

WORDS

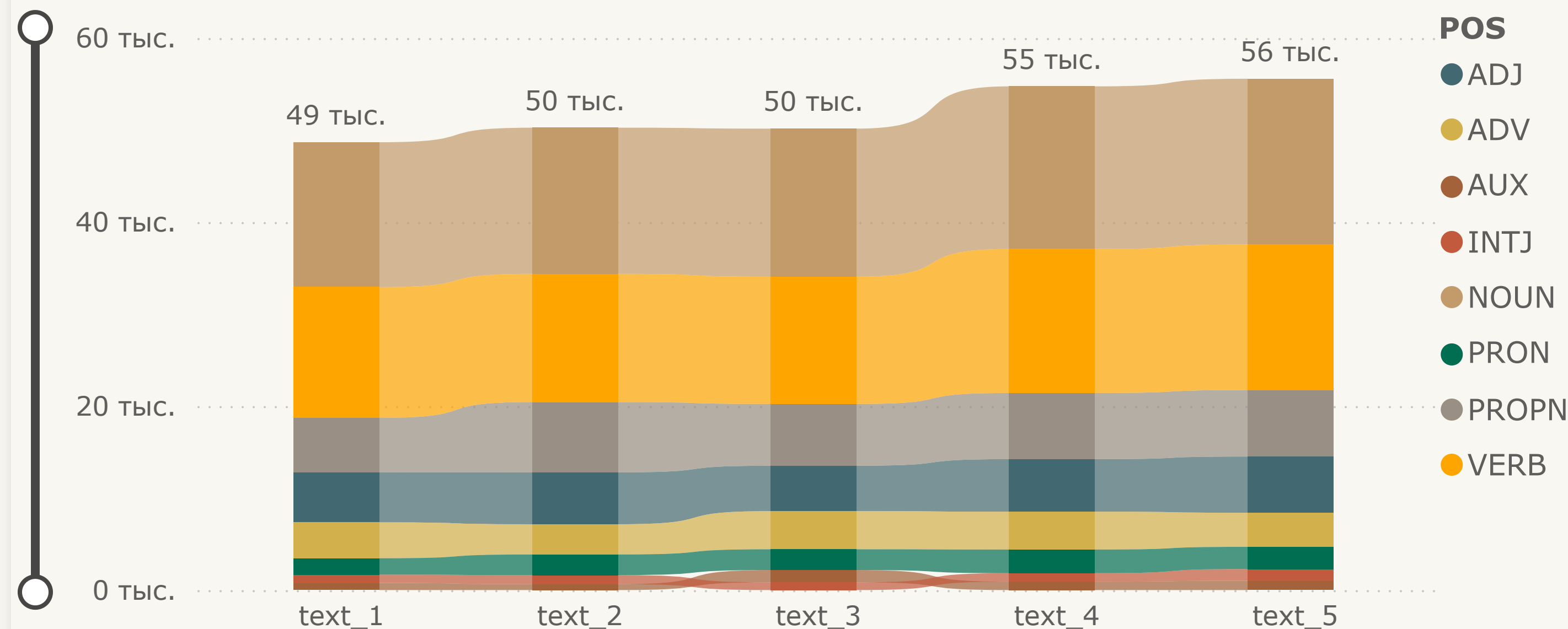
Text

Bce

POS

Bce

Сумма Frequency по Text и POS



Word	Frequency	TF Measure	TF	IDF	TF-IDF
say	9007	0,03	0,17	3,26	0,11
I	8326	0,03	0,15	3,26	0,09
Vimes	4000	0,01	0,07	3,26	0,04
get	2344	0,01	0,04	3,48	0,03
go	2141	0,01	0,04	3,26	0,03
sir	2048	0,01	0,04	6,52	0,02
Carrot	2031	0,01	0,04	4,87	0,03
look	2018	0,01	0,04	5,09	0,02
know	1870	0,01	0,04	3,26	0,02
think	1851	0,01	0,03	3,26	0,02
Bcero	259367	0,96	2,99	23 720,15	1,91



WORDS

Text

Bce



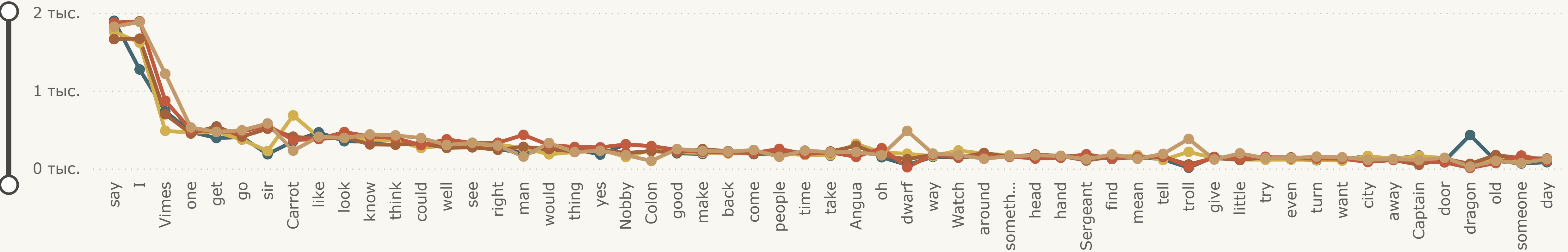
POS

Bce

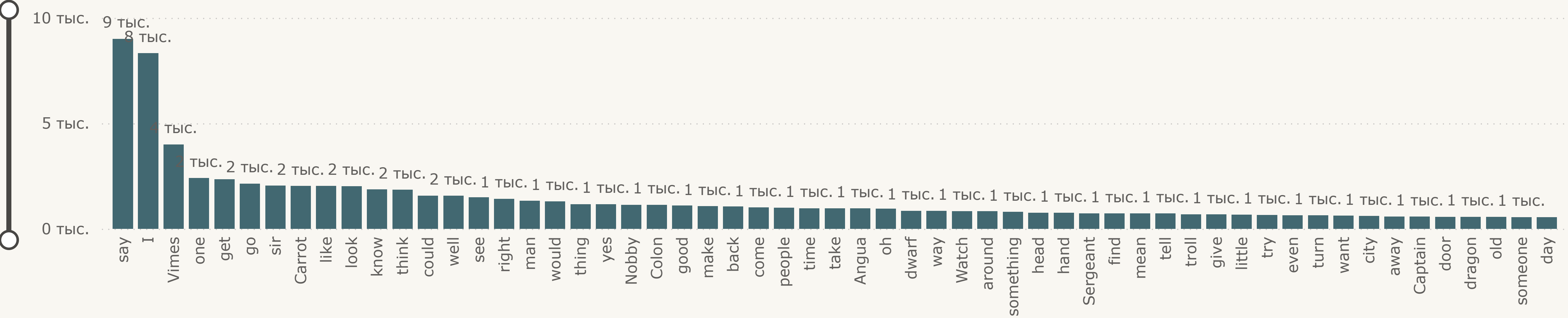


Сумма Frequency по Word и Text

Text text_1 text_2 text_3 text_4 text_5



Сумма Frequency по Word



BIGRAMS

Text

text_2

POS1_POS2

Bce

POS1_POS2 (rus)

Bce

Частоты биграмм + частичечный фильтр

Bigram	w1	w2	Сумма Frequency	Сумма w1 Frequency	Сумма w2 Frequency	POS1_POS2
('say', 'Carrot')	say	Carrot	258	1757	682	VERB_PROPN
('I', 'think')	I	think	146	1616	360	PRON_VERB
('say', 'Angua')	say	Angua	120	1757	315	VERB_PROPN
('Sergeant', 'Colon')	Sergeant	Colon	95	110	197	PROPN_PROPN
('say', 'Vimes')	say	Vimes	89	1757	484	VERB_PROPN
('say', 'Cuddy')	say	Cuddy	86	1757	245	VERB_PROPN
('I', 'mean')	I	mean	85	1616	169	PRON_VERB
('Ankh', 'Morpork')	Ankh	Morpork	83	103	90	PROPN_PROPN
('I', 'know')	I	know	82	1616	372	PRON_VERB
('lance', 'constable')	lance	constable	73	75	112	PROPN_PROPN
('I', 'say')	I	say	72	1616	1757	PRON_VERB
('say', 'Colon')	say	Colon	70	1757	197	VERB_PROPN
('say', 'Nobby')	say	Nobby	67	1757	147	VERB_PROPN
('say', 'I')	say	I	66	1757	1616	VERB_PRON
('say', 'Gaspode')	say	Gaspode	63	1757	151	VERB_PROPN
('Captain', 'Vimes')	Captain	Vimes	62	85	484	PROPN_PROPN
('Corporal', 'Carrot')	Corporal	Carrot	62	108	682	PROPN_PROPN
('say', 'Detritus')	say	detritus	61	1757	170	VERB_PROPN
('I', 'see')	I	see	52	1616	288	PRON_VERB
('I', 'get')	I	get	47	1616	469	PRON_VERB
('oh', 'yes')	oh	yes	44	203	245	INTJ_INTJ
('Corporal', 'Nobbs')	Corporal	Nobbs	43	108	48	PROPN_PROPN
('know', 'I')	know	I	43	372	1616	VERB_PRON
('say', 'sergeant')	say	sergeant	42	1757	110	VERB_PROPN
('right', 'say')	right	say	41	306	1757	ADJ VERB
Bcero			52167	4440454	5265032	

BIGRAMS

POS1_POS2

Bce

POS1_POS2 (rus)

Bce

Text

text_2

POS1_POS2

Сумма Frequency
14533

text_5
3142

text_4
3018

text_2
2935

text_1
2789

text_3
2649

VERB_PROPN
882

PROPN_PROPN
769

PRON_VERB
564

VERB_PRON
171

PROPN_PRON
62

PROPN_VERB
55

VERB_VERB
55

INTJ_PRON
54

BIGRAMS

Text

text_1

▼

POS1_POS2

Bce

▼

POS1_POS2 (rus)

Bce

▼

Bigram	Frequency	Likelihood_Ratio	PMI	Student_t	Rank_Likelihood_Ratio	Rank_PMI	Rank_Student_t	Total_Rank
('Lady', 'Ramkin')	0,29%	1 612,00	15,63	12,68	1	252	1	254
('Grand', 'Master')	0,24%	1 460,95	26,94	12,50	4 511	4 789	4 626	13 925
('Sergeant', 'Colon')	0,17%	663,46	31,26	11,90	465	504	465	1 434
('Supreme', 'Grand')	0,23%	1 378,85	15,63	11,49	3	252	3	258
('Ankh', 'Morpork')	0,17%	1 010,61	15,63	10,94	4	252	4	260
('Brother', 'Watchtower')	0,18%	934,39	15,63	10,32	5	252	5	262
('Brother', 'Plasterer')	0,12%	582,59	15,63	9,62	7	252	7	266
('say', 'Carrot')	0,35%	580,75	15,63	9,50	8	252	8	268
('say', 'Vimes')	0,47%	571,03	15,63	9,07	9	252	9	270
('say', 'Nobby')	0,31%	527,47	15,63	8,70	10	252	10	272
('say', 'Brother')	0,24%	427,60	16,32	8,67	38 418	38 659	38 417	115 493
('say', 'Colon')	0,21%	290,95	31,26	8,32	479	504	479	1 462
('Lord', 'Vetinari')	0,06%	418,32	15,63	8,17	12	252	12	276
('Captain', 'Vimes')	0,14%	410,81	15,63	8,06	13	252	13	278
('I', 'think')	0,24%	370,55	15,63	7,87	14	252	14	280
('I', 'mean')	0,16%	345,83	15,63	7,18	15	252	15	282
('Brother', 'Fingers')	0,06%	290,28	15,63	6,86	16	252	16	284
('Corporal', 'Nobbs')	0,04%	280,30	15,63	6,82	17	252	17	286
('lance', 'constable')	0,03%	200,41	31,26	6,64	482	504	481	1 467
('palace', 'guard')	0,04%	193,18	31,26	6,62	444	504	442	1 390
('Brother', 'Doorkeeper')	0,05%	243,00	15,63	6,26	19	252	19	290
('say', 'sergeant')	0,11%	138,77	31,26	5,97	226	504	226	956
('swamp', 'dragon')	0,04%	165,61	28,46	5,85	2 361	2 529	2 360	7 250
('Thieves', 'Guild')	0,04%	221,74	15,63	5,78	20	252	20	292
('Sgt', 'Colon')	0,04%	158,77	20,30	5,61	27 044	27 267	27 047	81 358
('sir', 'say')	0,13%	178,91	16,44	5,54	38 253	38 475	38 247	114 975
('say', 'Lady')	0,13%	186,13	15,92	5,37	38 920	39 162	38 931	117 012
('Mended', 'Drum')	0,02%	193,08	15,63	5,29	21	252	21	294
('...', '...')	0,000%	111,00	00,00	5,10	700	000	700	0 070