news

	<i>narod</i> 'people'			total		
	a	b,%	c,%	a	b,%	c,%
cnews	-	-	-	4	18,18	0,97
compulenta	-	-	-	-	-	-
computerra	2	0,73	0,48	8	2,92	1,92
membrana	-	-	-	-	-	-
PC week	-	-	-	-	-	-
<b>AVG IT NEWS</b>	<b>2</b>		<b>0,48</b>	<b>12</b>		<b>1,93</b>



Names of big groups of people (communities) with *gordy* in news and newspaper texts (24 nouns):

	<b><i>strana</i> 'country'</b>			<b>total</b>		
	a	b,%	c,%	a	b,%	c,%
RNC 1000	4	1,12		9	2,5	
<b>RNC news</b>	<b>16</b>	<b>1,57</b>		<b>57</b>	<b>5,6</b>	
<b>GICR news</b>	<b>21</b>	<b>2,49</b>		<b>127</b>	<b>15,1</b>	
lenta*	1	1,18	0,12	9	10,59	1,06
rian	3	1,09	0,35	26	9,42	3,05
regnum	11	4,10	1,29	59	22,01	6,92
rosbalt	6	2,69	0,70	33	14,80	3,87

# CoSyCo news

	<i>strana</i> 'country'			total		
	a	b,%	c,%	a	b,%	c,%
commerciant	9	3,53	0,45	23	9,1	1,15
rian	11	1,59	0,55	48	6,9	2,40
rg	2	0,87	0,10	13	5,6	0,65
utro	3	1,96	0,15	18	11,8	0,90
vz	2	1,10	0,10	9	4,9	0,45
nezavisimaya	1	0,45	0,05	13	6,0	0,65
lenta**	-	-	-	2	3,8	0,10
polit	3	3,85	0,15	12	15,4	0,60
rbc	4	4,26	0,20	5	5,3	0,25
ibusiness	1	2,27	0,05	2	5,0	0,10
<b>AVG CoSyCoNews</b>	<b>36</b>		<b>1,80</b>	<b>145</b>		<b>7,24</b>

# CoSyCo IT news

	<i>strana</i> 'country'			total		
	a	b,%	c,%	a	b,%	c,%
cnews	4	18,18	0,97	4	18,18	0,97
compulenta	-	-	-	-	-	-
computerra	2	0,73	0,48	8	2,92	1,92
membrana	-	-	-	-	-	-
PC week	-	-	-	-	-	-
<b>AVG IT NEWS</b>	<b>6</b>		<b>1,45</b>	<b>12</b>		<b>1,93</b>

Names of big groups of people (communities) with *gordy* in news and newspaper texts:

	<b><i>gosudarstvo</i></b> <b>'state'</b>			<b>total</b>		
	a	b,%	c,%	a	b,%	c,%
RNC 1000	-	-		9	2,5	
<b>RNC news</b>	<b>5</b>	<b>0,49</b>		<b>57</b>	<b>5,6</b>	
<b>GICR news</b>	<b>15</b>	<b>1,78</b>		<b>127</b>	<b>15,1</b>	
lenta*	-	-	-	9	10,59	1,06
rian	2	0,72	0,23	26	9,42	3,05
regnum	7	2,61	0,82	59	22,01	6,92
rosbalt	6	2,69	0,70	33	14,80	3,87

# CoSyCo news

	<i>gosudarstvo</i> 'state'			total		
	a	b,%	c,%	a	b,%	c,%
commerciant	1	0,39	0,05	23	9,1	1,15
rian	2	0,29	0,10	48	6,9	2,40
rg	1	0,43	0,05	13	5,6	0,65
utro	1	0,65	0,05	18	11,8	0,90
vz	1	0,55	0,05	9	4,9	0,45
nezavisimaya	-	-	-	13	6,0	0,65
lenta**	-	-	-	2	3,8	0,10
polit	1	1,28	0,05	12	15,4	0,60
rbc	-	-	-	5	5,3	0,25
ibusiness	-	-	-	2	5,0	0,10
<b>AVG CoSyCoNews</b>	<b>7</b>	<b>0,35</b>	<b>0,35</b>	<b>145</b>		<b>7,24</b>

More frequently used semantic zones of *gordy* in news and newspaper texts:

<b>Semantic group</b>	<b>Polythematic news sources</b>		<b>IT news sources</b>
<b>nouns naming a person according to nationality/ethnicity</b>	RNC news	5,0%	2,17%
	GICR news	5,7%	
	CoSyCo news	3,94%	
<b>big groups of people</b>	RNC news	5,6%	1,93%
	GICR news	15,1%	
	CoSyCo news	7,24%	
<b>representational objects</b>	RNC news	14,7%	<b>27,29%</b>
	GICR news	15,0%	
	CoSyCo news	13,72%	
<b>desemantised phrasemes</b>	RNC news	19,1%	17,63%
	GICR news	20,0%	
	CoSyCo news	16,47%	