Чего не хватает в «оцифрованном мире» лексикографу и социолингвисту

В. И. Беликов

Ещё у меня есть где-то гора ссылок по корпусной лингвистике ‹…›
dimkaguarani, 29 декабря 2002, 15:18

Примечание к эпиграфу: В НКРЯ словосочетание *корпусная лингвистика* не фиксируется, при том что четверть текстов с упоминанием *лингвистики* датировано 2003 г. и позднее.

Доклад посвящен специфическому ответвлению корпусной лингвистики — сегментно-статистическому методу анализа лексики (ССМ), содержащему богатый потенциал для лексикографии и оказывающемуся незаменимым для социолингвиста. Возможности этого метода демонстрируются на фоне того, что для тех же целей предоставляет НКРЯ, ставший уже традиционным инструментом в ходе разнообразных лингвистических исследований.

Результаты, полученные с помощью ССМ, излагались многократно, в том числе и на последних сессиях конференции «Диалог» (наиболее детально — в докладе 2009 г. в соавторстве с М. В. Ахметовой), однако ограничения на объем публикаций не дают возможности показать «методическую кухню», а поверхностно изложенная аргументация выглядит несколько декларативно. Настоящая электронная публикация доклада на Диалоге'11 снимает такого рода проблемы. В нескольких случаях по ходу изложения возникает необходимость сослаться на результаты, лишь косвенно связанные с основным содержанием доклада, но также требующие аргументации; такого рода информация вынесена в Прибавления, помещенные после основного текста.

1. Как реагирует НКРЯ на поиски *…реагировать* ... 1
2. Сегментно-статистическая методика как инструмент исследования лексики ... 5
3. На что рассчитан Национальный корпус, а на что — нет ... 10
3.0. НКРЯ и собственно лингвистика ... 10
3.1. Визит «лексикографа» в НКРЯ ... 11
3.2. НКРЯ vs. сегментная статистика на службе социолингвиста ... 13
1. *Училка* ... 13
2. *Химица* ... 13
3. *Студак* ... 14
4. *Туса* и производные ... 16
4. Итоги: чего лексикографу и социолингвисту все еще не хватает? ... 21
Литература ... 22

Начну с жизнеутверждающей цитаты. «В русском языке есть глагол несовершенного вида *реагировать*. Его коррелятами совершенного вида могут быть несколько разных приставочных глаголов: *прореагировать, отреагировать, среагировать* (явление нередкое, особенно среди заимствований). Какой из этих приставочных коррелятов употребляется чаще? К каким контекстам тяготеет каждый из этих приставочных коррелятов (например, какой из них охотнее сочетается с наречием *быстро*)? Наконец, в какой последовательности они появляются в современном языке — одновременно или по очереди? Различается ли частота их употребления в разные периоды?[1]» [Плунгян 2005: 302]. На эти вопросы «лингвист может ответить с помощью Корпуса буквально за считанные минуты» [Там же].

1. Как реагирует НКРЯ на поиски *…реагировать*

Сейчас трудно судить, каков был ответ на эти вопросы в 2005 г. С тех пор Национальный корпус значительно вырос в объеме, картина, выявляемая в январе 2011 г., представлена в Табл. 1 (число документов[2] с соответствующими глаголами в корпусе в целом, а для 1990-х и 2000-х гг. — отдельно в художественных текстах и публицистике) и в верхней части Табл. 3

[1] Не входя в детали, отмечу, что единица лексикона характеризуется частотностью трех типов: общее число вхождений в языковые произведения, число текстов, где она встречается, и число лиц, ее использующих; последняя представляется мне наиболее показательной. Исследовать эти частоты приходится по выборкам, настороженность к которым у меня не пропадает со школьных времен, когда я впервые взял в руки частотный словарь Э. А. Штейнфельдт [1963], обильно уснащенный пионерской атрибутикой. Есть понятие «случайная выборка», но как делать подлинно случайные и одновременно репрезентативные текстовые выборки, я не знаю.

[2] Произведения, представленные в Корпусе частями, считаются за один документ; это общее решение принято в связи с не всегда ясными принципами деления в НКРЯ единого текста на документы (см. прим. 8), а также для сопоставимости с поиском в корпусе толстых журналов, где романы, печатающиеся в двух и более номерах, я привык считать одним документом. Дублеты, встречающиеся в основном корпусе НКРЯ достаточно редко (при рассматриваемом поиске их не было), не учитываются. Датировки типа «1943-1999» обычно усредняются.

(число документов, где глаголы сочетаются с наречием *быстро*, вкл. форму *быстрее*)[3].

Таблица 1

	реагировать	*прореагировать*	*отреагировать*	*среагировать*
всего	2003	122	744	184
1800—1899	22	1	0	0
1900—1924	68	1	0	0
1925—1949	120	1	2	0
1950—1989	412	31	77	20
1990—1999	249	22	152	46
2000—	1132	66	513	118
худ., 1990-99	97	4	65	26
худ., 2000—	115	12	96	28
публ., 1990—99	137	14	84	22
публ., 2000—	746	41	366	72

Комментарии начну с употребления производящего глагола. Из 22 «документов» XIX в. 14 относятся к естественным наукам. Документ здесь — понятие довольно условное: все тексты А. М. Бутлерова из его трехтомника 1953—1958 гг. сведены в один документ «Теоретические и экспериментальные работы по химии», датируемый «1851-1886» (168482 словоформ, 43 вхождения глагола *реагировать*), а, например, четыре статьи физиолога Н. Е. Введенского, публикуемые по сборнику избранного (1950), рассматриваются как разные документы (*реагировать* в каждом встретилось по одному разу). Несколько искусственное занижение рейтинга химии можно счесть социально оправданным: физиологические опыты Базарова долго продолжали будоражить умы прогрессивной общественности, которая рассматривала работы Н. Е. Введенского как практическое пособие для их повторения[4].

Но для лингвиста пользы в этом я не вижу[5]. Приведение «сходных» текстов к некоему заданному объему — задача полезная, но вряд ли реализуемая на больших массивах документов. В любом случае делаться это должно по ясному и жесткому алгоритму; в НКРЯ это заведомо не так.

Вне естественнонаучной литературы примеры на *реагировать* появляются лишь с 1890 г., при этом в большинстве текстов, для которых в НКРЯ сфера функционирования определена как художественная, сами примеры достаточно физиологичны: *зрачок не реагировал на свет* (Мамин-Сибиряк. Три конца), *органическая ткань* ‹…› *должна реагировать на всякое раздражение* (Чехов. Палата № 6), *молодой неокрепший организм так быстро реагирует* ‹…› (Мамин-Сибиряк. Черты из жизни Пепко), *ослабела способность кожных нервов реагировать на температурные колебания* (Вересаев. Записки врача). В более обобщенном значении глагол фигурирует лишь у М. Горького в очерке «Мужик» (1899: *Сурков делает свои дерзости* ‹…› *на них никто не реагирует*) и в текстах, отнесенных к публицистике (В. С. Соловьев, «Оправдание добра», В. П. Мещерский, «Мои воспоминания», Л. И. Шестов, «Шекспир и его критик Брандес»).

Что касается глаголов совершенного вида, то *прореагировать* появляется первым (в «документе» с датой 1851-1886), и в течение длительного времени его можно считать узкопрофессиональным. В художественной прозе он фиксируется довольно поздно (Юлиан Семенов, «Семнадцать мгновений весны», 1968), а до этого однажды фигурирует в дневниковой записи

---

[3] Во всех поисках шум типа *«...просто не реагировать», — быстро сказал про себя Лужин* или *Ситуация* ‹…› *меняется очень быстро, реагировать на эти изменения надо моментально*, естественно, не учитывался.

[4] В паспорте к тексту Бутлерова в графе «Размер аудитории» указано «малая», а у довольно специальных статей Введенского — «О дыхании лягушки (Rana temporaria)» (1881), «О телефонических явлениях в мышце при искусственном и естественном тетанусе» (1883) и т. п. — «очень большая». Чтобы содержание второй из названных статей было яснее, приведу цитату из нее: *Телефонические явления при естественных сокращениях мне удалось прежде всего наблюдать на лягушке*.

[5] Есть и, условно говоря, «вред». Если тупо посчитать число вхождений глагола *реагировать* в тексты XIX в. по химии (добавятся работы Н. Д. Зелинского) и физиологии (Н. Е. Введенский, А. Я. Данилевский, И. И. Мечников), окажется, что у химиков на 4 документа 50 вхождений, у физиологов на 9 документов 16 вхождений, в среднем 12,5 словоупотреблений на документ против 1,7. Различие явное, но характеризует ли оно подъязыки соответствующих дисциплин? Конечно, по паспортам документов можно выяснять их объемы, тогда результаты будут более объективны, но ручная обработка материала не всегда привлекает.

от 16.11.1944 (сам дневник опубликован в 2002 г.)[6] и 10 раз в химических текстах (после Бутлерова однажды в 1908 и 8 раз в 1965—1966 гг.)[7]. Глагол *отреагировать* появляется в НКРЯ также как узкоспециальный — в первой половине XX в. он встретился лишь в двух текстах по психиатрии[8], а с 1950 г. проникает в художественную литературу. Несколько позднее (впервые — в «Живет такой парень» Шукшина, 1960—1964) в корпусе фиксируется глагол *среагировать*, который не имеет узкопрофессиональной специализации.

Отвлекаясь от профессиональных текстов, делаем вывод, что в повседневный узус все три глагола вошли **одновременно**, а даты появления в беллетристике — 1950 — 1960—1964 («1962») — 1968 — аккуратно объясняются различиями в частотности: разрыву в 12 лет соответствует соотношение 77/31, а разрыву в 6 лет — 31/20. Новизну глаголов подтверждают лексикографы: ни одного из трех нет не только в довоенном словаре под ред. Д. Н. Ушакова, их нет ни в тт. 8, 11, 14 БАСа, ни в МАСе.

О частоте употребления приставочных глаголов в разные периоды можно говорить лишь за последние 60 лет; если принять за единицу число текстов с наиболее редким глаголом *прореагировать*, то окажется, что *отреагировать* в «доперестроечном языке» появлялся в два с половиной раза чаще, а *среагировать* — на треть реже. В последующие годы употребимость двух более новых глаголов резко возросла: *отреагировать* стал абсолютным лидером, а тексты со *среагировать* стали в два раза более частотными, чем с *прореагировать*. Но сводная картина несколько маскирует реальное положение вещей. По числу текстов с глаголом несовершенного вида довольно ясно видно резкое увеличение доли публицистики: за 1950—1989 она составляла 36%, а в 2000-х годах — 66%, в то же время доля художественных текстов сократилась с 29% до 10%. Доля документов, относимых к прочим сферам функционирования языка, снизилась с 35% до 24%. Пользователю НКРЯ следует знать, что публицистика здесь понимается максимально широко: тут и газетные статьи от спорта до политики, и мемуары, и, например, описание маршрута клиентов харьковского турагентства «Стелима», размещенное на сайте агентства[9], и специализированные профессиональные тексты типа *По уравнению реакции это количество кислоты может прореагировать с 0,099 грамм-молекулы каждого из карбонатов, а в первый стакан внесено 0,2 грамм-молекулы* $CaCO_3$, *во второй — 0,24 грамм -молекулы* $MgCO_3$. («Химия и жизнь»).

Нужна ли корпусу текстов сбалансированность в профессиональном узусе? Не думаю, что ее вообще можно добиться.

Нужно ли стремиться избегать заведомого дисбаланса? Вот тут ответ мне кажется очевидным. Равномерно распределенное по временной шкале обсуждения проблем геодезии, металлургии меди, навигации, карточных игр, бухгалтерского учета, орфографии, овцеводства, прок-

---

[6] Формально можно считать, что в дневнике В. Катаняна, датированном 1943—1999, он также уже встретился, но там запись, содержащая глагол *прореагировать* датирована 13 июля 1968 г.

[7] Не приходится сомневаться, что и сейчас для химиков видовой парой к знач. ‘вступать в химическую реакцию’ выступает глагол *прореагировать*; формальным подтверждением может служить сайт журнала «Химия и Жизнь — XXI век», где в химических контекстах используется только *прореагировать*, а в остальных — исключительно *отреагировать* (*организм отреагирует, эмоционально отреагировал* и т. п.). Осмысленный толковый словарь должен отмечать это обстоятельство.

[8] Т. И. Юдин, «Психопатические конституции» (1926) и В. А. Гиляровский, «Психиатрия. Гл. 30-40» (1935). Главы 1—19 и 20—29 книги Вас. А. Гиляровского представлены в НКРЯ отдельными документами. Книга эта в целом толста (298310 словоформ), но документы «…Гл. 20-29» и «…Гл. 30-40» в сумме (185334 сл.) несколько уступают по объему документу «П. И. Мельников-Печерский. На горах. Книга первая» (186832 сл.), а все три документа «Психиатрия…» в сумме мало отличаются от единого документа «Братья Карамазовы» (296343 сл.). Иногда от того, дробится ли текст на несколько документов, зависит содержательный результат. Вполне понятно, почему за 1950—59 гг. лингвистика встретилась в одном документе, а *языкознание* — в 19. За 1960—89 *лингвистика* обогнала *языкознание* — 35:29. А что случилось в 1990-е неясно. Лингвистика упомянута в 36 документах, практически то же, что в предыдущие 30 лет. А со словом языкознание — то ли 24 документа, то ли 32; во первом случае падение по сравнению с предыдущим периодом, во втором — рост. А разница возникает из-за того, что отдельные главы «Книги воспоминаний» И. М. Дьяконова считаются в НКРЯ самостоятельными документами, их в выдаче за 1995 г. девять. Одни главы побольше, другие поменьше, есть и на 6535 словоформ, что совсем немного.

[9] *И Таня куда-то исчезла, на последние письма никак не отреагировала и не звонит* (Отчёт о велосипедном походе, 2001).

тологии, легкой атлетики, уголовного права и т. п. представить трудно, отвести текстам каждой из десяти названных сфер по десятилетию было бы совсем странно, — этого, конечно не происходит, но и стремления к равномерному распределению специализированных текстов не заметно.

Упоминание *орфографии* в 1956 и 1964—65 гг., также как *гимнастов, пловцов* и даже *саблистов* в год Олимпиады-80, оказалось бы вполне уместным. Однако ни одно вхождение слова *орфография* за 1950-е — 1960-е годы не касается ни происшедшей в этот период, ни несостоявшейся реформы, а в текстах, датированных 1980 г., не упоминаются не только названные спортивные специальности, но и *Олимпийские игры*[10]. «Бойкот» выражению *Олимпийские игры* распространяется в НКРЯ на период 1978—1982, но я не готов считать все только что перечисленное признаками несбалансированности НКРЯ: орфография и даже спорт — тематики специализированные.

Однако если расширить «±» вокруг 1980 г. до пяти, дисбаланс налицо: в текстах 1983—1985 *Олимпийские игры* поминаются один раз, а за 1975—77 — тринадцать, причем 12 из них в «Технике — молодежи». Не слишком ли часто не самое спортивное издание писало об олимпиадах — вопрос неуместный, а обсуждение равномерности насыщения этим журналом Национального корпуса вполне законно: обработка одного номера в год за все годы выхода журнала дала бы куда более объективную информацию о русском узусе. Неравномерно разложена по корпусу и другая периодика, результат довольно заметно сказался как раз на той лексике, поведение которой заинтересовало В. А. Плунгяна.

В Табл. 2 ее распределение по времени представлено на фоне двух других специализированных слов. Глагола совершенного вида не так много и несложно отдельно выделить его химическое значение. За сорокалетний период 1950—1989 он встретился в НКРЯ 31 раз, 12 из них — в химическом контексте. Мы уже знаем, что *прореагировать* — глагол новый, ожидать его равномерного распределения по годам не следует, но вот почему у гораздо лучше представленного бесприставочного глагола **четверть** всех содержащих его документов за 40 лет попала на 1965—66 гг.? Потому что в 1965 г. начал выходить журнал «Химия и жизнь», призванный разъяснить народу объявленный накануне лозунг *…плюс химизация народного хозяйства*? Отчасти поэтому, но в основном потому, что в НКРЯ включены чуть ли не все тексты этого журнал за два первых года издания, а позднее он появляется в корпусе редко. В строке «в т. ч. 1965—66» из 249 документов, содержащих рассматриваемые лексические единицы, лишь 9 *реагировать* из иных источников, остальные 240 — из «Химии и жизни»[11].

Таблица 2

	реагировать	*прореагировать* (в т. ч. хим. знач.)	*кристаллический*	*полимер*	*среагировать*
всего	2003	122 (25)	413	282	184
1850—1899	22	1 (1)	31	3	0
1900—1924	68	1 (1)	20	0	0
1925—1949	120	0	39	2	0
1950—1989	412	**31 (12)**	145	134	20
в т. ч. 1963—64	11	0	1	2	0
в т. ч. 1965—66	100	8 (7)	57	83	1
в т. ч. 1967—68	14	1 (0)	1	0	0
1990—1999	249	**22 (1)**	19	12	46
2000—2004	1106	64 (10)	152	128	115
2005—…	26	2 (0)	7	3	3

По сравнению с 1990-ми число текстов за 2000—2004 гг. в корпусе значительно возросло. Но главное — изменилась их содержательная структура, именно поэтому встречаемость *среагировать* за эти два периода соотносится как 1:2,5, а у *реагировать* — как 1:4,4. И дело не

[10] Возможно, максимальным достижением той Олимпиады следует считать не побитый за 30 лет рекорд стрелка из пистолета на 50 м Александра Мелентьева. *Стрелок* в текстах 1980 г. упомянут трижды: у Л. И. Брежнева («Возрождение»), В. Орлова («Альтист Данилов») и В. Черных («Москва слезам не верит»).

[11] В том числе и нехимические *Организм ‹…› не успевает активно прореагировать* и *Мозг среагировал на какой-то раздражитель*.

только в росте доли текстов связанных с химией: *Олимпийские игры* за 1990-е годы упомянуты в 29 текстах, за следующие пять лет — в 193, *орфография* — в 31 и в 66, *морфология* в лингвистическом смысле — в 10 и в 16[12], *морфология* в других значениях — в 9 и в 87; нелингвистическая морфология сначала чуть уступала лингвистической, а потом стала поминаться в пять с лишним раз чаще.

2. Сегментно-статистическая методика как инструмент исследования лексики

Попробуем выйти за пределы НКРЯ, может быть, узнаем что-то новое.

Насколько верно, что глагол *реагировать* начал употребляться учеными-естественниками, а в общее употребление попал не без воздействия докторов и писателей-медиков, причем лишь в самом конце XIX в.? Обращение к сегменту «Классика» Библиотеки Мошкова заставляет в этом усомниться.

Она не знала, что в этом кротком ребенке ее женственная ласка возбудит подземный огонь эротической страсти; а раз эта страсть возбудилась во мне — она реагировала на нее, и чем я больше рос, тем больше ее материнская любовь переходила в иную любовь, а моя детская привязанность в козлоногую похоть сатира, откровенничает Н. П. Огарев о своей гувернантке («Моя исповедь», 1862). До первых внеестественнонаучных фиксаций в НКРЯ глагол не так уж редко встречался и в отечественной беллетристике, и в художественных переводах — *Здоровье их расстроилось, реагировать против гибельного действия климата они не могли* («Приключения капитана Гаттераса» Жюль Верна, изд. 1882 г.)[13], и в судебных речах — *Когда человек, утопая, борется со смертью, реагирует и напрягает силы, чтобы дышать, тогда гортань, горло и бронхи содержат всегда пену* (Речь В. Д. Спасовича по Делу Давида и Николая Чхотуа и др., 1878), и в литературной критике — *Лишенная возможности реагировать на внешний мир, она* [мысль] *уходит в себя и давит всею своею тяжестью на внутренний мир человека* (П. Н. Ткачев, «Больные люди. "Бесы", роман Федора Достоевского», 1873). Видовая параллель, которую и в первой половине XX в. мы пока знаем лишь по психиатрическим текстам, столь же «древняя», ср.: *Прежде и мои «обидчики» отреагировали бы на укол в печати, а теперь все прошло без отклика* (Лесков. «Еврей в России», 1884); *Если я замечу, что кто-нибудь из наших отваливает, то я перережу ему горло, будь он сын моей матери! — сказал свирепый кузнец. — Ха, ха! — отреагировал Риенцо своим странным смехом* (Э. Бульвер-Литтон, «Кола ди Риенцо, последний римский трибун», изд. 1875 г.). Об укорененности этих глаголов в общелитературной норме говорит их использование в книгах для девочек: *Эти слова встревожили Эстер. Она прижала к себе сестру и почти не реагировала на ее лепетание* ‹…› *Никто из подруг не отреагировал на ее слова, и после минутного возбуждения от полученных новостей девочки замолчали* (Э. Мид-Смит, «Девичий мирок». Пер. М. А. Лялиной, 1900).

Глагол *прореагировать* попадает в детскую литературу также уже в начале XX в.: *Шакал не прореагировал; ему уже минуло три года, но нельзя же сердиться на оскорбление, нанесенное особой с клювом в ярд длины и сильным, как дротик* (Киплинг, «Вторая книга джунглей». Пер. Е. М. Чистяковой-Вэр, 1916)[14]. А вот глагола *среагировать* в сегменте «Классика» нет вообще; это уже дает основание считать, что он утвердился в языке позже начала XX в.; когда именно

---

[12] *Синтаксис* и *семантика* (в разных значениях) в конце XX и начале XXI оказались в НКРЯ представлены поразному: результат поиска на *синтаксис* | *синтаксический* с 45 документов вырос до 98, а *семантика* | *семантический* с тех же 45 до 197; результат для *фонетика* | *фонетический* изоморфен синтаксическому: рост с 32 до 68. А вот *лексикография* вместе прилагательным явно быстро переходит в разряд историзмов: за те же периоды число документов в НКРЯ упало с 10 до 15. Вот тут достоверность материалов корпуса отрицать было бы нелепо.

[13] В переводном романе *реагировать против* можно принять за кальку, но именно таково было некогда стандартное словоупотребление при отрицательной реакции на что-либо, ср. в авторском тексте того же времени: *Нельзя сказать, чтобы г. Андреевский* ‹…› *никогда не реагировал против уныния и бесплодных жалоб* (К. Арсеньев, «Новейшие поэты», 1887). В НКРЯ такое управление представлено в пяти документах 1892—1927 гг., сегмент «Классика» добавляет еще 8 примеров, последний — в «Воспоминаниях» П. Н. Милюкова (1943).

[14] Любопытно, что на сайтах, содержащих Энциклопедический словарь Брокгауза и Евфрона (vehi.net/brokgauz и clova.ru/d1) глаголы *отреагировать* и *среагировать* не находятся. Дублетность при поиске по этим сайтам очень мала, выдача всегда неполна: на clova.ru/d1 поиск *прореагировать* от 8.02.2011 и 21.03.2011 дал по пять статей, но одинаковых только три, на vehi.net/brokgauz (где обычно выдача меньше) поиск 21.03.2011 дал восемь статей.

— сказать будет можно лишь после появления достаточно представительного массива разнотипных оцифрованных русских текстов «средней половины» XX в.; пока его нет.

Рассмотрим результаты поисков на совместимость глаголов совершенного вида с наречием *быст ро* (см. верхнюю часть Табл. 3). Согласно Табл. 1 глагол *отреагировать* за 1990-е — 2000-е гг. встретился в 665 документах, а *среагировать* — в 164, то есть в 4,1 раза реже.

Таблица 3

быстро +	*прореагировать*	*отреагировать*	*среагировать*
НКРЯ, релевантные при поиске «/1» (релевантные при «/10»)			
всего	1 (1)	20 (23)	6 (6)
1950—1989	0 (0)	2 (2)	0 (0)
1990—1999	0 (0)	8 (10)	2 (2)
с 2000	1 (1)	10 (11)	4 (4)
Военная литература, тексты в 1950—1989 гг., релевантные при поиске «/1»			
	0	20	17
Журнальный Зал, релевантные при поиске «/1»			
всего	0	25	14
1995—1999	0	3	2
с 2000	0	22	12

За тот же период наречие *быстро* в непосредственном соположении с этими глаголами встретилось в 18 и 6 текстах соответственно (в три раза реже), однако при учете всех примеров, где наречие связано с глаголом (21/6) — лишь в 3,5 раза реже; десятые доли «разов» при таких цифрах вряд ли подлежат учету, так что намек на бóльшую «привязанность» наречия к глаголу *среагировать* выглядит сугубо формальным. Уровень риска каких бы то ни было рассуждений на эту тему для предшествующего периода очевиден, поэтому о динамике использования наречия с этими глаголами судить нельзя.

Но есть другие массивы оцифрованных текстов, где сходные задачи решаются успешнее, хотя работа с ними заметно более трудоемка, в первую очередь из-за отсутствия необходимого инструментария.

Сегмент «Военная литература» Библиотеки Мошкова (militera.lib.ru) содержит большой массив мемуарной, художественной, исторической и иной литературы (возможен поиск по соответствующим подмассивам), в основном второй половины XX — начала XXI в., но выявление датировки требует обращения к каждому тексту. При некотором усердии несложно выбрать подмассив, характеризующийся определенным словоупотреблением (например, содержащий один из рассматриваемых глаголов и *быстро* на расстоянии «/1») и датировать его, однако обратный процесс — отобрать среди текстов определенного периода те, которые обладают заданным свойством, — практически не реализуем. Легко сопоставлять можно лишь общее число текстов, удовлетворяющих нужным условиям (Табл. 4).

Таблица 4

Сегмент Военная литература, поиск 5.02.2011	*прореагировать*	*отреагировать*	*среагировать*
Все тексты[1]	147	«1126»[1]	328
Страницы, обновлявшиеся в 2010 г.	74	540	166
<глагол> /1 *быстро*, все тексты	1	59	30
<глагол> /1 *быстро*, страницы, обновлявшиеся в 2010 г.	1	29	20
<глагол> /1 *быстро*, тексты, датированные 1950—1989	0	20	17
Поиск 21.03.2011 (в скобках — <глагол> /1 *быстро*)			
Все тексты	147	«1096»	328
Мемуары (militera.lib.ru/memo), 1151 док.	44 (1)	353 (17)	169 (13)
Проза войны (militera.lib.ru/prose), 820 док.	32 (0)	180 (9)	65 (4)
Биографии (militera.lib.ru/bio), 307 док.	12 (0)	110 (5)	22 (3)
Военная история (militera.lib.ru/h), 573 док.	16 (0)	211 (9)	28 (6)

[1] Поиск через полтора месяца (21.03.2011) не выявил новых документов (*прореагировать*: 147, *среагировать*: 328; ситуация для поиска, вообще говоря, очень редкая), а оценка числа текстов с глаголом *отреагировать* уменьшилась: «1096».

Как видим, в сегменте в целом число документов с глаголом *отреагировать* троекратно

превышает число документов со *среагировать*, однако имеется явное жанровое различие: если в мемуарах разница чуть больше, чем в два раза, то в текстах по военной истории она составляет 7,5. Но при наличии обстоятельства *быстро* различие во всех случаях уменьшается.

Хороший срез современного литературного языка дает сегмент «Журнальный зал» (ЖЗ); это в основном тексты XXI в., частично также 1990-х гг., но в толстых журналах изредка публикуются и заметно более ранние произведения[15]. Здесь основная проблема в том, что подавляющее большинство текстов в ЖЗ продублировано дважды, некоторые более. Несколько лет назад поисковый алгоритм Яндекса это учитывал и первоначально предлагал каждый текст по одному разу, а полную выдачу давал лишь при реализации опции «еще с сайта» (при этом на первом этапе были некоторые потери полноты, которые ликвидировались лишь при полной выдаче). Сейчас выдается сразу все, так что близкое приближение к точному числу текстов можно получить делением брутто-результата[16] на два[17].

Общий недостаток поисковых машин — недостоверность объявленных цифр найденного, если они превышают 1000 — может быть компенсирован в этом сегменте поиском по отдельным журналам (см. Табл. 5)[18]. Другой недостаток — некоторые колебания в результатах поиска, зависящие от разных обстоятельств. По моим наблюдениям, в ЖЗ эти колебания происходят исключительно за счет разного объема дублирования, число релевантных документов не различается[19].

20 журналов из Табл. 5, покрывают 74% всех текстов сегмента ЖЗ, содержащих как глагол *прореагировать*, так *среагировать*; можно получить точный результат, просчитав употребимость глаголов и в других журналах, но соотношение итоговых сумм вряд ли будет заметно отличаться от уже полученного: соотношение текстов с глаголами *прореагировать* и *среагировать* и по всему массиву, и по 20 журналам составляет 1:2,4, а интересующее нас соотношение *отреагировать* к *среагировать* по 20 журналам составило 5,7:1.

Таблица 5

	прореагировать	*отреагировать*	*среагировать*
ЖЗ в целом	251	«2930»	591
ЖЗ, только инфинитив	41	442	96
ЖЗ, только прош. вр. ж. р.	48	620	99
«Новый Мир»	24	378	80
«Октябрь»	8	38	20
«Знамя»	21	295	58
«Дружба Народов»	15	272	26
«Континент»	16	152	42
«Новая Юность»	2	34	12

[15] В литературоведческих и работах (и не только) нередки цитаты из классической литературы.

[16] Так я называю собственно выдачу по запросу, с дублетами и цитатами; переход к нетто-результату требует ручной обработки.

[17] Мало пригодный для лингвистически ориентированного поиска Гугл организует выдачу в два этапа, но на первом небольшое дублирование при поиске в ЖЗ все же присутствует, а потеря релевантного часто составляет более половины. Потери в сравнении с Яндексом возможны и при «полной» выдаче.

[18] Как и при любом другом поиске, этот недостаток можно преодолевать поиском отдельных словоформ, но экстраполяция такого результата на статистику употребимости лексемы в целом не особо надежна.

[19] Мне не раз приходилось «подновлять» статистику, полученную по этому сегменту Интернета, неучтенные релевантные документы за более ранние годы встретились лишь однажды. В январе 2011 г. дополняя материалы конца 2009 на орфографию слова *пер*(*д*)*нуть*, кроме двух текстов 2009 и девяти 2010, я получил два за 2007 г., но они из харьковского журнала «©оюз Писателей», которого при предыдущем поиске в ЖЗ просто не было. Кстати, в НКРЯ в литературных текстах 2000-х годов орфография *перднуть* встретилась дважды, *пернуть* — 3 раза, в ЖЗ за те же годы — 17 и 40. Разница в уровне достоверности довольно заметная. Для оценки более редких явлений НКРЯ вообще не пригоден. Так, словоформа *блядями* появляется в 12 документах (с 1848 по 2002), а для *блядьми* — только одно вхождение (И. Бабель, 1930); поэтический корпус дает еще 2 и 3 примера, но они за 1811—1827 гг. В ЖЗ *блядями* встретилось в 32 текстах, *блядьми* — в 16; это уже статистика. Тем более нет шансов найти в НКРЯ междометное *блядь* в авторском тексте литературного критика (а у А. Левинсона в №104 «НЛО» за 2010 г., к глубокому моему изумлению, встретилось). Мне этот пример попался при выявлении распространения новой орфографии *блять* для междометия; в НКРЯ есть лишь два примера (блог и форум), что не позволяет делать никаких выводов. В ЖЗ новая орфография составляет 9% (15 примеров против 156), в любительской литературе — ок. 20%, в блогах — 2/3 и более.

«Зарубежные записки»	2	16	6
«Арион»	0	10	2
«Иностранная литература»	3	131	3
«НЛО»	9	140	13
«Отечественные записки»	10	58	8
«Вопросы литературы»	7	35	9
«Вестник Европы»	4	48	2
«Неприкосновенный запас»	2	78	4
«Звезда»	20	200	32
«Нева»	14	149	52
«Урал»	18	251	39
«Сибирские огни»	6	73	16
«День и ночь»	2	75	8
«Слово\Word»	2	44	4
всего по 20 журналам	185 (73,7% от ЖЗ в целом)	2477	436 (73,8%% от ЖЗ в целом)

Примечание: все цифры «брутто»; в редких случаях и лишь при очень малой выдаче они соответствуют или близки фактическому числу примеров. Так, для *прореагировать* «2» в «День и ночь» и «3» в «Иностранной литературе» отражают реальность, «8» в «Октябре» — это 6, «14» в «Неве» — это 8; а «10» в «Отечественных записках», «18» в «Урале», «20» в «Звезде» надо делить ровно пополам.

При достаточно большой выдаче число релевантных текстов составляет около половины найденного[20], но пока нас интересует соотношение, это не существенно, можно уверенно говорить, что *среагировать* в этом сегменте встречается в 5-6 раз реже, чем *отреагировать*, а в сочетании с наречием *быстро* (см. нижние строки Табл. 3, где фигурирует нетто-результат) — всего лишь в два раза. То есть сам глагол *среагировать* используется существенно реже, но его относительная сочетаемость с наречием в два, а то и в три раза выше, чем у *отреагировать*.

Можно ли что-то сказать о динамике процесса? Лингвистически полезным является раздел «Самиздат» Библиотеки Мошкова (zhurnal.lib.ru), где произведения размещает каждый, обнаруживший у себя творческие наклонности. Любые наметившиеся в языке изменения выражены в этом сегменте Интернета сильнее, чем в профессиональной литературе. Объемы Самиздата в целом слишком велики для получения достоверных данных, но здесь можно ограничить выдачу страницами, обновлявшимися за определенный период (см. Табл. 6).

Таблица 6

Дата обновления	*отреагировать*	*среагировать*	*среагир.* к *отреагир.*, %	*быстро* /1 *отреагир.*	*быстро* /1 *среагир.*	*среагир.* к *отреагир.*, %
с 1.01.2010 по 28.02.2010	959	503	52,5	17	40	235
с 1.09.2010 по 30.09.2010	810	423	52,2	29	36	124
с 1.12.2010 по 15.12.2010	869	440	50,6	20	39	195

Выдача в Самиздате имеет довольно большое и непредсказуемое дублирование, что́ в действительности означает факт обновления страниц в определенный период, остается неизвестным. Серьезного доверия эти результаты не вызывают, однако как указание на некую тенденцию их вполне можно рассматривать: относительная частота *среагировать* здесь явно выше, чем у тех, кто пишет профессионально, а в сочетании с наречием *быстро* глагол *среагировать* используется даже чаще, чем *отреагировать*.

Более важным оказывается учет материалов блогосферы. По стилистике блоги занимают промежуточное положение между традиционными письменными жанрами и повседневным разговорным узусом. Именно поэтому в них наиболее отчетливо проявляются все языковые инновации. Не трудно показать, что в этом сегменте популярность глагола *среагировать* несколько выше, чем в ЖЗ: число блогов[21], в которых он был использован, в разных регионах за разные

[20] Так что на начало 2011 г. ЖЗ превышает по объему тексты соответствующего периода из НКРЯ лишь вдвое, впрочем, если учитывать лишь художественные тексты и нон-фикшен, то превосходство ЖЗ будет как минимум троекратным.

[21] По блогосфере везде использованы данные о числе блогов в варианте «группировать по авторам», без блогов с «очень похожими сообщениями». При группировке «по авторам» все записи из одного блога (как блоггера-владельца так и его посетителей) автоматически собираются вместе и участвуют в подсчете один раз (при желании их можно просмотреть раздельно). «Очень похожими» обычно оказываются перепосты и другие цитаты (иногда

продолжение сноски см. на след. стр.

временные отрезки составляет четверть или несколько более от числа тех блогов, где фигурировал глагол *отреагировать* (см. Табл. 7). Впрочем, нужно иметь в виду, что данные по ЖЗ отражают число произведений, а не авторов.

Таблица 7

Регион, период	*отреагировать*	*среагировать*	*среагировать* к *отреагировать*, %
Москва, 09.2006	818	258	31,5
Петербург, 09—12.2006	769	202	26,3
Поволжье, 2006	830	260	31,3
Сибирь, 2006	661	168	25,4
Украина, 03—12.2006	965	297	30,8
Белоруссия, 2006	467	133	28,5
Москва, 5-24.12.2010	956	255	26,7
Петербург, 09—12.2010	752	204	27,1
Поволжье, 09—12.2010	723	199	27,5
Сибирь, 09—12.2010	748	173	23,1
Украина, 12.2010	603	148	24,5
Белоруссия, 2010	978	311	31,8

Результаты поиска *быстро* /1 [глагол] представлены в Табл. 8, они довольно причудливы: изначально *быстро* /1 *среагировать* преобладало незначительно, в 2006—2007 гг. его доминирование существенно возросло, а затем процесс пошел вспять, да так, что к концу 2010 г. статистика уже в пользу глагола *отреагировать*.

Таблица 8

быстро /1 + глагол: **все блоги**	2001—05	2006	2007	2008	01-06. 2009	07-12. 2009	01-04. 2010	05-08. 2010	09-12. 2010
реагировать	744	1065	1892	3519	2147	2393	2354	2462	3430
прореагировать	10	14	23	21	13	18	16	10	16
отреагировать	224	331	488	739	489	559	609	568	874
среагировать	280	499	699	868	539	681	612	613	824
среагировать к *отреагировать*, %	125	151	143	117	110	122	100	108	94
Москва	2001—05	2006—2007		2008—2009			2010		
реагировать	193	692		981			711		
прореагировать	2	7		6			4		
отреагировать	62	189		298			259		
среагировать	80	334		365			250		
среагировать к *отреагировать*, %	129	177		122			97		

Примечание. 2. Цифры, превышающие 1000, соответствуют показанному Яндексом на последней (сотой) странице выдачи. Интерпретировать их буквально не следует, хотя при очень небольшом превышении их можно уточнить сопоставлением с результатами выдачи по опции «по дате обратно».

Разгадка проста: возрастная структура блогосферы подвержена существенным изменениям, в 2006—2007 гг. в доля подростков была максимальной, потом их число снизилось в связи с массовым уходом в социальные сети[22], во многих регионах абсолютное число молодых блоггеров так и не превзошло пика, имевшегося в 2007 г., последующий рост блогосферы идет за счет среднего возраста. Но в отдельных регионах блогосфера и до сих пор оказывается в основном подростковой, и там такого рода статистика выглядит очень рельефно. В Астрахани за все годы по 2010 включительно *быстро /1 отреагировать* встретилось в 7 блогах, а *быстро /1 среагировать* — в 12, в Оренбурге, соответственно, 2 и 8; цифры мизерные, но показательные.

Не подлежит сомнению, что любые языковые инновации сильнее выражены в младших возрастах. Имея стратифицированную по возрасту синхронную статистику словоупотреблений, мы

продолжение сноски:

повторяющиеся десятками, а то и сотнями), а также идентичные записи одного человека, ведущего блоги на разных платформах, например, на livejournal.com и liveinternet.ru (таких блоггеров на удивление много).

[22] Ср. динамику студенческой молодежи в блогосфере Москвы (Табл. 12), школьников и студентов в блогосфере Белоруссии (Табл. 17). При этом важно показать, что такого же оттока не было среди более старших возрастов; об этом говорится в Прибавлении 1 в конце текста.

получим сведения о динамике языковых процессов. К сожалению, распределение блоггеров по возрасту, декларированное в расширенном поиске Яндекса, с лета 2008 г. практически не работает: при группировке по авторам или при переходе ко второму эшелону выдачи (а в блогах это записи 11—20) настройки сбиваются. Однако возрастная структура выявляется довольно отчетливо по частоте упоминания школьных реалий (показательно, что блоггерная активность подростков везде и регулярно резко снижается в июле месяце — «макушка лета»).

В ЖЗ, отражающем более «старый» язык, *отреагировать* встречается в 5-6 раз чаще, чем *среагировать*, у блоггеров только в три-четыре; популярность второго глагола явно растет. А резкий рост сочетания *быстро среагировать* по-настоящему впечатляет: в младших возрастах оно явно обогнало синонимичное *быстро отреагировать* не только относительно, но и в абсолютных цифрах. Разница в употреблении двух глаголов может быть связана с полом автора (в блогосфере доминируют женщины, а среди авторов письменных текстов их меньшинство), или же с различиями в тематике. Данные за разные периоды показывают, что блоггеры, не зависимо от пола, применяют глагол *среагировать* **к себе** в 2—3 раза реже, чем *отреагировать*, а **к власти** — в 4—6 раз реже. Соответствующая статистика по всей блогосфере за последний квартал 2010 г. есть в Табл. 9, там же для сопоставления приведены данные, которые можно получить из НКРЯ и ЖЗ. Довольно очевидно, что 30 текстов ЖЗ позволяют лишь выдвигать какие-то гипотезы, а из втрое меньшего стилистически разнородного материала НКРЯ вряд ли что можно извлечь.

Таблица 9

	Блоги, 10-12.2010	НКРЯ, 2000—…	ЖЗ, 2000—2010, нетто
"я отреагировал"	218	2: 2004 (2[1])	11 (2001—2010)
"я среагировал"	100	1: 2001[2]	6: 2004—2010
"я отреагировала"	260	2: 2003[3], 2004[4]	2: 2005, 2007
"я среагировала"	97	0	0
власть /3 отреагир., муж.	253	4[5]: 2002 (1), 2003 (3)	9 (2001—2010)
власть /3 отреагир., жен.	43	1[5]: 2001	2 (2006—2007)
власть /3 среагир., муж.	34[6]	0	0
власть /3 среагир., жен.	8	0	0

Примечания. 1. Мемуары А. Тарасова «Миллионер» и журнал «Вестник авиации и космонавтики». 2. «Формула», журнал об автогонках. 3. Газ. «Богатей» (Саратов). 4. Форум. 5. Журнал «Проблемы Дальнего Востока» (2002), остальное — газеты Москвы и Петербурга. Один пример из авторского текста, подписанного *Кирилл Пальшин, Ирина Подлесова* («Известия», 2002), в таблице не учтен. 6. Как и всегда, я пользуюсь статистикой без учета «очень похожих сообщений»; здесь удобно это пояснить. Общее число блогов, релевантных запросу, составило 77, из них 32 содержали фразу *Почему власти не среагировали на информацию, которая гуляла по интернету?* 13 — *Местные власти среагировали на указанные наиболее острые проблемы*. В приведенной в таблице цифре 34 учтено использование каждой из этих фраз лишь в одном блоге. Среди блогов, маркированных как женские, число содержащих «очень похожие сообщения» и без таковых совпадает (в них первое из гуляющих сообщений встретилось однажды, второго нет).

Разница в частотности рассматриваемых глаголов в разных контекстах особенно заметна на фоне того, что контекстно не привязанное соотношение употребимости этих глаголов находится как раз посредине (см. Табл. 7). Поведение наречия остается не вполне ясным; не исключено, что молодежь считает, что на любые перемены она способна *среагировать* **на порядок** *быстрее*, чем *отреагирует* власть.

3. На что рассчитан Национальный корпус, а на что — нет

3.0. НКРЯ и собственно лингвистика

Подведем промежуточный итог.

В цитате, с которой я начал, относительно трех глаголов были сформулированы четыре вопроса, ответы на которые в применении к текстам НКРЯ, действительно, можно получить «буквально за считанные минуты». Легко узнать, что чаще всего в корпусе встречается глагол *отреагировать*, он же охотнее двух других сочетается с наречием *быстро*, он же первым — в 1950 г. — появляется в беллетристике (в текстах по химии его опережает известное с XIX в.

прореагировать). На любом относительно большом отрезке времени примеров на этот глагол больше, чем на конкурирующие. Единственное исключение — дореволюционный период, когда глагола *отреагировать* еще не было, а *прореагировать* дважды было употреблено химиками.

Но эти ответы — про НКРЯ[23]. Выше было показано, что на те же вопросы ответы про русский язык могут заметно отличаться, если воспользоваться статистикой, полученной по разным относительно однородным массивам текстовых документов, представленных в разных сегментах Интернета. Этот подход к работе с языковым материалом я называю сегментно-статистическим методом (ССМ).

НКРЯ в совокупности является хорошим «макетом» русского языка: разметка и стоящая за нею идеология — это модель устройства языка (не во всем эксплицитная, заведомо не идеальная), а комплект текстов корпуса (создающийся как выборка из массива речевых произведений, порожденных в рамках реального русского языка) — экспериментальная база для проверки модели. НКРЯ создан лингвистами и для лингвистов, а лингвиста интересует грамматика. Часть ее «зарыта» в лексиконе, это продуктивные и уникальные модели грамматического поведения лексических единиц. Слова, ведущие себя стандартно, лингвиста интересуют очень мало: важно знать объем некоторого класса однотипно ведущих себя слов (степень продуктивности модели) и иметь под рукой несколько представителей каждого класса, чтобы строить примеры типа *Seymour sliced the salami with a knife*. Прагматическая ценность такого рода примеров лингвиста не интересует[24]. Собственно лексикографу (составителю общего словаря[25]) интересно **каждое** слово. Еще один тип специалиста по языку — социолингвист, на него НКРЯ также не рассчитан, поскольку ни в коей мере не может считаться сбалансированным с точки зрения решаемых им задач[26].

НКРЯ хорош для решения любых собственно лингвистических задач, часто полезен лексикографу и социолингвисту — в основном лишь для предварительной оценки положения вещей, хотя в отдельных случаях может служить и основанием вполне серьезных выводов. Но исследователи с лексикографическими и социолингвистическими интересами иногда склонны принимать его за полигон принятия решений в последней инстанции, что досадно. Приведу два недавних показательных примера.

3.1. Визит «лексикографа» в НКРЯ

«Русский орфографический словарь» подвергся разнообразным нападкам со стороны Н. В. Николенковой. Основная претензия — наличие в словнике единиц, «обнаружить наличие которых не удается даже при обращении к Национальному корпусу русского языка» [Николенкова 2011: 181]. Примеров необнаруженного нет, приведены лишь три слова, встретившиеся в

---

[23] Невольно напрашивается и ответ на имеющий форму вопроса заголовок цитированной статьи *Зачем мы делаем Национальный корпус русского языка?* — Чтобы за считанные минуты узнавать свойства размещенных в нем текстов. Это ответ ёрнический. Правильным ответом, вероятно, будет: *Чтобы иметь под рукой прагматически достоверные примеры для верификации отдельных положений лингвистической теории*.

[24] Это, конечно, полемическое упрощение: собственно лингвисты всегда были и останутся разными, как и те, чье основное академическое занятие считается социолингвистикой или лексикографией. Свой взгляд на взаимоотношение лингвиста и объекта его науки я достаточно детально излагал на I Международном симпозиуме по полевой лингвистике в 2003 г.; на той же конференции обсуждалась и проблема взаимоотношения исследователя и корпуса текстов. См. Прибавление 2.

[25] Существует довольно много специализированных словарей, где единицы словника подбираются в соответствии с некоторыми лингвистическими принципами (классические примеры — «Толково-комбинаторный словарь современного русского языка» или «Новый объяснительный словарь синонимов русского языка»). Их создатели выступают как лингвисты. Ничто не мешает лингвисту иметь параллельно другую ипостась: быть собственно лексикографом, поэтом или петь профессионально. Степень взаимосвязи этих видов деятельности с собственно лингвистикой различна, но это уже дело десятое: главное — занятия разные.

[26] Декларация создателей корпуса говорит о некоем идеале: «Национальный корпус должен быть прежде всего большим ‹…› Но, кроме того, он — и это даже важнее — должен быть представительным. Иначе говоря, он должен содержать все типы текстов, представленные в данном языке в данный исторический период, и при этом содержать их в правильной пропорции» [Плунгян 2005: 297]. Если бы была возможность сменить модальность с «должен содержать» на «содержит», интересы социолингвиста (но не лексикографа) были бы соблюдены полностью. Но это вряд ли достижимо, особенно когда «правильная пропорция» иначе как субъективно не определяется.

НКРЯ в единственных контекстах, тем самым, также недостойные помещения в орфографический словарь: *остервенить*, *окровенить* и *форсун*. Вообще-то довольно самонадеянно критиковать словарь, тому, кто иных словарей в руках не держал. Дело в том, что все три слова присутствуют и в Орфографическом словаре 1956 г., и в МАСе, и в последнем толковом словаре Н. Ю. Шведовой [2007]; мало того, во всех перечисленных есть и более редкое слово *форсунья*.

Обсуждать принципы составления орфографического словаря в контексте моей статьи неуместно, но на тот случай, если мой текст попадет на глаза Н. В. Николенковой или приравненным к ней специалистам, сообщу, что написание некоторых слов колеблется, например, одни пишут *окровенить*, другие — *окровянить*.[27] Кодификаторы пытаются навести унификацию, но не всегда успешно (см. ниже).

Не знаю, каким образом для трех упомянутых слов было найдено лишь по одному вхождению, хотя не исключено, что на заре создания НКРЯ именно так и было. К 2011 г. на *остервенить* находится 12 документов (даты создания: 1806-1818 — 1971), *окровенить* — 4 (1875-1881 — 1990-2000), на *форсун* — 5 (1925—1948); псевдоуникальность *окровенить* компенсируется находками для *окровянить* (21 документ за 1862—2002 гг.). Временное распределение текстов с глаголами указывает на их постепенную архаизацию, но для достоверных выводов данных мало.

А вот статистика по *форсуну* удивляет, пять документов — немного (особенно принимая во внимание, что два документа — это «Книга первая» и «Книга третья» эпопеи В. Я. Шишкова «Емельян Пугачев»), но все они попадают на слабее всего документированный период. Если относиться к статистике по НКРЯ с полным доверием, это должно означать что лексическая единица родилась в урожайные для жаргона второе-третье десятилетие XX в., а потом устарела. В самом деле, временно́е распределение текстов, где в НКРЯ встретился произв одящий глагол *форсить*, таково:

1855—1924: 35 документов (54%),
1925—1948: 6 документов (9%),
1949—2004: 24 документа (37%).

Верификация сегментно-статистическим методом дает совсем иные результаты. В разных разделах Библиотеки Мошкова *форсун* без труда находится (но — увы — лишь вручную паспортизируется) в пяти более ранних документах (1), в десяти более поздних (3) и (без учета совпадений с документами НКРЯ) в четырех, относящихся к периоду 1925—1948 (2):

(1) А. Ф. Писемский, «Люди сороковых годов» (1869), С. Т. Семенов, «Девичья погибель» (1894), Е. Н. Водовозова, «На заре жизни» (1911), В. Бласко-Ибаньес, «Обнаженная» (пер. Т. Герценштейн,1911), А. Е. Зарин, «Живой мертвец» (1915);

(2) К. С. Петров-Водкин, «Хлыновск» (1930), И. Ф. Наживин, «Глаголют стяги» (1935), А. Г. Малышкин, «Люди из захолустья» (1938), К. А. Федин, «Необыкновенное лето (1948)

(3) В. В. Карпов, «Вечный бой» (1967), Н. П. Воронов, «Юность в Железнодольске» (1972), Синклер Льюис, «Бэббит» (пер. Р. Райт-Ковалевой, 1973), М. В. Водопьянов, «Небо начинается с земли» (1976), Ф. Абрамов, «Дом» (1978), П. И. Федоров, «Генерал Доватор» (1979), Е. Ф. Богданов, «Прощайте паруса» (1981), В. Л. Кондратьев, «Встречи на Сретенке» (1983), А. П. Злобин, «Бонжур, Антуан!» (1986), В. П. Астафьев, «Забубенная головушка» (1992)

В ЖЗ *форсун* встретился пять раз (один — публикация в журнале «Урал» очерка «Отцы» Мамина-Сибиряка к 150-летию со дня рождения писателя). Посегментная статистика письменных текстов (Табл. 10) показывает, что *форсун* и *форсунья* всегда были словами редкими, первое в литературных текстах еще держится, второе скорее утрачивается. В блогосфере спонтанно (то есть вне цитат) каждое из этих слов появилось лишь по десятку раз (отмечается и их непонимание собеседниками)[28]; это явно свидетельствует, что для повседневного городского узуса они устарели.

---

[27] Иногда и в авторитетных справочных изданиях: в словаре под ред. Д. Н. Ушакова в ст. *кровяни́ть* сказано: «*несов.* (*к* окровянить)», хотя самого глагола *окровянить* нет (как и *окровенить*).

[28] В блогах *форсун* используется также в знач. ‘форсунка’: *Новые 49мм дросселя Mikuni обеспечивают работу 8-ми форсунов*.

Таблица 10

	Классика, XIX век	Классика, 1900—1924	ЖЗ	Самиздат
форсить	45	20	59	312[1]
форсун	2	3	4[2]	10
форсунья	2	0	1	0

Примечания: 1. Брутто. 2. без Д. Н. Мамина-Сибиряка

В ходе поисков в классической литературе обнаружилось три вхождения другого отглагольного имени, в прочих сегментах не встретившегося: *Но форсила все это проделывает* [бьет новичков] *«не изо всех сил» и не со зла* (Куприн, «На переломе», 1900), *Что, в самом деле, за хвастун выискался! Франт пике, нос в табаке! Форсило!* (Чарская, «Счастливчик» 1912); *Форсила несчастная!* (Аверченко, «Рождественский день у Киндяковых», 1914; обращено к мужчине). В словарях этого слова нет, нужно ли оно? На мой взгляд, в толковом словаре большого объема должна фиксироваться вся лексика, использовавшаяся в неузкоспециальных русских текстах послекарамзинского периода. Последнее время в моду вошло указывать на лексическую бедность русского языка в сравнении с английским, сопоставляются при этом словари. Отечественные словари действительно бедны и в сравнении с некоторыми иностранными, и с русским лексиконом. Включение в словарь «новых» (хоть и устаревших) слов полезно иногда и лингвистам, тем, например, кого интересует фактическая реализация словообразовательных потенций.

3.2. НКРЯ vs. сегментная статистика на службе социолингвиста

Исследователем НКРЯ «с точки зрения отражения социальной дифференциации языка» из слов «отражающих молодежный сленг школьников, студентов, были отобраны следующие: *училка*, *химица*, *студак*, *туса*» [Киеня 2010: 263—264]. «Наибольшее количество выдач было получено на запрос слова „туса“ и всех от него производных слов (80 — устный корпус, 2527 — газетный, 4137 — основной)» [264], на втором месте по встречаемости оказалась *училка*, а на *студак* и *химица* во всех подкорпусах нашлось лишь по одному примеру.

Каждое из четырех слов в этом контексте заслуживает отдельного комментария.

1. *Училка*

Это слово не стоит так однозначно относить к «молодежному сленгу школьников, студентов»; выдача основного корпуса НКРЯ в данном случае невелика (66 документов, 1923—2008 гг.), поэтому несложно вручную найти такие явно не жаргонные контексты:

> У лифта на них уставились две женщины, похожие на провинциальных **училок**, с одинаковыми чемоданами в руках (Ольга Новикова. «Женский роман», 1993).
>
> За этой маленькой чернявой «**училкой**» вставал образ восторженной идеалистки, беззаветно преданной школе, отдающей всю свою жизнь делу воспитания детей. (Эльдар Рязанов. «Подведенные итоги», 2000; о персонаже Ахеджаковой в фильме «Иронии судьбы»).

В БТС единственная помета при *училке* — *пренебр.*; ею вряд ли исчерпывается специфика использования этого слова в подростковом и взрослом дискурсе, да пренебрежительность и не всегда присутствует, ср.:

> Ольга СЕВЕРСКАЯ: ‹…› *На радио «Маяк» появилась программа «Грамотей»* ‹…› *«Русском радио» есть программа «Говорите по-русски»* ‹…› *По утрам на НТВ выходит программа, которая называется «Это правильно!», там* ‹…› *ужасно смешные диалоги и не очень строгие* ***училки****, а аудитория — от школьного возраста до пенсионного* (Круглый стол «Речевые стили в средствах массовой информации» // «Отечественные записки» 2005, №2).

Что именно хотелось выяснить С. Н. Киене о жаргонном употреблении этого слова в НКРЯ, удалось ли это сделать, осталось неизвестным.

2. *Химица*

То, что автор из Гродно собрался искать это слово в НКРЯ, не удивляет. Представление о том, что территориальные различия свойственны лишь сельским диалектам, а прочим языковым вариантам присуще единообразие на всем русскоязычном пространстве широко распространено; оно имплицитно заложено и в идеологии НКРЯ. При этом за общераспространенный

принимается язык своего окружения. Даже понимание индивидом наличия территориальной составляющей в социальной парадигме русского языка нисколько не мешает удивляться каждый раз, когда «свое» слово другим оказывается неизвестным. Вот что писала минчанка после шести месяцев участия в форуме «Городские диалекты»:

Я лично за свою минскую школьную жизнь не слышала никаких "русичек" (это было б очень уж ласкательное название для них) — только русица, и так со всеми предметами: белорусица, физица, биологица, географица, историца, ну и естественно, математица.
Вот уж точно в голову не пришло бы, что их так только в Беларуси называют...))

В действительности «…ички» в Белоруссии встречаются, но статистически заметно преобладают «…ицы». Число блогов из Белоруссии за 2007—2008, где встретились некоторые обозначения преподавательниц представлено в Табл. 11 (из орфографически ошибочных написаний включены лишь наиболее распространенные)[29].

Таблица 11

	… чка	*…ца*	соотношение
математи…	71	166	1:2,3
химии…	56	152	1:2,7
физи…	28	123	1:4,4
истори…	18	102	1:5,7
руси…	16	86	1:5,4
русси…	10	31	1:3,1
биологи…	15	71	1:4,7
географи…	17[1]	46	1:2,7
белоруси…	4	45	1:11,3
белорусси…	2	14	1:7
беларуси…	3	20	1:6,7
информати…	2	29	1:14,5
трудови…	1	8	1:8

1. Вхождения фразы *Приснилась географичка, на коне, с саблей и глобусом* из бродячего «Дневника двоечника» (12 блогов) исключены

Отдельный вопрос, считать ли *химицу* и ее аналоги жаргонизмами. *Химичка* и однотипные слова, попавшие в БТС, помечены *разг.*, но это, конечно, не аргумент. Однако когда в органе Совета министров Республики Беларусь об академике НАБ пишут *Федор Лахвич выбрал для поступления химфак БГУ, хотя — и все это знали — единственной учительницей, с которой у него не очень складывались отношения, была химица* (газ. «Республика», 3.05.2005), сомнение в однозначной квалификации этого слова как школьного жаргонизма возникают.

В этом контексте появление *химицы* (и *физицы*) в НКРЯ выглядит случайным, поскольку обусловлено исключительно включением в корпус повести Владимир Козлова[30] «Гопники»; попади вместо нее его же роман «Школа», место *химицы* заняла бы отсутствующая сейчас *математица*.

3. *Студак*

Единственный пример на это слово, да еще лишь в газетном корпусе — тоже случайность, но иного рода. Корпус делается силами москвичей, младшему поколению которых этот жаргонизм знаком не хуже, чем жителям Белоруссии, так что можно было ожидать бо́льшего числа *студаков*.

Как выясняется, заниматься в НКРЯ молодежными жаргонизмами, даже свойственными Москве, дело безнадежное. Это надо делать в блогосфере. Один из региональных синонимов для ‘студенческого билета’ — *студень*, который в силу наличия омонимов требует тщательной

[29] Написания *белорусичка* и *беларусичка* я считаю для регионального варианта русского Беларуси равноправными, а (*бело*)*руссичка* — ошибочным. Впрочем, *русички* (тем более *белорусички*) в словарях нет. В РОСе нет и *трудовички*. В БТС нет даже *математички* и *биологички* (при наличии *физички* и *химички*). Можно упрекнуть в несистематичности составления словника, но кто ж такой упрек примет? Значит, сексизм.

[30] Владимир Козлов родился в 1972 г. р., вырос в Могилеве, там же поступил в вуз, потом жил в Минске, с 2000 г. живет в Москве.

ручной обработки материала. В ноябре 2008 г. мне приходилось заниматься проверкой мифа, согласно которому московскому *холодцу* соответствует петербургский *студень*[31]. Воспользуюсь представленной статистикой за последний из рассматривавшихся там периодов (октябрь 2008 г.) [Беликов 2011: 293], дополнив ее новыми словами и регионами[32]. Для общей ориентировки в Табл. 12 включена и численность блогов, где встретилось слово *студент* за рассматриваемый период, а также за первую и последнюю недели октября 2008 г.

Таблица 12

	Белоруссия	Украина	Москва	Петербург	Екатеринбург
студенческий билет	7	6	27	13	2
студенческий[1]	1	4	4	0	0
студбилет	0	0	3	1	0
студак	6[2]	7[3]	76	1	0
студень ‘студенческий билет’	0	1	4	16	0
студень ‘студент’	1	0	4	5	1[4]
студик ‘студенческий билет’	0	0	2	0	4
студик ‘студент’	2	5	5	1	0
студент, окт 2008	383	1339	3790	1030	140
студент, 1—7.10	122	381	1011	271	39
студент, 25—31.10	104	371	984	271	33

Примечания. **1**. Отбор примеров с субстантивированным прилагательным слишком трудоемок; цифры отражают релевантную выдачу на запрос *(мой | твой | свой | его | ее) /+1 студенческий*. **2**. Кроме того, одна запись на белорусском яз. **2**. Кроме того, две записи на украинском яз. **3**. Кроме того, в двух блогах упоминается газета «Студень».

Для меня, студента сорокалетней давности, нормой является *студенческий* или *студбилет*; «редкость» в блогах субстантивированного *студенческий* — частично методически обусловлена: соположение его с притяжательным местоимением встречается не так уж часто, а многократное преобладание *студака* над *студбилетом* в московских блогах поразило. Динамика соотношения этих номинаций в московской блогосфере представлена в Табл. 13

Таблица 13

Блоги Москвы	*студак*	*студбилет*	соотношение
2001—2003	79	15	5,3:1
2004	212	17	12,5:1
2005	538	31	17,4:1
в т. ч. 1.01—19.08	200	18	11,1:1
20.08—30.09	137	6	22,8:1
окт-дек	243	10	24,3:1
2006, с 1.01 по 19.08	838	37	22,6:1
2006, с 20.08 по 30.09	606	22	27,5:1
2006, окт—дек	569	16	36,6:1
2007, янв-апр	747	20	37,4:1
2007, с 1.05 по 19.08	579	16	36,2:1
2007, 20.08—31.08	330	7	**47,1:1**
2007, сен	779	18	**43,3:1**
2007, окт-дек	591	13	**45,5:1**
2008	950	43	22,1:1

[31] В старших возрастах, действительно так, но среди знакомых мне петербуржцев 40—50 лет есть такие, кто в детстве соответствующее блюдо звал *студнем*, а позднее перешел на *холодец*. У петербургской молодежи *холодец* в этом значении преобладает.

[32] Строго говоря, сопоставлять статистику, собранную в разное время, не следует: число ранних релевантных записей постоянно снижается: удаленные записи при очередной переиндексации поисковой машиной уже не находятся, так что при поиске через два с лишним года число найденных *студней* в Москве и Петербурге было бы немного меньшим, но характер территориального распределения синонимов таков, что общей картины это не изменило бы.

в т. ч. 1.01—19.08	591	27	21,9:1
20.08—30.09	213	9	**23,7:1**
окт-дек	201	10	20,1:1
2009	673	36	18,7:1
в т. ч. 1.01—19.08	402	22	18,3:1
20.08—30.09	141	7	**20,1:1**
окт-дек	160	8	20,0:1
2010	517	38	13,6:1
в т. ч. 1.01—19.08	298	25	11,9:1
20.08—30.09	126	6	**21,0:1**
окт-дек	129	8	16,1:1

Студенческие билеты упоминают в основном студенты, и упоминание имеет довольно выраженный сезонный характер с резким подъемом в конце августа пиком в сентябре. Сопоставление данных за последние месяцы 2007 и 2008 гг. позволяет предположить, что за этот период активность студенческой молодежи в блогосфере снизилась в 3-4 раза; это результат массового оттока блоггеров в социальные сети.

4. *Туса* и производные

Не понимая, какие именно слова рассматривались как «производные от ***туса***», могу лишь выразить искреннее недоумение большим числом найденных документов (4137 док. в основном корпусе). Для социолингвиста словообразовательный процесс в этом гнезде шел таким образом: (*тусовка* / *тусоваться*) → (*туса* / *тусить*); что при этом происходило в скобках, вопрос отчасти академический, мне больше нравится *тусовка* → *тусоваться* и *тусить* → *туса*, но настаивать не буду: без возможности аккуратного распределения блоггеров по возрасту анализ блогосферы дает противоречивые результаты (см. Прибавление 3).

Более ранняя пара к концу прошлого века вышла за пределы жаргона[33], но пара *туса* / *тусить* возобладала в молодежном дискурсе не сразу; разумеется детали процесса социолингвисту интересны.

В основном корпусе НКРЯ *туса* (в этом значении) встречается в 10 текстах (2002—2006, разные форумы, роман, журнал «Хулиган»), *тусить* — в четырех (2004, 2005, 2007 — форумы, 2008 — блог). Единственным «полигоном», на котором можно сопоставить *тусу* с *тусовкой*, выглядит «Хулиган». В 2003 г. на каждое слово приходится по одному документу, в 2004 г. есть один документ с двумя словами, два с *тусой*, десять с *тусовкой*. Сделать осмысленный вывод вряд ли возможно[34].

У *тусоваться* временнóе распределение в основном корпусе таково[35]: 1980-е гг.: 6, 1990-е: 32, 2000—2003: 78, 2004: 16 (из них 6 — форумы), 2005—2008 гг. — 8 (только форумы). Выходит, сопоставлять можно лишь материалы форумов, но их ничтожно мало, да и паспортизация не всегда логична[36]. Общее соотношение документов с форумов с глаголами *тусоваться* и *ту-*

---

[33] В деталях часто отличный от своего прототипа электронный БТС на Грамоте.ру сохраняет при словах *тусоваться*, *тусовка*, *тусовочный*, *тусовщик* помету «жарг.» бумажного издания [1998], но Н. Ю. Шведова [2007] *тусовку* маркирует «разг.». Из новой пары слов глагол словарями не отмечается, существительное же есть только в РОСе, причем его подача явно свидетельствует о нормативности поясняющего слова: «тусá, -ы (*тусовка, жарг.*)».

[34] В газетном корпусе НКРЯ *тусить* нашлось в 7 документах за 2003—2007 гг., и *туса* — в 6; для сопоставления со «старой» парой *тусоваться* / *тусовка* нужна погодовая статистика, но во-первых, этот раздел НКРЯ не пригоден для погодовой статистики, а для некоторых слов и для статистики вообще — (см. Прибавление 4), а во-вторых, из семи текстов распределенных по пяти годам статистики не выжмешь.

[35] «Длинные датировки» распределены так: к 1980—89 гг. отнесены документы: Евгений Попов. Во времена моей молодости (1970-2000), Евгений Попов. Кто-то был, приходил и ушел (1970-2000), Михаил Козаков. Актерская книга (1978-1995), Михаил Мишин. Торжественный комплект (1985-1991); к 1990—99 гг. — Александр Вяльцев. Путешествия в одну сторону (1985-2000) и Сергей Каледин. Записки гробокопателя (1987-1999).

[36] Я использую сортировку по дате. Источник документа 130 обозначен как «[Наши дети: Малыши до года (форум) (2004)]», а у документа 131 — «[Наши дети: Подростки (2004)]», во втором случае приписки «форум» нет, но это рядоположенный первому форум с сайта eva.ru/static/forums. При документе 139 — «[Фрунзенский снесли заранее (форум) (2007)]»; по названию видно что это не форум. В действительности это пост 18059 в блоге электронного СМИ «Фонтанка.ру», обсуждавшийся 19—24.04.2007 и получивший 151 комментарий.

сить — 14 к 3, соотношение числа вхождений в этих документах — 33 к 3[37]. То есть в тех текстах типа «форум», где можно встретить глагол *тусоваться*, он в 2004—2007 гг. в среднем встречался 2,4 раза на «текст», а для *тусить* аналогичное соотношение — 1:1. К узусу эти статистические изыскания совсем никакого отношения не имеют, но легко проверить, как эта статистика соотносится с «правдой жизни».

Яндекс предоставляет возможность искать по форумам. При расширенном поиске по блогам предлагается выбрать «область поиска», где есть режим «только в форумах». При таком поиске «группировать по авторам» означает «группировать по форумам», когда не только посты, но и топики с одного форума объединяются[38], то есть можно получить сведения и о том, на скольких форумах употреблялась та или иная единица, и сколько раз она появлялась на этих форумах. Разумеется, самостоятельным форумом в данном случае считается сайт, относимый Яндексом к числу форумов и проиндексированный на момент поиска[39]. По состоянию на конец марта 2011 г. выясняется (Табл. 14), что в начале периода, для которого в НКРЯ имеется статистика по форумам (2004 г.), в Интернете на три форума, посетители которого употребляли глагол *тусоваться*, приходился один форум, на котором встречался глагол *тусить*, к концу периода (последний квартал 2007 г.) соотношение числа соответствующих форумов стало 1,5:1.

Таблица 14

	В НКРЯ на начало 2011				Результаты поиска в Интернете, 31.03.2011			
	тусоваться		*тусить*		*тусоваться*		*тусить*	
	текстов	вхождений	текстов	вхождений	форумы (записи)		форумы (записи)	
2001	0	0	0	0	8	(10)	5	(5)
2002	0	0	0	0	22	(45)	1	(1)
2003	0	0	0	0	53	(89)	5	(5)
2004:	6	14	1	1	100	(207)	31	(40)
2005:	5	15	1	1	457	(3425)	205	(1129)
2006:	1	1	0	0	1053	…	588	(5461)
2007:	2	3	1	1	1863	…	1084	…
окт-дек 2007					881	…	571	…

Примечание. В результатах интернет-поиска четырехзначные цифры числа форумов / словоупотреблений приведены по последней показываемой Яндексом странице на 31.03.2011. Пятизначные цифры не приводятся.

Подтверждается «гипотеза», что материалы форумов включены в НКРЯ с собственно лингвистическими целями: для быстрого подбора стилистически различающихся примеров. Но при нужде в примерах «форумной стилистики» их проще искать непосредственно на форумах. Так, за январь-март 2011 г. на форумах встретилось 42 поста со *студаком* (и один по-украински) и два с *химицей*, за «форумные годы» НКРЯ в целом (2004—2008), соответственно, 1359[40] и 25; число форумов — 246 и 9. Разница с нулевой выдачей с форумов на те же слова в НКРЯ вполне ощутима. Но содержательных выводов об узусе форумы делать не позволяют[41].

---

[37] Если для «обычных» текстов подсчет вхождений vs документов для некоторых задач оправдан (например, для составления частотного словаря), то для текстов с форумов принимать во внимание вхождения абсурдно, см. Прибавление 5.

[38] «Одного форума» — в понимании Яндекса; поскольку этот тип поиска для статистических разысканий я не считаю полезным, а присвоение ярлыка «форум» единице коллекции НКРЯ — разумным, мое сопоставление двух статистик имеет сугубо академический характер.

[39] Кое-что со временем удаляется из Интернета, кое-что после архивации, возможно, перестает индексироваться — во всяком случае, с ненахождением Яндексом/Гуглом ранее доступных фрагментов форумов приходится сталкиваться постоянно.

[40] Это сумма за разные годы, так что цифра точно соответствует выдаваемому Яндексом, но реально разных текстов наверняка чуть меньше, поскольку в выдаче бывает небольшое дублирование.

[41] Что ни в коей мере не означает, что эти материалы бесполезны для социолингвистики. На форумах, например, хорошо представлена регионально маркированная лексика с разным стилистическим статусом. Вот «экзотический» пример. Пироги-расстегаи, обычно рыбные, наиболее популярны на Урале, при этом этимологизируются

продолжение сноски см. на след. стр.

Статистическую обработку жаргонной лексики стоит вести в том сегменте, где она широко используется. Лучше всего подходит подростковый срез блогосферы.

Мне повезло заниматься глагольной парой *тусоваться*/*тусить* в мае 2008, когда Яндекс еще позволял получать статистику с учетом возраста (он указан в профилях примерно половины блоггеров). В Табл. 14 приведены результаты поиска по Москве (данные не сгруппированы по авторам, это количество з а п и с е й, а не блоггеров, как в других таблицах; работа с записями позволяет получать цифры больше 1000 суммированием данных за более короткие временный промежутки). Как и ожидалось, об экспансии нового варианта среди лиц старше 30 говорить не приходится, а младшая возрастная когорта в жаргонном дискурсе перешла от *тусоваться* к *тусить*. Не все молодежные записи в блогах делаются на жаргоне, поэтому ставший нормативным глагол *тусоваться* также остался в употреблении, но к 2008 г. занимал довольно скромное место.

Таблица 14

п е р и о д	*тусить*	*тусоваться*	с у м м а	*тусить*, %
2004:				
11—19 лет	7	8	15	46,6
20—30 лет	603	887	1490	40,5
старше 30	75	348	423	17,7
2005:				
11—19 лет	79	55	134	59,0
20—30 лет	1432	1592	3024	47,4
старше 30	108	520	628	31,0
2006, I кв.:				
11—19 лет	168	96	264	63,6
20—30 лет	646	612	1258	51,4
старше 30	52	169	221	23,5
2007, I кв.:				
11—19 лет	1227	533	1760	69,7
20—30 лет	1053	838	1891	55,7
старше 30	67	186	253	26,5
2008, I кв.:				
11—19 лет	1301	484	1785	72,9
20—30 лет	626	531	1157	54,1
старше 30	79	221	300	26,3

Динамика использования шести слов этого гнезда в блогосфере Белоруссии представлена в Табл. 15. На протяжении 2008 г. и в середине 2010 видно падение абсолютных цифр, означающее снижение числа блоггеров, пользующихся соответствующей лексикой.

Показательно, что на паре *тусовка/тусоваться* в первом случае оно отразилось менее значительно, а во втором не имело места (см. Табл. 16). Любопытно, что статистика по глаголу *тусоваться* несколько ближе к явно жаргонным *туса*, *тусня*, *тусить*. Это не случайно. Дело в том, что производство формы 1 л. ед. ч. наст. вр. от глагола *тусить* затруднено, поэтому вместо нее используется соответствующая форма глагола *тусоваться*[42].

п р о д о л ж е н и е с н о с к и :

там они не через *расстёгивать* (чего с этими пирогами не делается), а через *растягивать* (тесто действительно растягивается), соответственно, и орфографически часто пишутся как *растягай*; у пермских блоггеров такая орфография встречается чаще, чем у екатеринбургских. При поиске на форумах сайта «Екатеринбург Он-лайн» (*e1.ru*) к апрелю 2011 г. нашлось 19 постов с *растягаем* и 31 с *расстегаем*; дублеты и шум (*не расстегай порток*, фамилия *Расстегаев* и т. п.) исключены. Статистика вполне внушительная, но сравнить ее не с чем: в Перми аналогичного форумного сайта нет, с кулинарными форумами сопоставлять бессмысленно, а на других форумах это изделие упоминается слишком редко для статистики. Однако даже единичное упоминание на forum.revda-info.ru (*Так же и растягаи из слоёного теста делают...мм.. И разные пирожки.. Всё пошла на кухню!*) оказывается значимым, поскольку блогосфера в Ревде невелика (глаголом *испечь* пользовались 10 блоггеров в 18 контекстах) и *растягаи / расстегаи* не упоминались.

[42] Этому обстоятельству можно получить статистическое подтверждение, см. Прибавление 6.

Таблица 15

период	*тусовка*	*туса*	*тусня*	*тусняк*[1]	*тусить*	*тусоваться*	*туса* от *тусовка*, %	*тусить* от *тусоваться*, %
2001—2004	149	41	15	8	25	135	27,5	18,5
2005	229	80	19	8	79	215	34,9	36,7
2006	676	240	64	24	348	591	35,5	58,9
2007, янв-июн	553	198	49	18	384	491	35,8	78,2
2007, июл-дек	667	217	48	29	443	558	32,5	79,4
2008, янв-июн	564	160	38	19	334	459	28,4	72,8
2008, июл-дек	461	128	29	13	260	365	27,8	71,2
2009, янв-июн	521	143	35	21	266	«1090»	27,4	ок. 25
2009, июл-дек	611	269	44	24	428	781	44,0	54,8
2010, янв-июн	685	403	46	25	697	648	58,8	107,6
2010, июл-дек	740	346	44	31	662	660	46,8	100,3

[1]Для Белоруссии слово *тусняк* пока достаточно редкое, слишком малая статистика делает диахронические наблюдения рискованными. Центром распространения этой номинации, по-видимому, была Москва. По распространенности она везде, кроме Казахстана, уступает *тусне*.

Таблица 16

Период	*тусовка*	*туса*	*тусня*	*тусить*	*тусоваться*
2007, июл-дек	667	217	48	443	558
2008, июл-дек	461	128	29	260	365
2008 к 2007, %	69,1	59,0	60,4	58,7	65,4
2010, янв-июн	685	403	46	697	648
2010, июл-дек	740	346	44	662	660
2010-2 к 2010-1,%	108,0	85,9	95,7	95,0	101,9

Уменьшение числа блоггеров связано с уходом их в социальные сети. Первое снижение связано с ростом популярности «Одноклассников.ru», второе — с развитием сети «В Контакте». В разных регионах это происходило несколько по-разному, но две волны снижения числа активных блоггеров отмечается повсеместно, первая обычно начинается на рубеже 2007/2008 гг., затем после умеренного роста в течение большей части 2009 г. опять наступает спад.

При использовании ССМ в блогосфере важно понимать, каков прототипический автор анализируемых текстов. Яндекс позволяет стратифицировать записи в блогах по полу и возрасту авторов (второе разделение не работает, но, можно надеться, временно). Иногда удается найти своеобразные лексические индикаторы, позволяющие точнее разобраться в социальной структуре блогосферы в конкретном месте и времени (см. Прибавление 1). Собственно лингвистической ценности эта информация не имеет, но ее учет необходим, чтобы ясно понимать, язык какой социальной категории отражает словоупотребление, характерное для конкретного сегмента.

Для Белоруссии можно проследить, как меняется число тех, кто использует номинации учителей на *-ица* (а это в значительной степени школьники) и жаргонизм *студак* (почти исключительно студенты). Из Табл. 17 видно, что школьники уже в 2007 г. начали покидать блогосферу: имея в виду предшествующий быстрое увеличение числа блогов формальный рост на два блога за полугодие явно означает падение; и в самом деле, если посмотреть статистику за более дробные периоды, окажется, что за два предканикулярных месяца 2007 г. «ицы» фигурировали в 116 блогах, в сентябре—октябре — в 177 (явный рост), а в ноябре—декабре — только в 129. В течение 2008 и первой половины 2009 продолжается падение, минимальный показатель за двухмесячный период приходится на март—апрель 2009 (29 «иц», в 2006—2008 гг. ниже бывали только сезонные летние цифры за июль—август), затем начинается медленный рост с очередным пиком в январе—феврале 2010 (181), за которым опять следует падение.

Таблица 17

период	*…ица*[1]	*студак*
2001—2004	35	6
2005	85	27
2006, янв-июн	150	19
2006, июл-дек	213	46
2007, янв-июн	294	44

2007, июл-дек	296	56
2008, янв-июн	222	21
2008, июл-дек	93	25
2009, янв-июн	101	6
2009, июл-дек	197	16
2010, янв-июн	389	14
2010, июл-дек	309	17

[1] Поиск на дизъюнкцию 13 слов из Табл. 12.

Как говорилось выше, *студак* — слово сезонное, чаще всего упоминается в конце августа — сентябре. Тем не менее статистика появления его в блогах указывает на двукратный отток студентов из блогосферы с конца 2007 г., хотя столь точная временная привязка, как со школьниками, оказывается невозможной.

В Табл. 15 обращает на себя внимание сравнительная редкость слова *тусня*. Однако эта редкость становится явной лишь при синхронном сопоставлении с другими географическими сегментами блогосферы (см. Табл. 18).

Если даже социальные структуры блогосфер Москвы и Петербурга абсолютно изоморфны, имеющуюся разницу в словоупотреблении двух городов нельзя считать значимой. А сопоставление данных по Петербургу и Белоруссии явно отражает различие в жаргонном узусе. Столь же заметны различия между узусом Минска о остальной Белоруссии, особенно имея в виду, что у некоторой части блоггеров город в профиле не указан, так что в графу «Белоруссия без Минска» фактически попадает и некоторое количество блоггеров-минчан.

То, что региональные различия в молодежном жаргоне, как и в других социальных вариантах языка, не сводятся к территориально специфичной лексике, очевидно, но делать какие бы то ни было утверждения о стилистических предпочтениях без надежной статистики невозможно.

Таблица 18

2010		*туса*	*тусня*	соотношение
Петербург,	янв-июн	385	112	3,1:1
	июл-дек	361	126	2,9:1
Москва,	янв-фев	637	182	3,5:1
	июн-авг	731	222	3,3:1
	ноя-дек	771	203	3,8:1
Сибирь,	янв-июн	844	108	7,8:1
	июл-дек	894	121	7,4:1
Белоруссия,	янв-июн	403	46	8,8:1
	июл-дек	346	44	7,9:1
Минск,	янв-июн	150	24	6,3:1
	июл-дек	117	27	4,3:1
Бел. без Минска,	янв-июн	253	22	11,5:1
	июл-дек	229	17	13,5:1

Поводом для рассмотрения специфики молодежного жаргона Белоруссии послужила статья белорусского коллеги, занимавшегося изысканиями в Национальном корпусе. Автор приходит к заключению, «что национальные корпусы отражают все многообразие существования языка, генеральную совокупность языкового материала, что свидетельствует об их высокой социокультурной значимости» [Киеня 2010: 265]. Как обстоит дело с национальными корпусами «вообще», судить не берусь: не знаю[43]. Но НКРЯ социолингвисту ничем помочь не может.

[43] Однако ж в их универсальной перспективности сомневаюсь. Для «малых» языков (слабо представленных в Интернете) они безусловно полезны, да и идея их создания имеет большую историю: Corpus Inscriptionum Etruscarum существует с 1885 г. А что касается «больших» языков, то для специализированных целей нужны специализированные корпуса, причем нужны не только лингвистам, пример тому — большие базы СМИ, которые создаются с целью извлечения из них отнюдь не лингвистической информации. При решения многих частных задач в области собственно лингвистики и преподавания языка (и родного, и иностранного) универсальные корпуса крайне полезны, но при их наполнении надлежит стремиться не к увеличению объема, а к улучшению квалифицированно размеченной части. Недавно искал в НКРЯ *кулису*. На *кулиса* — 782 документа, из них 33 со снятой омонимией. На *кулисы* — 720 документов, из них 4 со снятой омонимией (как и положено, на первый запрос эти 4 не выдаются).

продолжение сноски см. на след. стр.

4. Итоги: чего лексикографу и социолингвисту все еще не хватает?

«Для современной лингвистики ‹…› все большее значение приобретает понятие корпуса текстов. ‹…› [И]сследователь языка все реже задает вопросы типа „Можно ли так сказать на языке L?“ и всё чаще — вопросы типа „Было ли когда-нибудь так сказано на языке L?“ А это означает, что исследователь обращается к корпусу текстов на данном языке» [Плунгян 2003: 79].

Следует иметь в виду что эта цитата — с конференции по полевой лингвистике, для которой протитипический «язык L» — бесписьменный и в социальном отношении обычно достаточно однородный; если же имеет место заметное территориальное варьирование или существует письменная фиксация сходной коммуникативной системы, то «язык L» для исследователя — это здесь и сейчас. У информанта с умеренной языковой языковой рефлексией и умеренным почтением к исследователю на вопрос «Можно ли так сказать…» есть только два ответа. Изредка встречается ответ типа «Можно, но у нас так не говорят». Излишне почтительного информанта, дающего такой ответ, стоит отбраковать. Но тот же ответ может быть обусловлен обостренным чувством нормы, которая свойственна и таким коммуникативным сиситемам, которые никогда не подвергались кодификации.

Довольно очевидно, что идея появления текстового корпуса, при обращении к которому можно получить абсолютно достоверный ответ на вопрос «Было ли когда-нибудь так сказано», утопична. Но для полевого лингвиста (для которого речевые произведеия на языке L — это «здесь и сейчас») утопичность тут чисто теоретическая: хотя идеал не достижим, стремление максимально к нему приблизиться вполне характерно для отечественной полевой лингвистики. Прототипический полевой лингвист параллельно с созданием чернового грамматического описания языка занимается и сбором корпуса текстов, который как раз и служит в дальнейшем для верификации грамматического описания. Чем полнее такой корпус, тем объективнее в конечном счете будет описан язык. Но слепая вера всему, что попало в корпус, лингвисту мешает. Чем полнее корпус, тем выше вероятность появления в нем как раз таких фрагментов, которые «правильный» носитель языка оценит «так сказать нельзя». Таким образом, вопрос «Было ли так сказано» носит вспомогательный характер по отношению к вопросу «Можно ли так сказать».

Если речь идет о сложно устроенных с социолингвистической точки зрения языках, имеющих хорошо документированную историю, как бы ни был велик корпус, ответы на вопросы «можно ли» и «было ли» относительно легко получить лишь про ядро языка, с периферией сложнее. Сложности многократно возрастают, когда вопросы касаются не грамматики, а словаря, особенно если речь идет о социально маркированном узусе любого типа — от детской лексики и жаргона до от профессиональной терминологии (см. Прибавление 8).

Все предшествующее изложение подсказывает: тому исследователю, который задается вопросом «Было ли когда-нибудь так сказано на русском языке», не вредно заглянуть в НКРЯ. Если «буквально за считанные минуты» достаточно ясного ответа не будет, стоит обратиться к сегментной статистике.

И по любомк нетривиальному вопросу СС-методика дает возможность получить более нетривиальные результаты и с гораздо большей надежностью, чем это позволяет НКРЯ. В национальном корпусе язык «отражается ‹…› таким, каким он на самом деле *является*» [Плунгян 2005: 304]. Это справедливо; но сегменты Интернета, фиксирующие язык прошлого, богаче, поэтому «отражение» языка получается заметно более четким. В том, что касается современных текстов, Интернет уже не отражение, а депозитарий, с каждым годом все более полный. Значительная часть текстов только здесь и присутствует.

Типология текстов в НКРЯ довольно богата, но заведомо неполна. Неполон и лексикон. Например, коллокация *корпусная лингвистика*, присутствующее в блогосфере с 2002 г. и уже дважды встретившееся в Журнальном зале[44], рано или поздно появится в НКРЯ. Но ведь пока его

продолжение сноски:

Радует тут только то, что лишь 11% снимающих омонимию не знают, что *кулисы* имеют ед. ч.; в других поисках бывало гораздо хуже, но деталей не помню.

[44] Узнать, как чаще ставится ударение в *ко́рпусно́й лингвистике*, по текстам не удастся, но для прилагательного в

продолжение сноски см. на след. стр.

там нет! СС-метод позволяет работать и со многими другими словоупотреблениями, (пока) не представленными в НКРЯ.

Но **главное** — на сегодняшний день я не вижу способов социолингвистического исследования русской лексики, отличных от ССМ.

Метод этот пока довольно убог. «Оцифрованный мир» дал социолингвисту (и лексикографу) необъятный языковой материал, который «самоорганизовался» в относительно однородные в стилистическом отношении сегменты. Сегменты эти динамично растут; то что объем ЖЗ увеличивается за счет вновь возникающих текстов, ясно, но, как мы видели, фонд этого сегмента пополняется и за прошедшие годы (см. Прим. 19). Менее очевидна регулярность приращений в сегменте «Классика», но там за два последних года объем информации вырос примерно на половину[45].

Однако для работы с этими сегментами специального инструмента нет, приходится пользоваться тем, что имеется. Грамматика Яндекса вполне удобна, поисковый алгоритм достаточно хорош, но, будучи рассчитан на стандартного «искателя информации», а не языковых единиц, он в отдельных случаях намеренно пренебрегает полнотой, дабы избежать большого шума, который стандартному пользователю Интернета будет заведомо неудобен[46].

Будучи оптимистом, надеюсь, что рано или поздно полнота описания языка, объявленного в России государственным, всерьез заинтересует кого-то из тех, кто в силах принять ответственное решение, ведущее к созданию специализированного инструмента для исследования больших массивов оцифрованных текстов.

А пока буду вручную доказывать необходимость такого инструмента. Gutta cavat lapidem.

Литература

БАС: Словарь современного русского литературного языка. — М.; Л.: Изд-во АН СССР, 1948—1965. — Т. 1-17. (Т. 8 — 1959, т. 11 — 1961, т. 14 — 1963).

Беликов В. И. Интернет и орфография // Компьютерная лингвистика и интеллектуальные технологии. Труды Международной конференции Диалог'2003. М.: Наука, 2003.

Беликов В. И. Полевая и неполевая идеология в лингвистике // Полевая лингвистика. Сборник материалов Первого международного симпозиума. М.: ИЯз РАН, 2007.

Беликов В. И. Стереотипы в понимании литературной нормы // Die Welt der Slaven. Sammelbände. Band 38. Русский язык в условиях культурной и языковой полифонии. München — Berlin. 2009. (Также опубликовано в: Стереотипы в языке, коммуникации и культуре. М.: РГГУ, 2009.)

Беликов В. И. Статистические различия в региональном лексическом узусе // Вопросы культуры речи. М.: АСТ-Пресс Книга, 2011.

Беликов В. И., Ахметова М. В. Статистическая оценка функциональных свойств лексики по материалам Интернета. // Компьютерная лингвистика и интеллектуальные технологии. Вып. 8 (15). По материалам Международной конференции Диалог 2009. М.: ИПИ РАН, 2009 (http://www.dialog-

---

продолжение сноски:

мужском роде для современного языка (с 1990 г.) НКРЯ дает 11 ясных контекстов для трех значений, при этом в 4 случаях форма противоречит тому, что предлагают БТС и новый БАС. Из 11 ясных случаев самое частое значение — военное, четырежды реализуется в варианте *корпусной*, но встретился и *корпусный генерал* (4 к 1 — преобладание, но 5 — не статистика). Обращение к корпусу со снятой омонимией все окончательно запутывает. ССМ, как обычно, помогает внести ясность (см. Прибавление 7).

[45] Например, в январе 2009 г. я уточнял детали изменения формы фразеологизма *на своих на двоих*, утрачивающего предлог *на* в середине; в сегменте «Классика» было 15 примеров с предлогом и 4 без него, в январе 2011 — 23 и 6. Характерно, в НКРЯ для XIX в. встречается только старый вариант, для XXI в. — только новый. Посегментная статистика дает большую гибкость: «По-новому» писали уже В. М. Гаршин («Петербургские письма», 1882) и В. В. Крестовский («Очерки кавалерийской жизни», 1892). В ЖЗ из 121 релевантного текста за 2000—2010 *на своих на двоих* присутствует в пяти. (В НКРЯ последние 23 примера относятся к 2000—2008 гг., в ЖЗ за тот же период 90 релевантных текстов, в четыре раза больше.)

[46] Так, при запросе *политика* (без кавычек) не находятся тексты, где встречаются только словоформы множественного числа, омонимичные формам существительного *политик* (одушевл. муж. р.). В результате на некоторых специализированных сайтах или в их разделах выдача на запрос *политика* (без кавычек) может быть меньше, чем на *"политик"* (хотя в последнем случае во всех выданных документах имеется в виду род. мн. от неодушевленного существительного). Примером может служить расширенный поиск по сайту panvasoft.com/rus/blog, где 4.01.2011 на запрос *политика* (без кавычек) «Нашлось 22 ответа», а на *"политик"* «Нашлось 29 ответов» (все они содержат словоформу род. мн. неодушевл.). Запросы на более редкие словоформы не так урожайны, но по ним выдаются иначе не находимые документы.

21.ru/dialog2009/materials/html/05.htm).
БТС: Большой толковый словарь русского языка. / Под ред. С. А. Кузнецова. — СПб.: Норинт, 1998.
Киеня С. Н. Корпусы в социолингвистическом аспекте // Наука-2010: сборник научных статей. Ч. 1. Гродно: ГрГУ им. Я.Купалы, 2010
МАС: Словарь русского языка: В 4 т. / Под ред. А. П. Евгеньевой. — 2-е изд., испр. и доп. — М.: Рус. яз., 1981—1984.
Мельчук И. А., Жолковский А. К., Апресян Ю. Д. и др. Толково-комбинаторный словарь современного русского языка: Опыты семантико-синтаксического описания русской лексики. Wien: Wiener slavistischer Almanach, 1984.
Николенкова Н. В. Орфографический словарь и кодификация современной нормы: проблемы несогласованности // Вопросы культуры речи. М.: АСТ-Пресс Книга, 2011.
Новый объяснительный словарь синонимов русского языка. / Под рук. Ю. Д. Апресяна. 2-е изд., испр. и доп. М.; Вена: Языки славянской культуры; Wiener slavistischer Almanach, 2004
Орфографический словарь русского языка / Под ред. С. И. Ожегова и А. Б. Шапиро. — М.: Гос. изд. иностр. и нац. словарей, 1956.
Плунгян В. А. Корпуса и полевая лингвистика // I Международный симпозиум по полевой лингвистике. Тезисы докладов. М.: ИЯз РАН, 2003.
Плунгян В. А. Зачем мы делаем Национальный корпус русского языка? // «Отечественные записки», 2005, № 2.
РОС: Русский орфографический словарь: ок. 180000 слов / Отв. ред. В. В. Лопатин. Изд. 2 — М.: ИРЯ РАН, 2007.
Толковый словарь русского языка: В 4 т. / Под ред. Д. Н. Ушакова. — М.: СЭ; ОГИЗ, 1935—1940.
Толковый словарь русского языка с включением сведений о происхождении слов / Ин-т рус. яз. РАН. Отв. ред. Н. Ю. Шведова. — М.: Азбуковник, 2007.
Штейнфельдт Э. А. Частотный словарь современного русского литературного языка. — Таллин: НИИ педагогики ЭССР, 1963.

Post scriptum

В ходе переписки с коллегой получил такую реакцию на один частный аспект настоящего текста: *Всё нормуль (для меня неожиданно, что нормуль, кажется, побивает нормалек)*. И для меня неожиданно.

Собственный узус коллеги мало отличается от моего, но у него есть сын-студент, а в трудовом коллективе он регулярно сталкивается с десятками лиц в возрасте до 30 лет, так что, будучи чутким к языку, регулярно доносит до меня разные новшества. Сам я про *нормуль* сказал бы «слышал, но и только-то».

Казалось бы, логичная история развития: сначала графическое закрепление естественного процесса в безударном вокализме при небрежном стиле произнесения — *нормалёк* → *нормулёк* (это если есть нужда в графической фиксации), потом де-деминутивизация — *нормулёк* → *нормуль*, но ниже будет показано, что фактического подтверждения такая эволюция не находит.

Проверять, как обстоят дела в узусе, можно по НКРЯ:

НКРЯ	1981—1989	1990—1999	2000—2002	2003—2007
нормалек	3	6	7	6
нормуль	0	0	0	4

Нормульков нет, что не удивляет. Газетный корпус добавляет четыре *нормалька* за 2001—2006 гг. Хронология *нормулей* в НКРЯ такова: сначала статья в журнале «Хулиган» (2003, № 12), затем девятая книга из сказочного сериала «Таня Гроттер» (2004; «для умеющего читать, но не желающего читать ребенка» от 8 лет, как сказано на сайте автора, Дмитрия Емца; москвич 1974 г. р., канд. филол. наук) и две цитаты с форумов (2005, 2007). *Нормалек* за те же годы встретился в детективах Д. Донцовой («Микстура от косоглазия» и «Уха из золотой рыбки»; 1952 г. р.) и С. Данилюка («Бизнес-класс» и «Рублевая зона»; 1952 г. р.), а также на двух форумах.

Строго говоря, информации хватает «к размышлению», а для достоверных выводов ее маловато. То, что *нормуль* появляется лишь в 2003—2004 гг. ничего не означает: тексты адресованы подросткам и детям, а их узус часто игнорируется как в текстах, написанных взрослыми, так и словарях. Сам я вряд ли до 2003 г. слышал это слово, но полагаться на интроспекцию давно отвык, особенно в том, что касается словарной периферии (см. Прибавление 9).

В ЖЗ на начало 2011 г. *нормалек* представлен 38 текстами, *нормуль* — четырьмя; по време-

ни публикации они распределены так:

	нормалек	*нормуль*
1990-е:	2	1
2000—2005	11	2
2006—2010	25	1

Никакой экспансии *нормулей* незаметно, если слепо верить в цифры, скорее наоборот, *нормальки* торжествуют. Не видно и молодежности, поскольку в текстах с *нормулями* расчет на молодежную аудиторию не просматривается, а года рождения авторов — 1928, 1952, 1959, 1976. Средний возраст, как сейчас принято говорить, «взрослый», на сегодня под 58 лет.

В современной блогосфере употребимость этих слов можно замерять лишь на кратких интервалах. За пять дней 12—16 января *"нормалек"* нашелся у 612 блоггеров, *"нормуль"* — у 976. При других замерах в начале 2011 г. соотношение обычно укладывается в интервал 1,4—2,0 в пользу *нормуля*. Это как раз подтверждает диагноз коллеги («нормуль, кажется, побивает нормалек»).

К сожалению, распределение блоггеров по возрасту сейчас не работает, но кое что можно выявить из общих цифр. Исторически ситуация выглядела следующим образом.

	"нормалек"	"нормуль"	соотношение
2001—2003	209	261	1,2
янв-июн 2004	212	458	2,2
июл-дек 2004	300	762	2,5
янв-мар 2005	217	696	3,2
апр-июн 2005	312	973	3,2
июл-авг 2005	235	689	2,9
сен-окт 2005	277	967	3,5
ноя 2005	187	704	3,8
дек 2005	274	1005	3,7

Пока рост блогосферы позволял получать полные цифры, статистика была достоверной и легко интерпретируемой. Среди блоггеров *нормуль* начал побеждать уже 10 лет назад, а к концу 2005 явно торжествовал. Особенно примечательно лето 2005: рост нормуля несколько затормозился. Но это как раз явное свидетельство того, что *нормуль* слово более молодое. То же подтверждается и статистикой по использованию словоформ, отличных от им.-вин. ед. ч.[47]. Тех, кто эти слова склоняет, немного, зато можно получить полные погодовые (то есть абсолютно достоверные) данные по всей блогосфере:

	нормуль склоняемый	*нормалек* склоняемый
2001—2004	10	5
2005	20	12
2006	80	47
2007	174	62
2008	88	34
2009	124	86
2010	327	180

Динамика *нормалька* более сглажена, восстановление после «провала» за счет ухода в социальные сети произошло быстрее. Именно это происходит с лексикой, использование которой по возрастам распределено равномернее.

В неизменяемой форме соотношение вариантов иное, но о причинах без точного распределения по возрасту и полу можно только гадать.

---

[47] Вот четыре синонимичные фразы из блогов:

у меня всё в нормуле))))
у мя усе нормули..!!!
у мя все нормуля))
Всё будет на нормулях!