

## Фонотактика корня в русском языке и ее потенциальные ресурсы

### Постановка вопроса

Проблема, которая обсуждается в этой статье, привлекла внимание автора, когда ему пришлось, как и многим, осваивать новую версию программы EXCEL (2007). Эта версия отличается от предыдущих, в частности, резким увеличением количества столбцов и строк. Если в старых версиях электронная таблица включала сотню столбцов, то в новой версии их 16384. Каждый из этих столбцов обозначается одной из 26 латинских букв или их комбинацией, при этом максимальное количество букв, используемых для различения столбцов, равно трем. Таким образом, трехбуквенные комбинации из 21 согласного и 5 гласных позволяют сформировать многотысячный список «слов».

Для специалиста по статистике этот факт – более чем банален: трехчленные комбинации из 26 букв дают  $26^3$ , то есть 17576 комбинаций<sup>1</sup>. Для «безнадежных гуманитариев» приведенный факт скорее удивителен, чем банален, в чем автор убедился, проведя мини-опрос среди коллег. Интуитивно многие лингвисты полагают, что, для создания словаря язык использует свой инвентарь фонем довольно активно – «процентов на 50%», учитывая, что некоторые комбинации фонем язык относит к «нежелательным». Статистически очевидный факт оказался неожиданным и для автора. Сколько раз доводилось читать, слышать (а иногда и говорить студентам), что-то вроде того, что «малое количество гласных в языке должно компенсироваться большим количеством согласных, чтобы язык мог различать корни» или что «одна из причин возникновения тонов связана с ограниченными возможностями различения корней – например, при запрете на сочетания согласных»! На чем основаны эти расхожие суждения?

Оценим, какие средства нужны языку для того, чтобы различать необходимое ему количество корней. Договоримся, что в нашей оценке не допускается существование омонимичных корней, хотя это вполне нормальное явление: русский язык без проблем различает 3 корня **рж** в *ржавый*, *ржаной* и *ржать*, 2 корня **треск** в *треск* и *треска*, 3 корня **порт** в *портной*, *напортачить* и *портовый*, 3 корня **губ** в *погубить*, *губа* и

---

<sup>1</sup> В EXCEL используются не все комбинации – программа останавливается на столбце XFD.

у*сугубить*, 5 корней **кут** в *кутать*, *кутить*, *закуток*, *кутья*, *кутенок*, 6 корней **пар** в *парилка*, *вспарывать*, *выспаривать*, *пара*, *парень*, *воспарить* и т.д.

В стандартном языке имеется 2000-3000 корней<sup>2</sup>. Какими средствами должен располагать язык для их различения? Рассмотрим несколько возможных сценариев.

*Первый сценарий.* Будем считать, что язык не допускает ни сочетаний согласных, ни закрытых корней, и в нем возможны только корни V, CV и CVCV. В этом варианте для того, чтобы построить 2500 корней, языку необходимо иметь всего 10 согласных и 5 гласных.

*Второй сценарий.* Наш язык моносиллабичен, при этом сочетания согласных в нем не допускаются. В нем возможны корни V, C, CV, VC, CVC, VCV. В этом варианте наличие 20 согласных и 5 гласных дает возможность построить 2725 корней.

В языке, в котором допускаются корни CCVCC (такой язык мог бы быть моносиллабичным!), а это случай русского языка, при 20-ти согласных и 5-ти гласных только в рамках этой структуры можно теоретически построить 800 000 корней! Заметим, что в русском языке зафиксировано лишь 88 корней этого типа, то есть 0,01%. Учитывая, что в русском (даже оставляя в стороне оппозицию согласных по твердости / мягкости) имеется 21 согласный, для русского языка это вполне точная цифра: в комбинации CCVCC русский язык использует одну комбинацию из каждых 10000 теоретически возможных. А ведь это далеко не единственная комбинация, характерная для русского корня – есть десятки других! Все это означает, что теоретическая матрица допускаемых структур корня в таких языках как русский, насчитывает миллиарды возможных теоретических корней, и из этого бесконечного мира возможностей реализованы лишь 2936 – именно столько корней в русском языке, если ориентироваться на Словарь морфем русского языка [Кузнецова–Ефремова 1986],

---

<sup>2</sup> «Интересно, что в этимологически хорошо разработанных семьях с большим ветвлением (типа индоевропейской, финно-угорской, алтайской, восточнокавказской, австронезийской, синотибетской и т. п.) число реконструируемых корней приближается к некоторой константе, колеблющейся около 2-2.5 тысяч. Это число, по-видимому, не зависит от временной глубины соответствующей семьи, поскольку совершенно аналогичные цифры мы получаем для молодых семей типа славянской, германской, тюркской и т.п. Наличие подобной константы можно объяснить только тем, что реконструируемые праязыки представляли собой реальные языковые образования со средним для любого языка числом морфем. Гипотеза о том, что любой язык имеет конечное число корней, колеблющееся между 2-3 тысячами, вполне вероятна, хотя и нуждается в дополнительной типологической проверке. Так, словарь морфем современного русского языка [Кузнецова, Ефремова 1986] насчитывает 4400 корней, хотя реальное число, видимо, существенно меньше (в качестве отдельных единиц в словарь включены варианты корней типа *оберт-* при *верт-* и т. п.)». [Старостин 2007, 792].

исключив из него многочисленные алломорфы<sup>3</sup>. А это означает, что шесть корней **кут** в русском языке – с вероятностной точки зрения, это чудо, в то время как отсутствие конкретного корня (например, корня **кет** или **куж**) – вполне ожидаемое явление.

Эти предварительные замечания подводят нас к следующей постановке вопроса. В области фонотактики давно и успешно изучаются как запреты, характерные для определенных языков, так и комбинации фонетических единиц, наиболее характерные для того или иного языка. Однако если статистически оценить «фонемный ресурс» и возможности его использования, мы видим, что: 1) часть выделяемых нами «запретов» может оказаться естественным следствием вероятностных закономерностей (так, например, в отсутствие корня **куф** в русском языке нет ничего «странного» (можно сказать иначе – жто не лингвистический вопрос), поскольку для корпуса в 3000 корней, ожидаемое число комбинаций **куф**, с учетом низкой частоты /ф/, равняется 1, а если исходить из всех возможных комбинаций в русском корне – то твердому нулю; 2) многие корни, которые не привлекают внимания лингвистов (не думаю, что в русистике когда-либо отмечалась частотная аномалия, связанная с корнем **кут**), оказываются достойными самого пристального внимания, поскольку кардинально отклоняются от вероятностных характеристик. Отсутствие корня **труз** – нормально (как, впрочем, и отсутствие любого другого конкретного корня), а вот наличие 3 корней **трус** (*трусливый, трусы, утруска*) – исключительно маловероятное событие.

К этому (на мой взгляд, кардинальному) вопросу фонотактической теории (и фонотактической типологии языков) мы вернемся в конце статьи. А сейчас обратимся к корневому словарю русского языка, чтобы на конкретном и хорошо нам знакомом материале оценить возможности вероятностной фонетической комбинаторики и их реальное проявление в языке. Надеюсь, что рассматриваемые ниже данные могут представлять самостоятельный интерес для русистов, вне зависимости от интерпретации полученных данных, предлагаемой в данной статье.

Последнее важное замечание. Для исследования фонотактической проблематики (тем более, статистического исследования) корневой словарь, по-видимому, подходит значительно лучше, чем самый полный словарь слов русского языка. Назову лишь одну принципиальную причину. Преимущество корневого словаря состоит в том, что в нем нет морфемных границ, а фонотактические правила очень чувствительны к этому

---

<sup>3</sup> Цифра 4400 корней, которая указана в Словаре морфем и которая часто цитируется (см, в частности, цитату в примечании 2) сильно завышена, и, в первую очередь, за счет включения в список алломорфов.

фактору – на морфемной границе вполне допускаются такие сочетания, которые не разрешаются внутри морфемы. Так, например, внутри морфемы в русском языке не допускается сочетание **пVф**, в то время как при наличии морфемной границы это сочетание вполне возможно: ср., *по-французски*, *пофартить*, *пофамильярничать*, *пофантазировать*, *пофехтовать*, *пофланговый*, *пофлиртовать*, *пофонемный*, *пофотографировать*. То же самое происходит с сочетанием **пVб** – практически любой глагол с начальным корневым **б-** может сочетаться с приставкой **по-** (*побежать*, *побыть* и т.д.), между тем, как внутри морфемы сочетание **пVб** практически не допускается.

Здесь мы сталкиваемся с чрезвычайно любопытным семиотическим законом: наш язык очень любит минимальные пары, состоящие из двух разных знаков – как **в парадигме** (*куда? – туда!*, *какой? – такой!*, *как? – так!*, *наши – ваши*, *notre – votre*, *where? – there!*, *what? – that!*), так и **в синтагме** (*такой-сякой*, *культур-мунтур*, *тик-так*, *там-сям*, *tel-quel*, *гоголь-моголь*): при этом, язык активно избегает минимальных пар внутри одного знака<sup>4</sup>. Для комбинации **бVр** в русском языке отмечены всего три слова – *бантист*, *биполярный* и *биплан* [Зализняк 1977]. Помимо того, что такой набор «русских» слов вызывает улыбку, отметим, что носитель, который включает в свой словарь эти слова, очевидно, ощущает прозрачную морфемную границу в двух последних словах<sup>5</sup>.

### Фонотактика русского корня

В Словаре морфем русского языка [Кузнецова–Ефремова 1986] (далее СМРЯ), если исключить из него алломорфы, – 2936 корней. С их полным списком можно ознакомиться на странице <http://pozdnjakov.free.fr>. На этом списке и основываются дальнейшие подсчеты. Рассмотрим несколько важных для нас аспектов, характеризующих русский корень.

#### Структура корня и слог.

Четырехсложных корней в СМРЯ нет. Из 2936 корней: 1893 – односложные или не имеют гласных, 941 – двусложные и 102 – трехсложные. Структура корней такова (в скобках для каждой структуры приводится количество корней):

<sup>4</sup> Подробно об этом см. [Поздняков 2009].

<sup>5</sup> Подробнее см. [Поздняков 2007].

- Односложные (или не имеющие гласных) корни – 18 типов корней: С (4), СС (15), ССС (3), ССССVC (1), СССV (1), СССVC (34), СССVCC (2), ССV (14), ССVC (406), ССVCC (88), CV (18), CVC (889), CVCC (330), CVCCC (22), CVCCCC (1), V (4), VC (40), VCC (20);
- Двусложные корни – 29 типов корней: СССVCVC (5), СССVVC (1), ССVСССVC (1), ССVССVC (13), ССVCV (1), ССVCVC (54), ССVCVCC (4), ССVVC (2), CVСССVC (1), CVСССVC (4), CVССV (9), CVССVC (134), CVССVCC (2), CVCV (11), CVCVC (525), CVCVCC (50), CVV (1), CVVC (6), CVVCC (3), VСССVC (1), VСССVCC (1), VССVC (27), VССVCC (4), VCV (3), VCVC (73), VCVCC (3), VCVCCC (1), VV (1), VVCC (1);
- Трехсложные корни – 13 типов корней: ССVCVCVC (6), CVССCVCCVC (2), CVССVCVC (1), CVССVVC (1), CVCVCССVC (9), CVCVCVC (61), CVCVCVCC (6), CVCVVC (1).

Таким образом, в СМРЯ – 60 типов корней. Чисто механически можно было бы предложить примитивный показатель вариативности корней в языке: отношение количества структур к количеству корней, то есть для русского корневого словаря – приблизительно 50 корней на одну структуру. Но это, конечно, неудачный критерий: во-первых, как мы видим, корни распределяются по структурам крайне неравномерно, во-вторых, многие структуры представлены одним корнем (16 структур), и при этом, большинство из них находятся на периферии русского языка, представляя собой заимствования или идеофоны (или же этимологически являются составными корнями). Приведем полный список корней, образующих «уникальные» структуры: ССССVC – **встрет**, СССV – **тпру**, CVСССС – **черств**, СССVVC – **страус**, ССVСССVC – **крестян**<sup>6</sup>, ССVCV – **стара**, CVСССVC – **завтра**, CVV – **мяу**, VСССVC – **остров**, VСССVCC – **искренн**, VCVССС – **электр**, VV – **ау**, VVCC – **аист**, CVССVCVC – **барвинок**, CVССVVC – **чемпион**, CVCVVC – **караул**.

Если о чем-то и говорит данный список, то лишь о том, что приведенные в нем корни нам не интересны – они интересны лишь своей уникальностью<sup>7</sup>. К явно маргинальным относятся и многие типы корней, которые представлены не одним, а десятком

<sup>6</sup> Твердый и мягкий знаки (в неконечной позиции) были автоматически заменены на **й**, интерпретированный как согласный. Корни типа *рояль* и *ед* интерпретировались как CVCVC и CVC соответственно. Заметим попутно, что, несмотря на ряд «фонетических поправок», здесь исследуется скорее графическая, чем фонетическая структура корня.

<sup>7</sup> В них можно играть – «Найдите единственный корень русского языка структуры CVССVVC!».

примеров. Так, например, структура CVCVCCVC представлена девятью корнями, причем, большинство этих корней несомненно находится на периферии русского языка. Вот их список: **басурман, виноград, карандаш, колошмат, коростел, кочевряг, монастыр, набекрень**. И все же, загадка остается, и нам приходится взять «басурмана» на «карандаш», поскольку с точки зрения статистики здесь все «набекрень»: девятикратное присутствие этой структуры в словаре – совершенно исключительное событие с вероятностной точки зрения. Какой Коловрат (CVCVCCVC) помогает этим странным корням сохраниться в русском языке?

Исключим механически из нашего списка структуры, частота которых меньше пяти. У нас останется 26 типов корней, которые с вероятностной точки зрения являются супер-частотными:

- Односложные (или не имеющие гласных) корни – 11 типов корней: CC (15), CCCVC (34), CCV (14), CCVC (406), CCVCC (88), CV (18), CVC (889), CVCC (330), CVCCC (22), VC (40), VCC (20);
- Двусложные корни – 11 типов корней: CCCVCVC (5), CCVCCVC (13), CCVCVC (54), CVCCV (9), CVCCVC (134), CVCV (11), CVCVC (525), CVCVCC (50), CVVC (6), VCCVC (27), VCVC (73);
- Трехсложные корни – 4 типа корней: CCVCVCVC (6), CVCVCCVC (9), CVCVCVC (61), CVCVCVCC (6).

Оценим, сколько корней можно построить, имея 21 согласный и 5 гласных в рамках этих – наиболее частотных – структур. Ответ: 1 781 389 932, то есть почти 2 миллиарда. Почему же при таком количестве возможностей, язык, вместо того, чтобы освоить хотя бы одну незанятую ячейку, предпочитает сохранять в одной ячейке шесть омонимичных корней **пар**?!

### **Звуки.**

В нашем корпусе – 2936 корней. Эти корни включают 12269 звуков, то есть в среднем русский корень состоит из 4,2 звуков. На 1 корень в среднем приходится 2,8 согласных и ровно в 2 раза меньше гласных, т.е. 1,4. Частоты согласных и гласных распределяются в словаре корней неравномерно.

Гласные (4060 гласных в корпусе) характеризуются следующими частотами (в %):

о/ё	а/я	е/э	у/ю	и/ы
26,3	26,1	20,0	15,0	12,8

Таким образом, более половины гласных в русских корнях – это гласные /о/ и /а/. Согласные в корнях (8209 согласных) имеют следующие частоты (в %):

<b>р</b>	13,5	<b>б</b>	5,0	<b>ш</b>	2,5
<b>л</b>	9,5	<b>в</b>	4,8	<b>ч</b>	2,4
<b>к</b>	9,4	<b>д</b>	4,7	<b>й</b>	2,1
<b>т</b>	9,0	<b>м</b>	4,4	<b>ж</b>	1,8
<b>с</b>	7,7	<b>г</b>	4,3	<b>щ</b>	0,9
<b>н</b>	5,9	<b>х</b>	2,8	<b>ф</b>	0,7
<b>п</b>	5,4	<b>з</b>	2,7	<b>ц</b>	0,6

Из данных таблицы видно, что 7 самых частотных согласных – **р, л, н; п, т, к** и **с** – составляют более половины консонантного фонда русской корневой базы. Каждый пятый согласный в русских корнях – плавный (**р** или **л**). Это означает, что, если мы хотим вывести наиболее общие фонотактические правила для русского корня, мы должны отдельно исследовать комбинаторику плавных.

### **Роль плавных в русском корне.**

**А. Плавные в структуре CVCCC.** Проиллюстрируем особую роль плавных на примере экзотической односложной структуры CVCCC. Прежде всего, отметим, что эта структура слога экзотична для большинства языков, но не для русского. В СМРЯ отмечено 22 односложных корня структуры CVCCC. Для трех тысяч корней это не просто много – это исключительно много. При этом корни этой сложной структуры не являются для русского языка маргинальными – если среди них и есть заимствования, то они вполне основательно интегрировались в русский язык. Вот список этих корней: **быстр, австр, верст, волхв, выспр, горст, жертв, жестк, костр, мездр, муштр, ноздр, перст, пестр, порск, сестр, толст, фортк, хандр, холст, шерст, шустр.**

Обратим внимание на важную деталь: в этой структуре практически нет **ни одного** трехконсонантного сочетания согласных без плавных – [р] или [л] (единственное исключение – корень **жестк**). Предполагая, что мы ничего не знаем об исторической фонетике русского языка (автору это нетрудно сделать), рассмотрим комбинаторику плавных в трехконсонантных сочетаниях других структур русского корня.

**Б. Плавные в составе комбинаций ССС.** Сочетания ССС отмечены в составе 79 корней. Из них только 9 корней не содержат ни [р], ни [л] (4 корня СКВ – **сквалыг, скверн, сквоз, сквор**; 2 корня ШКВ – **шквар, шкворен**; плюс 1 СТВ – **ствол**, 1 СТЙ – **крестян** и 1 СТК – **жестк**); соответственно 70 корней из 79 включают [р] или [л], а

одном случае (**верблюд** – РБЛ) – [р] и [л] одновременно. При этом, [р] в таких сочетаниях встречается на порядок чаще, чем [л]: в ССС 61 корень с [р] и только 8 – с [л] (3 ЛСТ – **галстук, толст, холст**; 3 СКЛ – **скл, склаб, склиз**; плюс МГЛ – **мгл** и ЛХВ – **волхв**). Еще одна интересная деталь: в сочетаниях ССС(С): только в 8 корнях [р] занимает начальную позицию (5 РСТ: **верст, горст, перст, шерст, черств**, 1 РСК – **порск**, 1 РТК – **фортк**, 1 РТВ – **жертв**). В средней позиции он не встречается вообще (как и [л])<sup>8</sup>, а в подавляющем большинстве случаев (53 корня) встречается в конечной позиции – причем, в крайне ограниченном наборе сочетаний, из которых львиная доля падает на СТР (36 корней), остальные же комбинации распределяются следующим образом: 6 СКР, 3 ШТР, 2 ЗДР и по одному – СПР, СБР, ТПР, КТР, ВТР, НДР.

В принципе [р] и [л] в русских корнях заслуживают особого внимания не только в сочетаниях согласных. Они играют ключевую роль и в комбинациях CVC. Рассмотрим этот вопрос подробнее.

**В. Плавные в составе комбинаций CVC.** В [Pozdniakov-Segerer 2007] [Поздняков 2007] [Поздняков 2010] подробно рассмотрены универсальные ограничения на сочетаемость согласных одного места образования в комбинациях CVC в самых разных языках (для этого явления мы предложили термин SPA – Similar Place Avoidance)<sup>9</sup>. В частности, на материале СМРЯ были получены следующие данные (таблица 1):

Таблица 1

1а Русский язык (5 локальных рядов)

I	Р	К	Т	С	С
Р	---		+		
К		---	+	-	-
Т	++	++	---	+	+
С	+	-	+	---	---
С				---	-

1б Русский язык (4 локальных ряда)

	Р	К	Т	С
Р	---		+	
К		---	+	-
Т	++	++	---	+
С				---

*Комментарий:* В (1а) согласные сгруппированы по 5-ти рядам: 1) Р – **п, б, ф, в, м**; 2) К – **к, г, х**; 3) Т – **т, д, л, р, н**; 4) С – **с, з, ц**; 5) С – **ч, ж, ш, щ, й**. В (1б) объединены данные по рядам S и C.

Представленные в таблице данные показывают системный характер отклонений. В корнях русского языка в комбинациях структуры CVC резко ограничиваются сочетания согласных одного места образования. Ограничиваются также сочетания между

<sup>8</sup> В связи с утратой между согласными \*ьг, \*ьл, \*ьр, \*ьс и \*гь, \*ль, \*гь, \*ль и их переходом соответственно в **ор, ол, ер, ел** и **ро, ло, ре, ле**, это вполне понятно.

<sup>9</sup> В [Mayer-Rohrdantz-Plank-Bak-Butt-Keim. 2010] явление SPA было подтверждено на материале 4500 языков, что дает основание отнести его к разряду универсалий. См. также [Mayer 2010].



препалатальными (**с,з,ц**) и палатальными (**ч,ж,ш,щ,й**)<sup>10</sup>. Несколько меньше ожидаемой нормы встречаются сочетания, в которых первый согласный велярный, а второй – (пре)палатальный. Наиболее предпочтительными комбинациями (знак «++») являются комбинации, в которых первый согласный «центральный» – Т (**т,д,л,р,н**), а второй – «периферийный» – лабиальный Р (**п,б,ф,в,м**) или велярный К (**к,г,х**): **TVР, TVК**. Следующие по предпочтительности комбинации также предполагают сочетание периферийных и центральных согласных, но в них вектор направлен от периферии к центру: **PVT, KVT** (знак «+»). Комбинации между периферийными согласными разного места образования (**PVK, KVP**) не ограничиваются, но и не относятся к предпочтительным (отсутствие значимых отклонений). Все эти тенденции практически полностью отвечают универсальной иерархии «запретов» и «разрешений» в языках мира (за исключением нетипичного отрицательного отклонения в комбинации **KVC**, а также положительного отклонения в **TVC**).

Для детального исследования этого явления по материалам СМРЯ была составлена таблица сочетаемости согласных в комбинациях CVC. В предыдущих публикациях она не приводилась. Приведем ее здесь (таблица 2).

Таблица 2

	п	б	ф	в	м	т	д	р	л	н	с	з	ц	ч	ж	ш	щ	й	к	г	х	%
п	7	0	0	3	3	24	10	55	40	14	24	7	2	7	2	2	1	7	14	9	10	6.4
б	0	6	1	1	1	21	18	64	44	15	14	12	1	6	1	5	0	9	17	8	7	6.7
ф	0	1	1	1	0	6	0	11	3	6	2	0	0	0	0	0	0	0	2	0	0	0.9
в	4	3	1	0	1	22	10	76	30	23	19	11	1	1	7	2	5	10	11	5	6	6.6
м	0	1	0	0	3	22	8	46	22	31	19	13	0	6	5	15	1	2	18	11	12	6.3
т	21	8	2	7	13	8	4	77	38	27	10	4	2	3	2	4	1	7	21	16	5	7.5
д	1	7	0	9	5	3	4	24	20	9	8	0	1	2	8	6	0	3	9	6	4	3.4
р	39	35	6	32	36	32	39	2	17	49	64	41	1	13	27	10	4	11	42	38	35	15.3
л	32	26	1	26	15	29	34	2	6	49	44	24	1	12	5	13	8	12	47	22	21	11.4
н	4	5	1	9	3	14	5	14	4	5	9	4	4	1	2	1	1	3	7	9	3	2.9
с	7	5	0	7	9	14	10	32	33	22	9	4	1	3	1	0	0	5	13	4	3	4.9
з	1	6	0	5	6	3	4	8	9	11	1	1	0	0	0	0	1	3	4	1	2	1.8
ц	6	0	1	1	0	0	1	5	2	4	0	0	1	0	0	0	0	1	2	1	1	0.7
ч	4	6	0	5	5	6	4	30	11	6	7	1	0	3	1	3	0	4	6	10	5	3.1
ж	0	2	0	2	4	3	3	15	17	8	3	1	0	2	2	0	0	0	1	0	2	1.7
ш	4	5	3	4	2	4	1	19	12	15	4	0	0	0	0	2	0	2	5	3	1	2.3
щ	4	4	0	1	1	3	3	8	3	4	0	0	0	1	0	0	0	0	6	2	0	1.1
й	0	4	0	3	5	2	6	18	7	9	9	5	0	2	2	0	4	1	3	5	3	2.3
к	19	15	4	22	18	27	14	76	57	31	24	7	1	8	3	10	1	5	9	2	1	9.4
г	1	4	0	7	5	10	5	31	27	15	8	5	0	1	3	0	1	4	3	5	0	3.6
х	1	3	0	0	3	6	4	16	18	6	0	1	0	0	1	0	0	1	0	0	4	1.7
%	4.1	3.9	0.6	3.9	3.7	6.9	5	16.8	11.2	9.6	7.4	3.8	0.4	1.9	1.9	1.9	0.7	2.4	6.4	4.2	3.3	3750

В таблице строки обозначают C1 в сочетании CVC, а столбцы – C2. Так, в словаре 49 комбинаций лVн и 4 комбинации нVл.

<sup>10</sup> Один из аргументов в пользу слияния этих рядов, как это и сделано в *1б*.

Нетрудно заметить, что комбинации CVC с участием [р] и [л] – самые многочисленны. Именно они как правило занимают 1-ый или 2-ой ранги как по строке, так и по столбцу, а часто – по строке и по столбцу одновременно: см., например, ячейки **вVр, тVр, кVр, рVс, лVн, лVк** и др. Из 3750 комбинаций CVC в СМРЯ, комбинации с участием [р] и [л] составляют 2024, т.е. 54%.

Можно, конечно, заметить, что в русском языке плавные согласные характеризуются очень высокой частотой, что и определяет их активность в сочетаниях CVC. Отчасти это так, но дело в том, что в комбинациях CVC их частоты существенно выше: в корневом словаре в целом составляют 13,5% для [р] и 9,5% для [л] в группе согласных (в сумме 22%), а в комбинациях CVC, как это видно из приведенной таблицы – 27% в сумме по строке (первый согласный) и 28% – по столбцу (2-ой согласный), при том, что друг с другом они практически не сочетаются.

Приведем полные данные запретов и предпочтений сочетания согласных в комбинации CVC с учетом частот каждого согласного в словаре (таблица 3).

Напомним, что плюс означает существенное превышение комбинаций по сравнению с «нормой», рассчитанной в соответствии с их частотами в словаре; минус – существенно меньшее количество комбинации относительно нормы; ноль означает, что комбинаций приблизительно столько, сколько мы ожидали; отсутствие знака означает, что норма настолько мала (речь идет о комбинациях с участием редких согласных), что мы не можем оценить в интересующем нас смысле словарные данные.

Таблица 3

	п	б	ф	в	м	т	д	р	л	н	с	з	ц	ч	ж	ш	щ	й,ь	к	г	х
п	—	—		—	—	+	—	+	+	—	+	—						+	0	0	+
б	—	—		—	—	+	+	+	+	—	—	+						+	0	—	—
ф								+													
в	—	—		—	—	+	—	+	0	0	0	+						+	—	—	—
м	—	—		—	—	+	—	+	—	+	0	+						—	+	0	+
т	+	—		—	+	—	—	+	+	0	—	—		—	—	—		0	+	+	—
д	—	+				—	—	0	+	—	—								0	0	
р	+	+		+	+	—	+	—	—	0	+	+		+	+	0		—	0	+	+
л	+	+		+	0	0	+	—	—	+	+	+		+	—	+		+	+	+	+
н						+	0	—	—	—	0								0		
с	0	—		0	+	0	0	0	+	+	—	—							0	—	—
з								0	+	+											
ц																					
ч						—	—	+	—	—	—									—	
ж								+	+	+											
ш						—		+	+	+	—								0		
щ								+													
й						—		+	—	0	+									—	
к	+	0		+	+	0	—	+	+	0	0	—		+	—	+	—	—	—	—	—
г	—	—		+		0	—	+	+	+	—	0							—	0	
х								+	+	0											

В морфемном словаре лишь 17 корней, включающих комбинацию **рVл** (например, в *куролес-ить*), при том, что мы бы ожидали найти 64 таких комбинации. В [Поздняков 2007] на многочисленных примерах иллюстрировалось следующее наблюдение: «запретные» фонетические структуры и сочетания звуков активно осваиваются экспрессивной лексикой. В частности, слова с редкими корнями часто являются идеофонами. **Сама экзотичность определенного сочетания фонем активизирует иконическую составляющую знака.** И здесь «культурные» факторы часто оказываются сильнее «природных», физиологических.

Близкое соседство двух плавных согласных, по крайней мере, для русского языка, явление «нежелательное». Язык стремится что-то сделать с такими сочетаниями – если ему предлагают сочетания одинаковых плавных, находящихся в «опасной» близости (в структурах **СС** или **CVC**), он пытается их расподобить (и в результате *\*рурь* превращается в *руль*, точно так же, как *\*arbor* в испанском дает нерегулярное изменение  $> arbol$ ). Если у нас два разных плавных, это тоже плохо, и язык стремится их уподобить, ассимилировать (и в результате *лаборатория* превращается в просторечной форме в *\*лаболаторию*). Если же ничего нельзя сделать, то слово или исчезает (рискну сказать, что слова *брыластый* и *мерлуша* – структуры **рVл** и **рл** – практически исчезли, хотя они и фиксируются в СМРЯ), или же оно отодвигается на языковую периферию (*бирюльки* – **рVл**, *куролесить* – **рVл**, *лелеять* – **лVл**, *стерлядь*<sup>11</sup> – **рл**), или же, напротив, остается в основном фонде, благодаря тому, что приобретает экспрессивную окраску. Так, закреплению в языке (в качестве фразеологизма) лермонтовской строки *без руля и без ветрил*, возможно, способствовала дополнительная эмоциональная окраска в двух «запретных» сочетаниях плавных<sup>12</sup>.

Можно заметить, что, даже с учетом частот [р] и [л], в выделенных серым цветом столбцах и строках очень много плюсов и очень мало минусов (минусы системно проявляются в сочетаниях плавных между собой и в сочетаниях [н] с зубными сонантами – **нVр**, **нVл**, **нVн**).

**Г. Плавные в составе односложных корней.** Закрепив особый статус за плавными согласными, рассмотрим частоты различных корневых структур в СМРЯ, ограничив наш список лишь односложными структурами (с одним гласным), а также структурами,

<sup>11</sup> Впрочем, в постепенном исчезновении слова *стерлядь* из нашей жизни язык, по-видимому, не виноват.

<sup>12</sup> Примечательно, что Мандельштам, перефразируя Лермонтова, сохраняет двойную комбинацию **рVл** строки оригинала: «Научи меня, ласточка хилая, / Разучившаяся летать, / Как мне с этой воздушной могилой / Без **руля** и **крыла** совладать». У Мандельштама обе комбинации **рVл** входят в корни слов.

не имеющими гласных. Из 2936 корней словаря к таким структурам относится 1892 корня, то есть большинство (64,4%)<sup>13</sup> при следующем распределении (таблица 4).

Таблица 4

Односложные корни в СМРЯ

	без [l], [r]	с [l]		с [r]		с [l] и [r]		сум
С	4							4
СС	7	Сл	2	Ср,рС	6			15
ССС		ССл	2	ССр	1			3
СССVC				СССрVC	1			1
СССV				ССрV	1			1
СССVC	1	СССVл,ССлVC	3	СССVр,ССрVC	29	ССрVл	1	34
СССVCC				СССVрC,ССрVCC	2			2
ССV	6	СлV	3	СрV	5			14
ССVC	97	ССVл,СлVC	134	ССVр,СрVC,рCVC	170	СрVл	5	406
ССVCC	9	ССVСл,ССVлC, СлVCC	25	ССVСр,ССVрC, ССVрC	52	ССVрл,СрVСл	2	88
CV	17			рV	1			18
CVC	558	CVл,лVC,лVл	166	CVр,рVC,	163	рVл,лVр	2	889
CVCC	113	CVCл,CVлC, лVCC	70	CVCр,CVрC,рVCC	138	CVрл,лVCр, рVСл	9	330
CVCCC	1	CVлCC	3	CVCCр,CVрCC,	18			22
CVCCCC				CVрCCC	1			1
V	4							4
VC	34	Vл	4	Vр	2			40
VCC	11	VлC	2	VCр,VрC	7			20
	862		414		597		19	1892

Упорядочим эти структуры по мере убывания количества корней (таблица 5).

Таблица 5

CVC	558	VC	34	CCV	6	СССVрC, ССрVCC	2	ССрV	1
ССVр,СрVC,рCVC	170	СССVр,ССрVC	29	Ср,рC	6	ССл	2	ССрVл	1
CVл,лVC,лVл	166	ССVСл,ССVлC, СлVCC	25	СрV	5	ССVрл, СрVСл	2	CVCCC	1
CVр,рVC,	163	CVCCр,CVрCC,	18	СрVл	5	Сл	2	CVрCCC	1
CVCр,CVрC,рVCC	138	CV	17	С	4	рVл,лVр	2	рV	1
ССVл,СлVC	134	VCC	11	V	4	VлC	2	ССС	0
CVCC	113	ССVCC	9	Vл	4	Vр	2	СССVC	0
ССVC	97	CVрл,лVCр, рVСл	9	СССVл, ССлVC	3	СССрVC	1	СССV	0
CVCл,CVлC,лVCC	70	СС	7	СлV	3	СССVC	1	СССVCC	0
ССVСр,ССVрC,ССVрC	52	VCр,VрC	7	CVлCC	3	ССр	1	CVCCCC	0

<sup>13</sup> Приведем попутно статистику по открытым и закрытым односложным корням: из 1892 односложных корней – 1) 1826 – закрытые, 66 – открытые; 2) 1806 прикрытые, 60 – неприкрытые; 3) 4 корня имеют структуру V и 22 корня – структуру C или CC. Таким образом, подавляющее большинство корней начинается с согласного и им заканчивается.

### Двучленные сочетания согласных.

Известно, что в русском языке нет корней с сочетанием **нл**, и тот факт, что в соответствии с частотами **н** и **л** в корневом словаре, включающем 1488 парных сочетаний согласных, мы ожидали бы встретить 8 корней с **нл**, нам собственно неважно – мы знаем, что прстореч. *Фильянский вокзал* «естественнее» для русского языка, чем *Финляндский*. Статистика интересна в менее очевидных случаях, поскольку она привлекает внимание к системно неизученным явлениям фонотактики. Специалисту полезно знать, что 12 корней с **рн** в русском корневом фонде – это ровно столько, сколько мы бы ожидали в соответствии с их частотами, а 12 корней с **тв** – это ровно в 2 раза больше ожидаемого (ожидали бы 6). В словаре 10 корней с **хв** и 10 корней с **рс**, но, при этом, 10 корней с **хв** – это очень много (ожидание – 2, то есть в 5 раз меньше), а 10 корней с **рс** – это очень мало: /р/ и /с/ – частотные фонемы, и мы бы ожидали встретить в словаре не 10 сочетаний **рс**, а 16.

Приведем полную таблицу двучленных комбинаций согласных в русской корневой базе (таблица 6). В таблице 7 выделены отклонения от статистически ожидаемого количества комбинаций.

Прежде, чем проанализировать интересные детали, связанные с конкретными комбинациями согласных (таблицы 6 и 7), оценим столбцы и строки таблицы в целом (таблица 8).

#### *Комментарий к таблице 8а.*

В двух колонках отмечены для каждого согласного отклонения (в %) от статистически ожидаемых – в первой позиции (левая колонка) и во второй позиции (правая колонка).

*В группе лабиальных:* 1) ни в одной из позиций ни один лабиальный существенно не отклоняется в положительную сторону от «нормы» (несколько больше ожидаемой частота **в** во второй позиции бинарных консонантных сочетаний (+24%), но это не очень много. 2) интересное дополнительное распределение наблюдается в группе лабиальных в отрицательных отклонениях: сонант **м** и звонкий фрикативный **в** избегают первой позиции, а глухие **п**, **ф** и звонкий смычный **б** – второй позиции.

*В группе дентальных:* 1) количество **т**, **р**, **л** во второй позиции существенно превышает норму. Так, **т** в качестве второго члена входит в 281 комбинацию, а должно бы входить в 134, превышая норму более, чем на 100%. 2) Плавный **л** активно включается в комбинации во второй позиции, но избегает их в первой.

Таблица 6

	п	ф	б	в	м	т	д	р	н	л	ц	с	з	ч	ш	щ	ж	й	к	х	г	СУМ
п						5		37		38		1		2	2				2			87
ф					1	3		4		4												12
б			1				1	42		24			1			1			3			73
в						2	1	9	2			1		1	2			1	2		1	22
м	5		2					2	2	4			2	2	1				1		2	23
т	1			12				86	1	4				1		2		1	2			110
д				6	1			43	2	2			1					2	1			58
р	7	2	12	5	19	32	13	1	12	13	2	10	7	14	9	3	7	2	25	8	12	215
н		1		2	1	20	18		2		1			5			3		6		3	62
л	4	1	5	3	3	12	11		7			3	1	1	2			1	12	2	5	73
ц				1																		1
с	11		1	21	13	170			13	28		2		1					68			328
з			1	9	2		18	1	6	4							1	2			14	58
ч				2	1	3				1									2			9
ш	9			2	6	17		1	4	10									12			61
щ																						
ж			1		1		7	1	1	1							2		2		1	17
й			1		1	1	1			1	1								5		1	12
к				14		6		62	3	44		2	2									133
х				10	6	10		17	1	19												63
г				2			3	38	12	16												71
СУМ	37	4	24	89	55	281	73	344	68	213	4	19	14	27	16	6	13	9	143	10	39	1488

Таблица 7

	п	ф	б	в	м	т	д	р	н	л	ц	с	з	ч	ш	щ	ж	й	к	х	г
п	--		--	--	--		--	++	--	++		-	--					--	'-	--	--
ф																					
б	--		--	--	--	--	-	++	--	++		--	-	--	--			--	-	--	--
в	--		--	--	--	--	-		-	--		-	--	-				-	-	--	-
м	+		-	--	--	--	--	-	-	-		--			-				-	--	
т	-		--	++	--	--	--	++	--	-		--	--	-	--		--	-	-	--	--
д	--		--	++	--	--	--	++	+	-		--	-	--	--				-	--	--
р	-			-	+	++	+	--	-	-		-		++	+	+	+	-		+	+
н	--		--	--	--	++	++	--	+	--		--	--	++	--		+	--		--	
л	--			--	--		+	--		--		-	-	-	-		--	-		-	
ц																					
с	++		--	++	+	+++	--	--	++	++		-	--	-	--		--	--	++	--	--
з	--		--	++		--	+++	--	++			--							--		++
ч	--		--		-		--	--	--	-		--							-		--
ш	++		--		+	+++	--	--	+	++		--							++		--
щ																					
ж								-	-	-		--									
й	--		-	--		--		--	--	-		--							+		
к	--		--	++	--	--	--	++	-	++		-	-	--	--		--	--	--	--	--
х	--		--	++	++	+	--	++	-	++		--							--		--
г	--		--		--	--		++	++	+		--	--	--	--				--	--	--

*Условные обозначения.* Плюс означает существенное превышение фактического числа комбинаций. Особенно крупные положительные отклонения отмечены двумя плюсами. Соответственно отрицательные отклонения от статистически ожидаемого числа комбинаций отмечены минусами (двумя минусами отмечен практически полный запрет на сочетаемость согласных). Серым цветом помечены клетки, в которых статистически ожидаемая «норма» равна нулю или единице (при комбинациях согласных, имеющих низкие частоты в словаре). Таким образом, серый цвет обозначает скорее

отсутствие интересующей нас информации. Незаполненными остаются клетки, в которых фактическое число приблизительно равно ожидаемой норме.

Таблица 8

<i>8a</i>	1 <sup>ый</sup>	2 <sup>ой</sup>
п	9	-54
ф	18	-61
б	-2	-68
в	-69	24
м	-65	-15
т	-18	110
д	-16	5
р	7	71
н	-30	-23
л	-49	50
ц	-88	-52
с	185	-84
з	46	-65
ч	-75	-24
ш	63	-57
щ	-100	-53
ж	-37	-52
й	-62	-71
к	-5	2
х	50	-76
г	12	-39

<i>8б</i>	1 <sup>ый</sup>	2 <sup>ой</sup>
с	185	-84
ш	63	-57
х	50	-76
з	46	-65
ф	18	-61
г	12	-39
п	9	-54
р	7	71
б	-2	-68
к	-5	2
д	-16	5
т	-18	110
н	-30	-23
ж	-37	-52
л	-49	50
й	-62	-71
м	-65	-15
в	-69	24
ч	-75	-24
ц	-88	-52
щ	-100	-53

<i>8в</i>	2 <sup>ой</sup>	1 <sup>ый</sup>
т	110	-18
р	71	7
л	50	-49
в	24	-69
д	5	-16
к	2	-5
м	-15	-65
н	-23	-30
ч	-24	-75
г	-39	12
ж	-52	-37
ц	-52	-88
щ	-53	-100
п	-54	9
ш	-57	63
ф	-61	18
з	-65	46
б	-68	-2
й	-71	-62
х	-76	50
с	-84	185

*В группе препалатальных:* 1) существенное превышение (и только в первой позиции) показывают **с** и **з**, причем в случае **с** отмечено огромное положительное отклонение – 328 при ожидании 115; 2) и в этой группе для **с** и **з** мы наблюдаем дополнительное распределение по позициям – во второй позиции эти согласные входят в существенно меньше ожидаемого число комбинаций; употребление **ц** в бинарных комбинациях ограничено в обеих позициях.

*В группе палатальных:* 1) единственное положительное отклонение отмечено для **ш** в первой позиции. 2) Все остальные палатальные избегают участия в бинарных комбинациях в обеих позициях, обнаруживая существенные отрицательные отклонения.

*В группе велярных:* 1) единственное положительное отклонение – **х** в первой позиции. Таким образом, одинаковыми свойствами характеризуются согласные **с,ш,х,з** – активно участвуя в начале комбинаций, они ограничено употребляются в конце.

Отсортируем отклонения по их величине в первой позиции (таблица **8б**). Указанные 4 согласных – **с,ш,х,з**, – оказываются единственными согласными, обнаруживающими

существенно превышение частот в первой позиции. Избегают употребления в первой позиции аффрикаты, сонанты, а также **в**.

Наиболее обобщенные данные по второй позиции представлены в таблице **8в** (она отличается от двух предыдущих таблиц только сортировкой). Существенно больше нормы встречаются здесь дентальные **т,р,л** и **только** они. Половина согласных встречается существенно реже ожидаемого, то есть ограничение на употребление в качестве второго члена комбинаций распространяются на существенно большее число согласных, чем это происходит в первой позиции. Внизу таблицы оказываются **с** и **х**, то есть именно те согласные, которые наиболее активны в первой позиции.

Вернемся к таблице 7, в которой обобщены статистические результаты по конкретным комбинациям. Здесь много интересных нюансов:

С губным **в** во второй позиции охотно сочетаются шумные других рядов **т** и **д**, **с** и **з**, **к** и **х**, но не губные.

С дентальными плавными **р** и **л** во второй позиции охотно сочетаются начальные лабиальные **п,б** и веллярные **к, г, х**, но не дентальные – **т** и **д** активно открывают комбинации с **р**, но не с **л**. Напротив, палатальные **с** и **ш** очень активно сочетаются с **л**, но не с **р**.

При этом, **с** и **ш** активно сочетаются с глухими смычными **п** и **к**, в то время как звонкий **з** обнаруживает повышенную сочетаемость с **г** в комбинации **зг**.

С сонантом **н** во второй позиции активно сочетаются начальные звонкие **г,д,з**, а также **с, ш**. Отметим, что единственный плюс «в диагонали», фиксирующей сочетания двух одинаковых согласных, отмечен в клетке **нн**.

В группе дентальных отчетливо наблюдается дополнительное распределение по признаку «смычный»: смычные дентальные в первой позиции сочетаются только с несмычными дентальными во второй позиции: – **тр, др, дн**, а несмычные дентальные в первой позиции активнее, чем мы бы ожидали, сочетаются со смычными дентальными во второй позиции – **рт, рд, нт, нд, лд**.

В суперчастотных комбинациях препалатальных с дентальными доминантной оказывается корреляция по звонкости: **ст** и **зд**.

Начальный **р** активно комбинируется с конечными **х** и **г**.

И, наконец, отметим повышенную сочетаемость **р** и **н** (но не **л**!) с палатальными: **рч, рш, рщ, рж, нч, нж**.



## Сходства и различия комбинаций СС и CVC в русском корне.

Комбинации СС и CVC различаются радикально в том, что в первой комбинации согласные комбинируются непосредственно, а во второй – запреты на ту ли иную комбинацию согласных снимаются, благодаря тому, что между согласными появляется определенная дистанция, заданная гласным. С другой стороны, предпочтения – положительные отклонения от статистически ожидаемых частот – могут совпадать в структурах СС и CVC, а могут не совпадать. Рассмотрим предпочтения и ограничения в структурах СС и CVC по отдельности, а затем сопоставим распределение предпочтений и ограничений в этих двух структурах, чтобы получить наиболее обобщенные характеристики сочетаемости согласных в русском корне (таблица 9).

В матрице таблицы – 441 ячейка. Если исключить из нее «серые» ячейки, остается 270 ячеек, которые делятся на четыре типа: 0, 1, 2, 1+2. Из них – 53 фиксируют существенные положительные отклонения в СС (1) и 80 – в CVC (2)<sup>14</sup>. При этих исходных данных, если предпочтения в структурах СС и CVC никак не связаны между собой, мы бы ожидали получить 16 «совмещенных» ячеек, в которых положительные отклонения отмечены как для комбинаций СС, так и для CVC ( $270 * 53/270 * 80/270 = 16$ ). Реально мы наблюдаем 26 таких ячеек (индекс «12»), то есть существенно больше. Это означает, что некоторые комбинации согласных являются для русского языка предпочтительными настолько, что языку неважно, встречаются ли эти согласные в непосредственной близости или же они разделены гласным. Можно ли выделить такие комбинации в статистическом «шуме»? Вероятно, можно. Критерием может служить «фонологическая системность» распределения индекса «12». Представим себе невероятный пример, в котором из 26-ти индексов («12») 21 попадает в одну строку, не имеющую серых клеток, а 5 – распределены несистемно. Очевидно, что вероятность этого события настолько ничтожно мала, что «сигнал» легко отделяется от «шума». В менее экстремальных (то есть в реальных) примерах, системность распределений также работает, поскольку каждое системное распределение резко понижает вероятность случайности. В нашем конкретном случае, очевидно, что системные положительные отклонения определяются особой ролью согласных **р** и **л**, о чем много говорилось выше. И для СС, и для CVC, в первой позиции существенно выше ожидаемого

---

<sup>14</sup> Отметим, что положительные отклонений в комбинациях CVC существенно больше, чем в СС, что вполне естественно.

отмечено участие **р** (6 ячеек с индексом «12» в строке Р), а для второй позиции – **р** и **л** (по 6 ячеек в соответствующих столбцах).

Таблица 9

Положительные отклонения

	п	ф	б	в	м	т	д	р	н	л	ц	с	з	ч	ш	щ	ж	й	к	х	г
п						2		12		12								2		2	
ф																					
б						2	2	12		12			2					2			
в						2		2					2								
м	1					2		2	2				2						2	2	
т	2			1	2			12		2									2		2
д			2	1				1	1	2											
р	2		2	2	12	1	12					2	2	12	1	1	12			12	12
н						12	1		1					1			1				
л	2		2	2			12		2			2	2	2	2			2	2	2	2
ц																					
с	1			1	12	1			12	12									1		
з				1			1		12	2											1
ч								2													
ш	1				1			2	12	12									1		
щ																					
ж								2	2	2											
й								2				2							1		
к	2			12	2			12	2	1				2	2						
х				12	1	1		12	2	12											
г								12	1	12											

Условные обозначения: Индекс 1 обозначает повышенную частоту комбинаций согласных в структуре СС; 2 – повышенную частоту комбинаций согласных в структуре CVC; серым цветом отмечены комбинации с редкими согласными, то есть комбинации, которые трудно оценить статистически.

Характер распределения положительных отклонений в 1, а также в 2, был рассмотрен выше. Остановимся на тех строках и столбцах, в которых эти индексы обнаруживают дополнительное распределение. Такое распределение характерно для строк М, Д и Й, а также для некоторых столбцов, в частности, П и К. Распределение в последних двух столбцах имеет слишком много общих черт, чтобы отнести его к разряду случайных. Ни в столбце П, ни в К, нет индекса «12», то есть положительные отклонения в СС и в CVC всегда имеют разнонаправленный характер. Распределение подводит к следующему выводу: занимая последнюю позицию, периферийные глухие смычные (**п** и **к**) повышенно сочетаются с дентальными в первой позиции в CVC (**тVп**, **лVп**, **рVп**, **кVп**, **лVп**), а в СС системно проявляется другое правило: глухие смычные активно

сочетаются с глухими фрикативными препалатального и палатального ряда – **сп, ск, шп, шк**, (ср. также **йк**).

Таблица 10

*Отрицательные отклонения*

	п	ф	б	в	м	т	д	р	н	л	ц	с	з	ч	ш	щ	ж	й	к	х	г
п	12		12	12	12		12		12			1	12					1	1	1	1
ф																					
б	12		12	12	12	1	1		12			12	1	1	1			1	1	12	12
в	12		12	12	12	1	12		1	1		1	1	1				1	12	12	12
м	2		12	12	12	1	12	1	1	12		1			1			2	1	1	
т	1		12	2	1	12	12		1	1		12	12	12	12		12	1	1	12	1
д	12		1		1	12	12		2	1		12	1	1	1				1	1	1
р	1			1		2		12	1	12		1						12			
н	1		1	1	1			12	2	12		1	1		1			1		1	
л	1			1	1			12		12		1	1	1	1		12	1		1	
ц																					
с			12				1	1				12	12	1	1		1	1		12	12
з	1		1			1		1				1							1		
ч	1		1		1	2	12	1	12	12		12							12		1
ш			1			2	1	1				12									1
щ																					
ж								1	1	1		1									
й	1		1	1		12		1	1	12		1							2		
к	1		1		1	1	12		1			1	12	1	1		12	12	12	12	12
х	1		1				1		1			1							1		1
г	12		12		1	1	2					12	1	1	1				12	1	1

*Условные обозначения:* Индекс 1 обозначает пониженную частоту комбинаций согласных в структуре СС; 2 – пониженную частоту комбинаций согласных в структуре CVC; серым цветом отмечены комбинации с редкими согласными, то есть комбинации, которые трудно оценить статистически.

Из 270-ти ячеек таблицы («серые ячейки» не учитываются) для 190 отмечены ограничения сочетаемости в СС (индекс 1), для 83 – в CVC (индекс 2). Если для CVC количество отрицательных и положительных отклонений приблизительно равно (83 и 80 соответственно), то в структуре СС расхождение отрицательных и положительных отклонений огромно – 190 отрицательных отклонений и лишь 53 положительных, то есть структура СС в целом характеризуется, прежде всего, системными запретами на сочетания большинства согласных. В соответствии с приведенными исходными данными, мы ожидали бы (по чисто статистическим причинам) обнаружить 58 комбинаций с индексом «12». Реально их имеется 73, то есть некоторые ограничения определяются общими фонотактическими особенностями комбинаций СС и CVC.

В системности фонотактических ограничений легко убедиться, выделив, например, следующий фрагмент таблицы, включающей ячейки с индексом 12 (таблица 11).

Таблица 11

	п	ф	б	в	м	ч	ш	щ	ж	й
п	12		12	12	12					
ф										
б	12		12	12	12					
в	12		12	12	12					
м			12	12	12					
ч										
ш										
щ										
ж										
й										

Характер сочетаний отдельных палатальных между собой невозможно оценить статистически. В тех же, клетках, которые не закрашены серым цветом, распределение ограничений поразительно системно: ни одной ячейки с индексом 12 в секторах «палатальный – губной» и «губной – палатальный» и практически тотальный запрет на сочетание «губной – губной» не только в СС, но и в CVC.

Не менее системны двойные ограничения внутри сектора «дентальный – дентальный» (таблица 12).

Таблица 12

	т	д	р	н	л
т	12	12			
д	12	12			
р			12		12
н			12		12
л			12		12

Не разрешается ни в СС, ни в CVC сочетать внутри дентальных ни смычные (в любых возможных комбинациях – **тт, тVт, тд, тVд, дт, дVт, дд, дVд**), ни несмычные (**рр, рVр, рл, рVл, нр, нVр, нл, нVл, лр, лVр, лл, лVл**). Исключение: отсутствие двойного запрета на сочетания начальных несмычных с конечным **н**.

### Сопоставление положительных и отрицательных отклонений (СС плюс CVC)

В заключение анализа статистических данных по фонотактике русского корня построим таблицу, представляющую в наиболее концентрированном виде системные предпочтения и ограничения в сочетаниях согласных в русском корне (таблица 13).

В определенном смысле сходство запретов и предпочтений в сочетаниях согласных в СС и CVC (а это базовые структуры в русском корне) можно рассматривать как наиболее существенный признак консонантной комбинаторики. По отдельности распределения плюсов минусов были рассмотрено выше. Здесь нас будет, в первую

очередь, интересовать соотношение плюсов и минусов в различных секторах таблицы. Остановимся лишь на некоторых интересных деталях.

Наиболее «нейтральными» в интересующем нас смысле являются «зеркальные» сектора «губной – палатальный» и «палатальный – губной»: в них нет ни одного плюса и ни одного минуса.

Таблица 13

	п	ф	б	в	м	т	д	р	н	л	ц	с	з	ч	ш	щ	ж	й	к	х	г	
п			-	-	-		-	+	-	+			-									
ф																						
б	-		-	-	-			+	-	+		-								-	-	-
в	-		-	-	-		-													-	-	-
м			-	-	-		-			-												
т			-			-	-	+				-	-	-	-		-				-	
д	-					-	-					-										
р					+		+	-		-				+			+	-		+	+	
н						+		-		-												
л							+	-		-							-					
ц																						
с			-		+				+	+		-	-								-	-
з									+													
ч							-		-	-		-								-		
ш									+	+		-										
щ																						
ж																						
й						-				-												
к				+			-	+					-				-	-		-	-	-
х				+				+		+												
г	-		-					+		+		-								-		

Условные обозначения: знак /+/ обозначает существенно повышенную частоту конкретной комбинации согласных в структурах СС и одновременно CVC; знак /-/ – несомненное ограничение на сочетание согласных одновременно в обеих структурах; серый цвет – комбинации с низкочастотными согласными.

В столбце препалатальных нет ни одного плюса, и это единственный среди колонок и строк сегмент, характеризующийся этим свойством.

Повторим, что подавляющее большинство плюсов отмечено в столбцах (строках) **р** и **л**. Исключение составляют 7 комбинаций, а именно: **н-т, с-м, с-н, з-н, ш-н, к-в, х-в**.

В «диагонали» секторов (например, в диагонали «губной – губной») мы наблюдаем только минусы (за исключением согласных дентального ряда). Такое распределение является еще одной иллюстрацией универсального для языков правила, ограничивающего комбинации согласных одного места образования. Повторю, что «исключения» в секторе дентальных на самом деле закономерны: здесь наблюдается

вполне четкое дополнительное распределение – смычные дентальные активно сочетаются с несмычными, плавными: **т-р**, а несмычные дентальные – со смычными: **р-д, н-т, л-д**. Вне «диагонали» наибольшая концентрация минусов обнаруживается в секторе «губной – велярный», что также согласуется с универсальным правилом, предполагающим не только ограничение сочетаний согласных одного места образования (например, «губной – губной»), но и ограничение на сочетания согласных в рамках «суперклассов» – периферийного и центрального, например: «губной – велярный» или «дентальный – палатальный» (подробнее см. [Pozdniakov-Segerer 2007]).

### Заключение

Рассмотрев основные фонотактические предпочтения и запреты в русском корне, вернемся к отправной точки статьи. В таблице 14 выделены 10 наиболее частотных односложных структур русского корня.

Таблица 14

<b>структура</b>	<b>количество</b>	<b>ресурс</b>
CVC	558	2205
CCVp,CpVC,pCVC	170	6615
CVл,лVC,лVл	166	225
CVp,pVC,	163	220
CVCp,CVpC,pVCC	138	6615
CCVл,ClVC	134	4410
CVCC	113	46305
CCVC	97	46305
CVCл,CVлC,лVCC	70	6615
CCVCp,CCVpC,CCVpC	52	138915
<b>СУММА</b>	<b>1661</b>	<b>258430</b>

Корни приведенных в таблице структур (1661 корень) составляют 88% всех односложных корней русского языка. В столбце «ресурс» приведено «теоретическое» количество корней каждой структуры при 21 согласном и 5 гласных. Оказывается, что в рамках только этих структур (а в русском корне, как мы видели, есть множество других структур, причем, не только односложных) теоретически возможно построить с таким инвентарем фонем 258 430 корней (напомним, что оппозиция согласных по твердости / мягкости здесь не учитывается!). Это означает, что даже в самых «хороших» для языка корнях, он использует лишь 0,6% возможных ресурсов, то есть ничтожно мало.

Осознание этого факта отрывает, как представляется, новые перспективы в построении фонотактической типологии языков. У двух неродственных языков с вероятностной

точки зрения есть ничтожно мало шансов поместить имеющиеся в них 2000 корней в одни и те же клетки огромной «теоретической» матрицы. Для корней структур CV, CVp, лVл, V вероятность такого события, конечно, велика, поскольку в таких корнях самые разные языки, в том числе и русский, используют очень высокий процент возможных ресурсов. Но уже для корней CVC, CVCV, CCVC процент используемых теоретических возможностей резко падает – как видно из таблицы русский язык использует лишь четверть возможностей своего инвентаря для построения корней структуры CVC. В более сложных корнях совпадения между языками вообще должны были бы практически отсутствовать. И тем не менее, их, как правило, оказывается довольно много, благодаря разнообразным общим тенденциям в фонотактике корней.

Рассмотрим конкретный пример. Сопоставляя положительные отклонения в комбинациях CC в русском корне и положительные отклонения в комбинациях CC в корнях языка волоф (атлантическая группа Нигер-Конго), мы обнаруживаем 78 несовпадений (положительное отклонение либо в русском, либо в волоф) и 13 совпадений, а именно: **mp,nt,nd,nč; st,sk; bl,gl,sl; rm, rt,rg; nn**. Первые 4 соответствия определяются наличием в волоф двух серий преназализованных смычных фонем (как глухих, так и звонких). Повышенные частоты **st,sk** наблюдаются в самых разных языках мира и требуют отдельного изучения. В волоф на порядок меньше комбинаций CC, чем в русском языке: типичная структура корня в волоф – CVC. Тем не менее, комбинации с плавными допускаются, в том случае если [l] занимает вторую позицию (отсюда совпадение **bl,gl,sl**), а [r] – вторую (французское слово *fromage* ‘сыр’, восходящее к *\*formage*, попав в язык волоф, почти восстановило утраченную французским языком исходную форму – *[formaas]*). Отсюда совпадение **rm,rt,rg**. И, наконец, единственная высокочастотная комбинация одинаковых согласных в русском корне – **nn** – находит параллель в волоф по вполне конкретной причине: волоф противопоставляет простые согласные геминированным, и естественно, что в диагонали таблицы волоф, фиксирующей комбинации одинаковых согласных, отмечены массовые положительные отклонения.

Если изложенные здесь исходные посылки верны, сопоставление фонотактических характеристик корней разных языков на большой общей матрице представляется «беспроектным проектом»: если клетки системно совпадают, значит, мы обнаруживаем общие тенденции, свойственные человеческому языку (именно так следует интерпретировать, в частности, ограничение сочетаний согласных одного ряда

в структуре SVC); если же они системно выделяются в нескольких разных зонах (а так и должно было бы быть при столько ограниченном использовании имеющихся ресурсов), открывается возможность системно выделить в языках несколько основных фонотактических типов и проследить их связь со структурой корня, с наличием геминат и другими фонологическими факторами.

### Литература:

[Зализняк 1977] А.А.Зализняк. Грамматический словарь русского языка. М.: Русский язык, 1977.

[Кузнецова–Ефремова 1986] А.И.Кузнецова; Т.Ф.Ефремова. Словарь морфем русского языка. М.: Русский язык, 1986.

[Поздняков 2007] К.И.Поздняков. *О «плохих» и «хороших» словах в русском языке* // Сборник к 60-летию А.К.Байбурина. СПб: Европейский Университет, 2007, с. 78-119.

[Поздняков 2007] К.И.Поздняков. *О природе и функциях неморфемных знаков* // Вопросы языкознания, 2009, 6, с.35-64.

[Поздняков 2010] К.И.Поздняков. *Ограничения дистантной сочетаемости согласных в русском языке* // «В пространстве языка и культуры: звук, знак, смысл». Сборник статей в честь 70-летия В.А.Виноградова. М.: «Языки славянских культур», 2010, с.113-126.

[Старостин 2007] С.А.Старостин *О доказательстве языкового родства* // С.А.Старостин, Труды по языкознанию. М.: Языки славянских культур, 2007, с. 779-793.

[Mayer–Rohrdantz–Plank–Bak–Butt–Keim 2010] Th.Mayer, Ch. Rohrdantz, F. Plank, P.Bak, M. Butt, D.A. Keim. *Consonant Co-occurrence in Stems Across Languages: Automatic Analysis and Visualization of a Phonotactic Constraint* // Proceedings of the 2010 Workshop on NLP and Linguistics: Finding the Common Ground, ACL 2010, Uppsala, Sweden, 2010, p. 70-78 <http://www.aclweb.org/anthology/W/W10/W10-2110.pdf>

[Mayer 2010] Th.Mayer. *Automatically extracting place features from the distribution of consonants in corpora*. <http://www.benszm.net/BSBWWS/Mayer.pdf>

[Pozdniakov-Segerer 2007] K.Pozdniakov – G. Segerer. *Similar Place Avoidance: A statistical universal*. // Linguistic Typology 2007, 11-2, p.307–348.