

Динамика лексического состава русской художественной прозы

(на материале частотных словарей Корпуса
русских рассказов 1900-1930)

Т. Г. Скребцова (СПбГУ),

А. О. Гребенников (СПбГУ),

Т. Ю. Шерстинова (НИУ ВШЭ)

Грант РФФИ

«Русский язык на рубеже радикальных исторических перемен: исследование языка и стиля предреволюционной, революционной и постреволюционной художественной прозы методами математической и компьютерной лингвистики (на материале русского рассказа)»
(грант РФФИ № 17-29-09173 офи_м)

Корпус русских рассказов (1900-1930)

- Причины выбора жанра рассказа
- Несколько тысяч рассказов, написанных в России и Советском Союзе
- Несколько тысяч авторов, как столичных, так и провинциальных
- Корпус разделен на три периода:
 - а) 1900-1913
 - б) 1914-1922
 - в) 1923-1930
- Выборка из 310 рассказов как начальный полигон для исследований

Наиболее известные авторы

- 1900-1913: Л.Н.Андреев, М.П.Арцыбашев, А.Т.Аверченко, К.Д.Бальмонт, А.Белый, В.Я.Брюсов, И.А.Бунин, А.П.Чехов, З.Н.Гиппиус, М.Горький, А.С.Грин, В.Г.Короленко, Ф.Д.Крюков, А.И.Куприн, А.М.Ремизов, А.С.Серафимович, Ф.К.Сологуб, Н.А.Тэффи, Л.Н.Толстой, В.В.Вересаев
- 1914-1922: А.А.Блок, И.А.Бунин, К.К.Вагинов, В.В.Вересаев, Ф.В.Гладков, М.Горький, С.А.Есенин, Е.И.Замятин, М.М.Зощенко, В.А.Каверин, В.П.Катаев, М.А.Кузмин, Б.А.Пильняк, Л.Н.Сейфуллина, Н.С.Тихонов, К.А.Тренев, К.А.Федин, И.С.Шмелев
- 1923-1930: И.Э.Бабель, М.А.Булгаков, А.П.Гайдар, Л.И.Добычин, В.В.Иванов, В.М.Инбер, Ю.К.Олеша, А.М.Коллонтай, Б.А.Лавренев, Л.М.Леонов, О.Д.Форш, К.Г.Паустовский, А.П.Платонов, М.М.Пришвин, А.Н.Толстой, А.А.Фадеев, М.А.Шолохов

Исследования Корпуса

- Предмет: не только язык и стиль рассказов, но также их тематика и композиция
- Метод: Сочетание качественного и количественного (в том числе статистического) анализа
- Анализ динамики изменений (сравнение различных показателей по периодам)

Частотные словари по выборке

- 1-й период (1900-1913): 24 316 лексем,
376 513 словоформ
- 2-й период (1914-1922): 24 617 лексем,
303 588 словоформ
- 3-й период (1923-1930): 30 560 лексем,
383 430 словоформ
- Суммарно по всей выборке: 124 081 лексема,
1 077 970 словоформ

Сравнение частотного распределения лексики

- Знаменательные слова, расположенные в верхних зонах частотного распределения (с частотой выше 100)
- Для каждого периода их количество превысило 200, и составило около 800 для выборки в целом
- Сравнение рангов лексем (ввиду разницы в объемах полученных словарей)

«Инварианты» частотного распределения

слово	1900-1913	1914-1922	1923-1930
сказать	1	2	3
один	2	3	4
глаз	3	4	2
говорить	4	1	5
рука	5	5	1
мочь	6	6	6

2-й период (1914-1922) vs. 1-й период (1900-1913)

слово	1900-1913	1914-1922	слово	1900-1913	1914-1922
офицер	–	169	теперь	–	20
русский	–	182	наконец	–	186
дьякон	–	83	красивый	143	–
писать	–	181	веселый	152	–
бог	89	65	улыбаться	173	–
солдат	116	62	счастье	228	–
письмо	190	117	тихий	128	–
можно	80	–	ребенок	213	–
должный	–	76	дети	69	107
нельзя	–	147	работать	218	–
сегодня	–	172	рабочий	202	–
сейчас	–	71	студент	217	–

3-й период (1923-1930) vs. предыдущие: общие тенденции

- Вхождение в верхнюю зону конкретных существительных (сельская жизнь, части тела, технический прогресс):
- Возвращение некоторых слов, связанных с мирной жизнью и «пропавших» во 2-й период
- Изменение характера социальных связей

3-й период (1923-1930) vs. предыдущие: иллюстрации

Слово	1900-1913	1914-1922	1923-1930
машина	–	–	250
поезд	–	–	256
вперед	–	–	170
книга	–	–	208
можно	80	–	95
нельзя	–	147	–
должный	–	76	67
офицер	–	169	–
бог	89	65	211
рабочий	202	–	112
работать	218	–	109
знакомый	179	193	–
гость	272	165	–

Поступательный рост ранга

Слово	1900-1913	1914-1922	1923-1930
товарищ	198	105	38
красный	163	110	54
деревня	250	192	150
народ	224	198	164
мужик	252	144	61
баба	—	143	81
лошадь	176	139	93
бежать	169	134	85
дорога	209	168	134
ветер	216	149	113

Поступательное снижение ранга

Слово	1900-1913	1914-1922	1923-1930
любить	39	43	69
сердце	47	54	77
чувствовать	54	152	197
мысль	62	102	130
смеяться	131	183	189
жена	52	84	139
муж	146	166	236
дети	69	107	198
казаться	14	51	58
деньги	133	178	249

Стабильно высокие ранги

- Глаголы движения (*идти, ходить, выйти, пойти, уйти*); позы (*сидеть, стоять, лежать*), речи (*говорить, сказать, спросить, молчать*), чувственного восприятия (*видеть, смотреть*), а также глаголы *жить, взять, спать, хотеть, любить, казаться*
- Прилагательные *последний, большой, маленький, старый, новый, белый, черный + другой, каждый*
- Существительные *город – улица – дом, комната – стол, человек – люди, время – год – час*, названия частей тела (*голова, лицо, глаз, нога, рука*), времени суток (*день, ночь, утро, вечер*), а также лексемы *жизнь, место, сторона, земля, дело, отец*

Выводы

- Построенные частотные распределения могут служить хорошим индикатором динамики лексического состава художественной прозы отдельной эпохи;
- Они позволяют проследить влияние крупномасштабных политических изменений на словарный состав языка художественной литературы.