

О «плохих» и «хороших» словах в русском языке

Работая в области африканского сравнительно-исторического языкознания, мы с соавтором, Г.Сежерером, обнаружили одну фонетическую закономерность, показавшуюся нам странной (Pozdniakov, Segerer, 2007). Наши материалы показывали, что в подавляющем большинстве языков атлантической группы (а в эту группу входят приблизительно сорок языков, которые в генетическом плане отстоят друг от друга дальше, чем, например, русский и французский языки), действует специфический (как нам представлялось вначале) запрет на сочетаемость согласных одинакового места образования в корнях структуры CVC: ни в одном из языков губные практически не сочетаются с губными, зубные редко сочетаются с зубными, палатальные – с палатальными, а велярные – с велярными. Поскольку в этих языках практически нет сочетаний согласных CC и структура корня выражается формулой CVC (согласный – гласный – согласный), речь шла о запрете на дистантное сочетание согласных одного ряда, когда ограничение распространяется на два согласных, между которыми имеется любая гласная фонема. Так, например, в атлантических языках крайне редко встречаются корни вроде **mip, far, bom** (губные) или же **guk, keh, kik** (велярные).

Такого рода ограничения широко известны в семитских языках (в которых они вызывают своеобразные метатезы в трехконсонантном корне). Сегодня их выявлением активно занимается оптималистское направление в фонологии (см., например, Frisch, Pierrehumbert & M.Вгое, 2004). Большинство подобных ограничений, которые были выявлены в языках мира, связаны с сочетаниями плавных согласных [l] и [r] (иногда здесь важна последовательность: к примеру, сочетание **IVr** в составе слова допускается, а сочетание **rVI** – не допускается), а также в группе сибиллянтов – в сочетаниях [s] и [ʃ].

В обоих случаях мы имеем дело с очень близкими согласными (каждая из пар описывается только одним различительным признаком), которые при этом, если и не выпадают из фонологической системы (очень часто, например, в фонологических схемах [l] и [r] занимают изолированное положение), то характеризуются рядом важных отличительных черт. «Опасная» близость двух согласных в условиях их общего маргинального положения в системе создает определенное «поле напряжения» – когда они рядом, их трудно произносить, и язык стремится по мере возможности избегать таких сочетаний.

Практически универсальное «неудобство» сочетаний [l] и [r], с одной стороны, и [s] и [ʃ], с другой (даже если между ними помещается гласный), проявляется, в частности, в том, что в большинстве языков именно на основе этих сочетаний порождаются скороговорки, то есть трудновыговариваемые мини-тексты.

Именно это ограничение объясняет, почему Карл у Клары украл именно кораллы, а не жемчуга, а Клара у Карла украла кларнет, а не баян (вариант: «Королева Клара строго карала Карла за кражу коралла»). Ср, также : «У перепела и перепелки пять перепелят», «Была у Фрола, Фролу на Лавра наврала, пойду к Лавру, Лавру на Фрола навру», и в новоязе «В Кабардино-Балкарии валокордин из Болгарии» (сайт Скороговорки – <http://orator.kiev.ua/>). Выберем из указанного источника некоторые типичные примеры из других языков:

Украинский: «Король орел, орел король», «Пилип прилип, прилип Пилип. Пилип плаче. Пилип посіав просо, Просо поспіло, Пташки прилетіли, Просо поіли».

Французский: «Les vers verts levèrent le verre vert vers le ver vert», «La roue sur la rue roule; la rue sous la roue reste», «La grosse cloche sonne», «L'Arabe Ali est mort au lit» (вариант: «Moralité: Maure Ali, t'es mort alité»).

Английский: «Red lolly, yellow lolly», «Clowns grow glowing crowns».

Испанский: «Col, caracol y ajo; ajo, caracol y col».

Хорошо известно и ограничение на сочетание близких сибилантов, различающихся лишь одним признаком, например, [с] и [ш]. Эти согласные «слишком» близки, и поэтому при их линейной близости (в синтагме или в слове) возникают трудности с их произнесением. На этом факторе построены сотни известных скороговорок в самых разных языках, начиная с русского Саши в «Шел Саша по шоссе и сосал сушку» и кончая Сашей французским в «Sachez, mon cher Sasha, que Natasha n'attacha pas son chat!».

Ср, также, в русском языке: «Шли сорок мышей, несли сорок грошей, две мыши поплоче несли по два гроша», «Даже шею, даже уши ты испачкал в черной туши. Становись скорей под душ. Смой с ушей под душем тушь. Смой и с шеи тушь под душем. После душа вытрись суше. Шею суше, суше уши, и не пачкай больше уши», «Веселей, Савелий, сено пошевеливай».

Украинский: «Шишки на сосні, шашки на столі».

Французский: «Le chasseur sachant chasser sans son chien, est un bon chasseur», «Les chaussettes de l'Archiduchesse sont-elles sèches, archi-sèches?», «Cinq chiens chassent six chats», «Un pâtissier qui pâtissait chez un tapissier qui tapissait, dit un jour au tapissier qui tapissait: vaut-il mieux pâtisser chez un tapissier qui tapisse ou tapisser chez un pâtissier qui pâtisse?», «Suis-je bien chez ce cher Serge?», «Son chat chante sa chanson», «Seize jacinthes sèchent dans seize sachets secs», «Un chasseur sachant chasser doit savoir chasser sans son chien de chasse», «Chouette chaussures!», «Douze douches douces», «Seize chaises sèchent», «Ce chasseur sait chasser sans son chien dit le sage garde-chasse, chasseur sachez chasser sans chien!».

Английский: «She slits the sheet she sits on», «Miss Smith's fish-sauce shop seldom sells shellfish», «Sunshine city, sunshine city, sunshine city, ...», «The soldier's shoulder surely hurts!», «She sees seas slapping shores» (Там же).

Слова *Саша*, *сушка*, *шоссе*, *сушить* (русский), *chasse*, *chaussette*, *duchesse*, *sèche*, *chanson*, *sachet*, *chaussures* (французский), *sunshine*, *shores* (английский) – «неудачные», «плохие» в плане звуковой комбинаторики, и, как показывают статистические данные, большинство языков стремятся их избегать: частота таких слов как правило существенно ниже, чем она должна была бы быть в соответствии с частотами [с] и [ш] в словаре.

Другими словами, сам факт некоторых ограничений, которые налагаются на сочетания определенных согласных в самых разных языках, известен и известен давно. Неожиданным для нас было то, что в атлантических языках мы обнаружили ограничения на сочетаемость согласных одинакового места образования, а не способа образования, причем оказалось, что в каждом из атлантических языков эти ограничения проявляются системно – мы столкнулись с системным ограничением на сочетание любых согласных внутри любого локального ряда.

Это был для нас очень важный результат, потому что мы восприняли его как диахронический. Дело в том, что правомерность выделения так называемой западно-атлантической группы языков (в последнее время утвердилось название *атлантическая группа*, поскольку восточноатлантических языков не существует) не очевидна, – до сегодняшнего дня никому из исследователей не удалось представить доказательств ее существования. На языке компаративистов это означает, что никому из исследователей не удалось обнаружить атлантическую фонему, морфему, корень, синтаксическую конструкцию, характерную **только** для атлантических языков и не характерную для нигеро-конголезских языков других ветвей. По-прежнему, как и 100 лет назад, остается, следовательно, вероятность, что атлантические языки – это не генетическая, а географическая общность (замечу, что для такой интерпретации атлантических языков в

последние годы появляется все больше аргументов – все чаще приходится встречать термин «так называемые атлантические языки»). В этом контексте – обнаружение свойства, которое отличало бы атлантические языки от всех прочих языков нигеро-конголезской макросемьи, означало бы революционный прорыв в африканском сравнительно-историческом языкознании.

Но революции не получилось. И не получилось по простой причине – оказалось, что указанное выше ограничение в равной степени характерно для языков всех прочих нигеро-конголезских ветвей.

Более того, ограничения сочетаемости согласных одного локального ряда оказались ярко акцентированными не только в нигеро-конголезских языках, но и в языках всех других семей, которые мы обследовали. Статистический анализ был проведен нами на материале 15-ти генетических общностей. Были проанализированы более 30-ти языков 15-ти генетических общностей, а именно: атлантические, иджо, манде, убангийские, банту, сара-бонго-багирми, австронезийские, чадские, индоевропейские, алтайские, ностратический, баскский, кечуа, камиларой, пиджины (на английской основе).

Результаты анализа оказались поразительно сходными, что позволило нам выдвинуть гипотезу об универсальном характере выявленных ограничений.

Подчеркну, что в каждом языке мы включали в базу данных не слова, а все последовательности сочетаний согласных (разделенных гласным), имеющиеся в словаре. Рассмотрим эту на русском французском примерах. Оказалось, что в русском языке, если ориентироваться на словарь А.А.Зализняка, (Зализняк, 1977) самое длинное (графически) слово состоит из 25-ти букв – *частнопредпринимательский*. Из этого слова в анализируемый корпус вошли 8 последовательностей согласных, разделенных гласным, а именно: Ч-С, Н-П, Р-Д, Р-Н, Н-М, М-Т, Т-Л, К-Й. Самое длинное французское слово также состоит из 25-ти букв и 8-ми слогов. Это слово (любопытно, что французы свое самое длинное слово давно открыли, и оно знакомо каждому) – *anticonstituellement* ‘антиконституционно’. Из него в корпус французских сочетаний были включены последовательности Т-К, К-С, Т-Т, L-М (там, где было возможно, учитывалась фонетическая, а не графическая структура слова).

Рассмотрим обобщенные результаты по трем языкам, из 33 языков, которые были проанализированы. На приведенных ниже схемах, знак «-» означает, что зафиксированных в словаре комбинаций существенно меньше, чем ожидалось бы в соответствии с их частотами (двумя минусами обозначаются особенно существенные отрицательные отклонения от ожидаемого результата), знак «+» означает, что соответствующих последовательностей оказалось существенно больше, чем ожидалось (двумя плюсами обозначаются особенно существенные положительные отклонения), отсутствие знака означает приблизительное соответствие реального количества комбинаций вычисленной норме. Символ Р обозначает любой губной согласный, К – велярный, Т – дентальный, С – палатальный. В строках указаны отклонения частот согласных в первой позиции в комбинации CVC, в столбцах – отклонения во второй позиции. В скобках над схемами указано общее количество (N) интересующих нас комбинаций в словаре, использованном в качестве источника.

Язык балант, атлантическая группа (N=904)

		C ₂			
		Р	К	Т	С
C ₁	Р	--	--	++	+
	К		--	+	
	Т	+	++	-	
	С	++	++	-	-

Классический монгольский язык (N=66407)

		C ₂			
		Р	К	Т	С
C ₁	Р	--	-	+	++
	К		--	++	
	Т		++	--	
	С	+			-

Праиндоевропейский язык (N=3085)

		C ₂			
		Р	К	Т	С
C ₁	Р	--		+	
	К		--	+	
	Т	++	++	--	
	С	+	+		-

Чтобы было легче воспринимать схемы, интерпретируем отдельные фрагменты отклонений по языку балант. По данным схемы, в языке балант мы, в частности, видим, что существенно реже, чем следовало бы ожидать, встречаются слова с комбинацией «палатальный согласный + гласный + дентальный согласный» (С-Т со знаком «→»). Напротив, комбинация согласных «лабиальный + дентальный» (Р-Т) встречается «слишком» часто (знак «++») – процент таких слов исключительно высок по отношению к рассчитанной норме.

Приведенные языки существенно различаются по объему представленной лексики. Так, корпус комбинаций в монгольском языке (66 407) был составлен на основе доступного в Интернете монгольского словаря, включающего более 25 000 слов, а корпус по языку балант составлен на основе словника, в котором менее 1000 слов. В праиндоевропейском корпусе вообще собраны не реально засвидетельствованные комбинации, а их реконструкции. Тем не менее, три схемы поразительно похожи. На каждой из них, в частности, прослеживается диагональ с более чем существенными отрицательными отклонениями. Такое распределение интерпретируется однозначно: во всех языках существенно ограничиваются комбинации согласных одного места образования, какой бы гласный ни стоял между ними.

Остановлюсь лишь на самых общих выводах, к которым приводит статистический анализ комбинаций согласных в языках мира. Выясняется, что есть фонетически «плохие» слова, есть «очень плохие» слова, а есть фонетически «хорошие» слова, и при этом консонантная структура «плохих» и «хороших» слов всюду одинакова: и в баскском языке, и в языке кечуа. Во всех обследованных нами языках проступает достаточно сложная, но вполне четкая иерархия запретов и разрешений на сочетаемость согласных в интересующей нас позиции. Так, английские слова *dog* ‘собака’ и *cat* ‘кошка’ – «очень хорошие» (в большинстве языков – «хорошо» сочетать дентальные согласные с велярными), *bug* ‘клоп’ – «нехорошее» слово (в большинстве языков накладываются ограничения и на сочетания лабиальных с велярными), а *toad* ‘жаба’ – «совсем плохое» слово (ограничения на сочетания согласных одного места образования прослеживаются, по нашим данным, во всех языках, и особенно ярко они проявляются в сочетании согласных, различающихся только по одному признаку, например, «глухость» ~ «звонкость» – [t] ~ [d]).

Рассмотрим в в плане общей типологии «запретов» и «разрешений» некоторые аспекты проявления обозначенных универсальных тенденций в русском языке.

Вначале несколько слов методике подсчетов. Она подробно рассматривается в: Поздняков, 1993. Обратимся к конкретному примеру. В словаре А.А.Зализняка, который использовался в качестве источника, 97328 слов (в электронной версии, полученной нами от С.А.Старостина (Зализняк, 1977)). Ограничимся в нашем примере сочетаниями CVC- в начале слова. Их существенно меньше, поскольку в русском языке много сочетаний структуры CCV-, VCV-, VCC- и проч. Тем не менее, в словаре Зализняка интересующих нас сочетаний – 58301. Рассмотрим, сколько мы ожидали бы встретить в словаре сочетаний **фVп-** в начале русских слов. Мы знаем, что начальный **ф-** в русском языке редкий звук. Тем не менее, в словаре 1216 слов с начальным **ф-**. Мы знаем, что **-п-** – частый звук. Действительно, в словаре Зализняка – 3069 слов, в которых в сочетаниях **фVC-** в качестве C представлен **-п-**. Вопрос: сколько мы ожидали бы встретить в словаре слов структуры **фVп-**, если входящие в это сочетание согласные не влияют друг на друга?

Один из наиболее «прозрачных» способов подсчета. Всего комбинаций CVC- – 58301. В них начальных **ф-** – 2,1%, а **-п-** в позиции второго согласного – 5,3%. Это означает, что если появление каждого из двух согласных в интересующей нас комбинации независимо, то в словаре мы ожидали встретить $58301 \times 0,021 \times 0,053 = 65$ примеров с интересующей нас комбинацией **фVп-**. В словаре их 0. Нет в словаре Зализняка ни одного (!) примера указанной комбинации. Можно спорить о том, какую статистическую методику использовать, чтобы наиболее правильно сопоставить фактический 0 с «нормой» 65, но любая осмысленная методика подсчета скажет, что вероятность случайности данного отклонения ничтожно мала. Вот такие отрицательные отклонения от нормы и обозначаются на приводимых здесь схемах двумя минусами.

Пример интересен не только тем, что в данном случае мы сталкиваемся с полным запретом на сочетание **фVп-**. Пример ярко высвечивает и другую проблему: нам известно, что практически все слова с начальным **ф-** в русском языке представляют собой заимствования. Среди них очень много заимствований из французского языка. Почему же французские слова структуры **fVp-** не вошли в русский язык? Потому что русский язык их не принимал? Потому что русский язык их видоизменял? Потому что во французском языке их тоже мало? И вот ответ: нельзя сказать, что во французском языке таких слов мало. Дело в том, что их просто нет. Ни одного! Нет во французском языке слов, которые начинались бы с **fap-, fer-, fir-, for-, foep- fup-**. Нет во французской лингвистике и работ, которые описывали бы этот абсолютный запрет, существующий и в русском языке. А ведь во французском, в отличие от русского, начальный согласный **f-** имеет очень высокую частоту. Вспомним, что праиндоевропейские звонкие придыхательные перешли в латинском языке именно в **f-**: ***bh > f** (ср. фр. *frère* и русск. *брат*), ***dh > f** (ср. фр. *fumée* и русск. *дым*).

Но тогда почему бы словам структуры **фVп-** не перейти в русский из английского? Ведь английских заимствований в русском языке тоже много. Это могло бы произойти, если бы в русский язык было заимствовано единственное (!) английское слово данной структуры – *for* ‘фат, щеголь, хлыщ’ (Мюллер, 1990). Это слово вошло в английский язык в середине 15-ого века и означало изначально ‘глупец’. То есть с самого начала и до настоящего времени у этого единственного английского слова интересующей нас структуры была ярко выраженная модальная окраска.

Итак, в трех «больших» языках (в русском, во французском и в английском) нашлось единственное слово, начинающееся с **fVp-**, хотя, в соответствии с частотами задействованных в этой комбинации согласных, таких слов мы ожидали бы тысячи. Конечно же, нельзя забывать, что все три языка родственные (все-таки, и русский, и

французский, и английский сохраняют по 50% индоевропейского словаря базовой лексики М.Сводеша). Но все же совершенно очевидно, что мы имеем дело не с генетическим феноменом, а с типологическим (см. упомянутую выше трансформацию индоевропейских придыхательных в латинском языке, которая привела к резкому увеличению количества начальных **f-**, и, тем не менее, сочетаний **f-** и **-p-** не появилось).

Здесь есть еще один интересный аспект – восприятие «легитимности» сочетания согласных носителем. Возьмем комбинацию без заимствованных фонем. Носитель русского языка ощущает, что звук [б] в русском языке употребляется часто, по крайней мере, чаще, чем [ф], и звук [п] употребляется часто, по крайней мере, чаще, чем [щ]. Носителю известно, что в русском языке слов очень много. Опираясь на эту информацию (в данном случае речь идет об интуитивно ощущаемых статистических закономерностях), на вопрос «Возможны ли в русском языке слова, в которых есть [б], а за ним после гласного идет [п]?» «статистический» носитель русского языка отвечает: «Почему бы и нет?». Замечу, что мой мини-опрос проводился в среде носителей, профессионально ориентированных на изучение языка, а, говоря проще, среди русскоязычных лингвистов. Когда же просишь привести примеры таких слов, дальше слова *бантист* дело не идет. Самые профессиональные люди вспоминают о *билане* и о *биполярности* (в этих словах очевидна морфемная граница).

Но дело в том, что в словаре А.А.Зализняка, в котором почти 100 000 русских слов и приблизительно 58000, интересующих нас комбинаций, приведенными тремя словами и ограничивается список! А в соответствии с частотами [б] в первой позиции и [п] во второй, мы должны бы были ожидать, что в словаре, в котором 58301 комбинации согласных интересующей нас структуры, указанным условиям будет соответствовать 161 слово! Какую бы статистическую формулу мы ни применяли, ясно, что отрицательное отклонение в данном случае надо описывать сплошными минусами, а говоря строже – вероятность случайности в данном случае практически нулевая.

Совершенно непонятный для меня вопрос – почему же носитель не ощущает этого запрета? И второй вопрос – почему в русской лингвистике, где, казалось бы, описано все (по крайней мере, то, что «лежит на поверхности»), нет, насколько мне известно, ни одной работы, в которой бы этот запрет описывался (как, впрочем, и во французской лингвистике, в которой не только не интерпретирован, но и не описан запрет на сочетания структуры **фVп-** или **бVп-**). А ведь мы увидим дальше, что речь не идет в данном случае о «частном» запрете на сочетаемость [б] и [п]. Речь идет об универсальном (и для русского языка, как и для других языков – системном!) запрете на сочетаемость согласных одного места образования.

Как отмечалось, в словаре русского языка сочетаний структуры **бVп-** практически нет. Рассмотрим обратные сочетания: **пVб-**. В соответствии с позиционными частотами данных согласных в словаре, в котором слов структуры **СVC-** – 58301, мы ожидали бы встретить приблизительно 382 интересующих нас комбинаций. Реально же имеются *публика*, *публицити* и *пуб* (столь же «исконно русские», как *бантист*, *билан* и *биполярность*). Кроме же этих дублетных заимствованных форм, словарь включает только разряд слов с префиксом *по-* и с начальным корневым **б-**: *по-бежать*, *по-быть*, *по-барахтаться* и. т.д. Таким образом, мы имеем основание говорить о практически полном запрете комбинаций **бVп-** и **пVб-**, за исключением контекста, в котором комбинация двух губных согласных разорвана морфемной границей – в этом случае, ограничение на сочетаемость **п** и **б** в русском языке разрешается снять. Тем не менее, несмотря на «разрешение» сочетать префикс *по-* с любым корнем, в том числе и корнем с начальным **б-**, в словаре Зализняка имеется лишь 161 такое слово, в то время как мы бы ожидали (учитывая частоту **п-** в позиции первого согласного и частоту **-б-** в позиции второго согласного) найти в словаре 362 слова, имеющих структуру **пVб-**, то есть интересующих нас слов в словаре оказывается более чем в два раза меньше ожидаемой нормы.

Правило морфемной границы разрешает сочетание префикса *по-* с корнями с начальным **ф**-, например, *по-фартить*, *по-философствовать*, *по-фантазировать*. И тем не менее, слов структуры **пVф**- в словаре почти в 4 раза меньше рассчитанной нормы (26 слов при норме 104). Без префикса *по-* мы практически ничего не находим в словаре, за исключением *пифагорейца*, *пифии*, *пафоса*, *пуфика* и *пиф-паф*. Последнее слово особенно интересно, поскольку, по нашим данным, языки «допускают» «плохие» сочетания не только, когда между согласными проходит морфемная граница, но и в случаях, когда мы сталкиваемся с экспрессивной лексикой.

Еще меньше слов – структуры **бVв**: их 17 вместо 93. Это конечно же *бивак* и *бивуак*, *бювар* и *баварец*, а также редкие односложные глаголы с суффиксом *-ива*: *бивать*, *бывать* (с производными вроде *бывальщина* и *бывало*).

Еще меньше слов структуры **вVп**-, если не считать сочетаний префиксов *во-* и *вы-* с основами на **п**:- их только 8 при норме 100: это *вапориметр* и *вапоризация*, *винпер*, *венс*, идеофон *вопить* с производными и два действительно неправильных русских слова: *вепрь* и *выпь*. Заметим, что оба слова оказываются сегодня на периферии языкового употребления.

Практически полный запрет обнаруживается для сочетания **мVп**:- :в русском языке имеется лишь 2 таких слова при норме 142! Это слова *мопс* и *мопед*. Еще раз подчеркну, что «статистический» носитель, к разряду которых относит себя автор, не воспринимает этого практически абсолютного запрета на сочетаемость **м** и **п**, разделенных гласным. Судя по отсутствию публикаций об этом, не воспринимают этот запрет и русисты. По крайней мере, сомнительно, что его ощущали авторы аббревиатуры **МАПРЯЛ** (Международная ассоциация преподавателей русского языка и литературы), в которой нарушается абсолютный запрет на комбинацию **мVп**- в начале слова. Кстати, если рассматривать аббревиатуру как слово, оно вдвойне неудачное, поскольку включает не только запрещенную для русского языка комбинацию **мVп**- , но и комбинацию плавных согласных **рVл**: 75 слов в словаре из 100000 слов при норме 578! Было бы совсем плохо, если бы в название международного форума русистов была включена комбинация **рVр**. Такая комбинация в буквальном смысле является раритетной: она встречается в русском языке лишь в одном слове *рапитет* при норме 867(!) слов.

Последние примеры, связанные, в частности, с ограничением в сочетании плавных согласных, достаточно хорошо известны. Сочетать **р** и **л** (во всех четырех возможных комбинациях) – достаточно деликатная операция. Между этими фонемами («слишком» близкими, но не тождественными) существует постоянно напряжение: в каких-то случаях проявляется тенденции к их ассимиляции (русск. просторечн. **лаболатория*), в других случаях – к их диссимиляции (в испанском ‘дерево’ **arbor* > *arbol* – нерегулярное диахроническое изменение).

Чтобы дополнить материалы по сочетаниям губных согласных, приведу все слова с комбинацией **мVв**- в начале слова. Перечень не займет много места: в этой позиции встречается только три основы: *мавзолей*, *мавр* (с производными) и *моветон*. Еще раз подчеркну, что речь идет о сочетании согласных, каждый из которых имеет в русском языке очень высокую частоту: в соответствии с частотами [м] и [в] в соответствующих позициях мы ожидали бы встретить в словаре 112 слов структуры **мVв**-!

В словаре А.А.Зализняка 397 слов, начинающихся с комбинации **зас**-, и все эти слова без исключения содержат префикс *за-*. Других префиксов структуры **зV**- в русском языке нет, а поэтому нет и ни одного (!) слова с **зес**-, **зис**-, **зос**-, **зус**, **зыс**-. , хотя без префикса *за-* мы ожидали бы обнаружить 48 слов в словаре Зализняка. И даже, если бы в словарь включались личные имена, ни Зосима, ни Зося не спасли бы положение. Сокращение ЗИС («Завод имени Сталина») в этом смысле значительно «сильнее», чем его предшественник ЗИЛ («Завод имени Ленина»), своей «фонетической неправильностью» воздействуя на

носителей так, как призваны воздействовать идеофоны. Не менее экспрессивно в своей неправильности и сочетание СИЗО.

Абсолютно та же картина наблюдается с велярными согласными. Слов, в которых они комбинируются в указанном контексте, крайне мало. Особенно их мало в тех «ячейках», в которых велярные согласные различаются только одним признаком, например, различаются по глухости–звонкости. Приведу все слова на **гVк-**: это архаичное *гак* и производные (в неспециальной лексике слово *гак* ‘крюк’ задержалось только в клише *с гаком*, приобретя, возможно, экспрессивный компонент), это слова *гекзаметр* и *гектар* (плюс слова того же происхождения), а также два идеофона с производными – *гикать* и *зукать*. В соответствии с фонемными частотами мы бы ожидали встретить не меньше 80 слов обозначенной структуры. Таким, образом, фонетические слова ГВЧП и КГБ оказываются в русском языке практически уникальными, а следовательно имеют особо ярко выраженную экспрессивность (ср. также «неправильные» аббревиатуры КУГИ и ГЭК).

Убедимся в этом, рассмотрев все слова с комбинацией **кVг-** в начале слова. Их несколько больше, чем слов структуры **гVк-**, но и здесь мы имеем дело практически только с заимствованиями, причем, за редкими исключениями (*кагор*) – маргинальными заимствованиями: *кагал*, *кагуляр*, *кагат*, *каганат*, *каганец*, *кигелия*, *когорта*, *куга*, *кугуар*, *кег(е)ль* ~ *кегли* и их производные, *когнитивный*, *когерентность* (в двух последних случаях имеется прозрачная морфологическая граница) и, наконец, это *коготь* (< праслав. ***когъть**) с производными (слово, которое в конечном счете восходит к тому же корню, что и *гак* ‘крюк’). Конечно, в говорах встречаются и другие слова с сочетанием **кVг-**, но они относятся к специфическим слоям лексики. Очень интересно слово *кагуны* (мн.) – «прозвище воронежцев, потому что они говорят **кагуб**, **ягуб** (род. п. ед.ч. от *кто*, *он*)» (Фасмер, 2004). Переход [г] > [в] в окончании родительного падежа, отсутствие которого в ряде говоров и вызвало к жизни появление слов типа *кагуны* и *егуны*, объясняется по-разному. Р. Якобсон объяснял его, например, субморфными изменениями по аналогии в системе окончаний русского склонения: «Фонетическое изменение сочетания /ogó/ в /ovó/ получило широкое распространение только в окончании Р. падежа и было перенесено по аналогии и в формы без конечного ударения. Стимулом к такой экспансии, если не к самому изменению, была, надо думать, возможность обобщения -v- в роли приметы Р. падежа» (Якобсон, 1985, с. 197). Не оспаривая ни одной из многочисленных гипотез, выскажу осторожное предположение: не является ли практически полный запрет на сочетание **кVг-** и фонетическая «неправильность» формы *кого* дополнительным фактором, обусловившим указанное фонетическое изменение во флексии родительного падежа?

Наиболее частотным словом с комбинацией **кVг-** является конечно же местоимение *когда*. Трудно сказать, ощущают ли носители современного русского языка «этимологическую границу» между *ко-* (исторически – местоимение) и *-гда* (< ***gьda** < ***goda**) или же они скорее воспринимают в данном случае оппозицию *к-огда* / *т-огда* в противопоставлении «К-слов» и «Т-слов», как их удачно назвал А.А.Реформатский, говоря о субморфах в русском языке, которые он называл «звуковыми метками»: «Следует отметить еще одно наблюдение Есперсена, это то, что местоимениям присуща пристрастность к "звуковым меткам". <...> то, что, как это отметил Есперсен, в английском языке со звонкого *th* могут начинаться только местоимения (*that*, *they*, *there* и т.д.) – это факт не случайный (проверено по словарю; знаменательные слова в английском со звонкого *th* не начинаются!). В русском языке начальное не "йотированное" э именно прежде всего является в местоимениях (*это*, *этот*, в диалектах: *эвона*, *эва*, *энтот*, *эсто* и т.д.). Такого же рода особенности имеют в русском языке вопросительные местоимения, так называемые *к-слова* и утвердительные их пары – *т-слова*: *куда?* – *туда*, *какой?* – *такой*, *где* (из: *къдь*) – *тут*, *там*, *когда?* – *тогда* и т.п. <...> В “пространственно-

невопросительных” местоимениях характерно для русского языка соотношение *m*-лексем и *s*-лексем, где *m*- означает – →, а *s*- – ← <...>. Ясно, что никакой “естественной” связи понятий “тамошности, тудашности” с *m*- и понятий “здешности, сюдашности” с *s*- нет; связь эта чисто конвенциональная, но она неоспоримо указывает на то, что *m*-слова – это →, а *s*-слова – это ← » [Реформатский, 1979, 73-75]. В этом случае можно следовательно говорить о **субморфной** границе в слове *когда*.

Замечу, что слово *когда* столь же «неправильное», что и слово *туда*, включающее сочетание двух дентальных с одним различительным признаком. Ведь кроме слова *туда*, в очень полном русском словаре Зализняка мы находим только *таджика* и *таджичку*, да и то в слове *таджик* (еще раз подчеркну – единственном в словаре!) фонетически возможно выделять комбинацию **tVдж-**, а не **tVд-**.

На фоне рассмотренных выше комбинаций (практически табуированных в русском языке), комбинация **dVт-** оказывается почти благоприятной: мы находим целый десяток заимствованных корней (*дата* с производными, *датский*, *детерминизм*, *деталь*, *детектив* и *детектор*, *детонатор* – нехорошее слово! – *детрит*, *дотация*, *дот*, *дутар*). Мы находим также сочетания с префиксом *до-*, то есть с морфемной границей между согласными) типа *дотащиться*, *дотла*, *дотуда* (**д-т-д** – дважды неправильное слово, но опять с морфемной границей!), целых три глагола – *дать*, *деть* и *дуть* (заметим, что все они односложные, причем в современном русском языке между согласными проходит морфемная граница, разделяющая основу и окончание, корни *да-*, *де-*, *ду-*) и наконец *дитя* / *дети* с многочисленными производными.

Приведенные примеры в основном касались комбинаций, в которых согласные совпадают по месту образования, но различаются одним признаком по способу образования (глухой / звонкий (**tVд-**), смычный / фрикативный (**пVф-**), назальный / оральный (**мVв-**). В этом контексте согласные всех пяти рядов ведут себя одинаково: дистантные комбинации согласных здесь или не допускаются вовсе, или резко ограничиваются.

Иначе обстоит дело с комбинациями одинаковых согласных: **фVф-**, **tVт-**, **шVш-**, **гVг-** и т.д. Сочетания одинаковых **дентальных** согласных (**дVд-**, **рVр-**, **лVл-**, **нVн-**, **сVс-**, **зVз-**) встречаются существенно реже ожидаемого (только частота сочетаний **tVт-** соответствует рассчитанной норме). Одинаковые **лабиальные** согласные вполне допускаются в сочетаниях (частота сочетаний **пVп-**, **бVб-**, **фVф-** соответствует норме), впрочем, комбинации сонорных (**вVв-**, **мVм-**) резко ограничиваются. Комбинации одинаковых **велярных** в сочетаниях **кVк-** и **гVг-** близки к норме. Комбинации же **цVц-**, **шVш-** и **хVх-** встречаются в начале слова существенно чаще, чем предполагают частотные характеристики каждого из этих согласных в отдельности. При этом, частота комбинаций **жVж-** и **чVч-**, хоть и не превышает норму, но и не является заниженной, в отличие от комбинаций одинаковых дентальных. Такое распределение позволяет сделать вывод, что одинаковые **палатальные** согласные (а также [x]) в указанном контексте – это единственные согласные, которые «с удовольствием» сочетаются друг с другом. В чем причина такого «консонантного магнетизма»? Выскажу осторожную гипотезу. Анализ данных позволяет предположить, что дистантный повтор согласных [ц] [ш] [х] является (по крайней мере, в современном русском языке) формальным маркером экспрессивной лексики.

Рассмотрим такие комбинации подробнее.

В словаре мы ожидали бы найти 12 слов с комбинацией **шVш-**, в то время как реальное их количество составляет 39. Помимо заимствований *шашлык* и *шашка* (холодное оружие), а также слова *шашки* (< *шах*), в котором комбинация **шVш-** задается правилами морфонологических чередований, слова этой структуры имеют ярко выраженный экспрессивный компонент: ср. *шашни*, *шашель*, *шашень*, *шшиига*, *шшиика*, *шшишмор*, *шшиш*, *шушура*, *шушукаться*, *шушпан*, *шушун*, *шуваль* и др. М.Фасмер, наряду с

шушукаться, приводит и другое звукоподражательное слово – *шишикать*. У М.Фасмера можно найти и другие слова (архаичные и диалектные), которые не вошли в словарь А.А.Зализняка. Примечательно, что этимология большинства этих слов неясна. Даже в тех случаях, когда за пределами славянских языков этим словам находятся соблазнительные параллели, этимологу приходится отметить их, поскольку слова с данной фонетической структурой обнаруживают нерегулярные звуковые соответствия. Приведу выборку из словаря Фасмера, подтверждающую высказанное наблюдение: *шашал* / *шашел* ‘моль’ – «неясно», *шешень* ‘рисовые лепешки’ – «неясно», *шишабарник* ‘растение (sp.)’ – «неясно», *шишать* / *шишит* / *шишикать* / *шишилять* ‘копаться, возиться, мешкать’ – «неясно», *шушера* ‘хлам, старье’ / *шушульки* ‘старое тряпье’ – «неясно», *шушукаться* / *шушукаться* ‘мешкать’ – «неясно», *шиши* ‘разбойник, бродяга’, ‘островерхая кладь, копна сена’, ‘шишка’, ‘кукиш’, диал. ‘черт’ – «**этимология затруднительна**», *шашмура* / *шашмура* ‘головной убор замужней женщины’ «**темное слово**», *шушун* – «**по-видимому, иноязычное**», *шашни* / *шашень* ‘осторожный шаг назад’ / *шашал* ‘дрянной человек’ – «**недостаточно сближение с хахаль**», *шишимола* ‘большая шишка’ – «**образование представляется необычным**» (Фасмер, 2004, с. 416-494). Нам остается только радоваться, что не все слова этимологизируются подобным образом.

Нетрудно заметить, что в значении большинства этих слов содержится ярко выраженный модальный компонент (пейоративность): «вредное насекомое», «дрянной человек», «недостойные отношения», «осторожный шаг назад», «разбойник», «черт», «мешкать», «хлам». Иными словами, сочетание **шVш-** маркирует преимущественно идеофоны, то есть экспрессивную лексику. Возможно, именно потому, что язык в целом стремится избегать сочетаний согласных одного места образования, нарушение этого принципа модально окрашивается. А если это так, то в истории некоторых «темных» русских слов, в частности, слов, начинающихся с **шVш-**, следует, возможно, ориентироваться не на закономерные диахронические изменения, формирующие регулярные фонетические соответствия между родственными языками, и не на изменения по аналогии, а на третий тип диахронических (нерегулярных) изменений, характерных для иконических (в частности, экспрессивных) знаков.

Это в полной мере относится и к словам с комбинацией **хVх-**. Об экспрессивности русской фонемы /х/ написано очень много (см., в частности, один из последних обзоров проблемы [х] в исторической перспективе в: Маслова, 2004, с. 196-244). Слова типа *хохотать*, *хихикать*, *хрустеть*, *хранеть*, многократно рассматривались в контексте дискуссии об источниках происхождения [х] в словах, в которых этот звук не выводится из индоевропейского *s. Часто слова типа *соха* рассматриваются как аргумент в пользу реконструкции глухих придыхательных в индоевропейском ([х] выводится в таких словах из *k^h). Сбои в других словах с неясным [х] объясняются экспрессивным характером этого звука в русском языке. Однако, по нашим статистическим данным, повышенной частотностью характеризуется в словаре не сам звук, а именно комбинация **хVх-**.

Сочетание **цVц-**. Согласный [ц] в русском языке относится к редким согласным. Точнее, в корпусе, составившем словарь Зализняка, мы бы ожидали встретить 3 слова с указанной комбинацией. В словаре их 8, причем слова *цаца*, *цацка*, *цацкаться*, *цуцик*, *цыц* безусловно относятся к экспрессивным. Даже заимствование *цеце* содержит экспрессивный компонент. Подчеркну, что речь идет не о звуковом символизме того или иного согласного, в данном случае [ц]. Статистические данные говорят о том, что комбинации согласных, характеризующихся определенными признаками, избегаются в русском языке, как и в большинстве языков мира. И именно там, где существует запрет, возникает возможность маркировать дополнительный семантический признак нарушением этого запрета. Носитель прекрасно ощущает, что [ц] – редкий согласный. Но это не мешает нам ожидать, что в корпусе Зализняка будет, в частности, обнаружено 51 слово структуры **цVс-**, если конечно это сочетание не относится к «запрещенным». Мы же

находим в словаре только 11 слов, а точнее *цистерну, цистит, цистоскоп и цесаря с цесаркой*. Это комбинация, в которой согласные отличаются только одним признаком, и она практически запрещена. При этом, ни в одном из слов не просматривается экспрессивный компонент.

Естественно, фонетические особенности идеофонов в русском языке изучались, и в этой области многое известно. Здесь мне хотелось бы высказать лишь одно методологическое соображение, касающееся исследования экспрессивной лексики. Подсчет вероятностей сочетания фонем на основе их частот и анализ фактических отклонений от этих вероятностей позволяет вскрыть и системно проанализировать субморфные приметы экспрессивных слов. Понятно, что такой анализ связан с определенными сложностями, поскольку в статистическом распределении отражается много разных факторов, действующих в языке одновременно: чисто фонетические запреты на дистантную сочетаемость определенных согласных, фактор морфемной границы, фактор морфонологических чередований, фонетическая маркировка слов с определенной семантикой, частотность отдельных лексических основ и афиксов. Так, например, исключительно высокая частота комбинации **pVз-** в начале слова определяется тем, в русском языке высокую продуктивность имеет префикс *раз-*. Безусловно повышенная частота комбинации **вVщ-** определяется большим количеством производных слов от *вещь*. Большинство подобных сложностей, впрочем, преодолимы. Так, например, последние два фактора не проявляются при подсчете частот сочетаемости фонем в словаре морфем русского языка (Кузнецова, Ефремова, 1986).

Приведу фрагмент распределения сочетаний согласных, полученного по Словарю морфем (напомню, что плюсами отмечаются положительные отклонения от «нормы», минусами – отрицательные. Отсутствие знака обозначает приблизительно точное соответствие ожидаемому числу). В словаре выделяется 3750 комбинаций структуры CVC. Подсчет частот согласных в каждой позиции и фактических отклонений от этих частот дает следующие результаты:

Русский язык (5 локальных рядов)

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

Русский язык (4 локальных ряда)

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

Комментарий: В первой таблице 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**. Во второй таблице объединены данные по рядам S и С.

Представленные в таблице данные показывают системный характер отклонений. В корнях русского языка в комбинациях структуры CVC резко ограничиваются сочетания согласных одного места образования. Ограничиваются и сочетания между препалатальными (**с, з, ц**) и палатальными (**ч, ж, ш, щ, й**). 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**).

Распределение частот консонантных комбинаций в русском языке, по-видимому, не является результатом специфических процессов в исторической фонетике русского языка. В этом можно убедиться, сопоставив таблицу с приведенной выше таблицей по индоевропейскому словарю. Данные, полученные по самым различным языкам мира, поразительно сходны и свидетельствуют скорее об универсальной иерархии «плохих» и «хороших» фонетических слов в синхронии. Сопоставим, например, иерархию признаком в русском и в баскском (Pozdniakov, Segerer, 2007):

<i>Русский язык</i>					<i>Баскский язык</i>				
	P	K	T	C		P	K	T	C
P	--		+		P	--			
K		--	+	-	K	--	--	+	
T	++	++	--	+	T	+	++	-	+
C				--	C	++			--

В данном случае, очевидно, что мы имеем дело не с диахроническим процессом, а с типологической универсалией в синхронии. Вместе с тем, представленная на схеме диагональ постоянно «размывается» за счет разнообразных диахронических факторов, многие из которых рассмотрены выше (разнообразные примеры использования редупликации в языке, формирующей комбинации одинаковых согласных, рассмотренные примеры формирования у фонетически «плохих» слов значения экспрессивности и т.д.). Но коль скоро запрет на сочетание согласных одного места образования всюду проявляется, это значит, что в самых разных языках действуют механизмы «подстройки диагонали». Такие механизмы, а они практически не изучены, представляют несомненный интерес для теории сравнительно-исторического языкознания, и, в частности, для типологии диахронических фонетических изменений. Это отдельная большая тема. Здесь приведу лишь один пример. Суффикс *-ок* в русском языке сочетается с самыми разнообразными лексическими основами, образуя комбинацию **-C-ок**. При этом, сочетаясь с конечными корневыми веллярными, этот суффикс обуславливает консонантное чередование в корне. В результате в русском языке есть слова *боч-ок* (**чVк**), *бож-ок* (**жVк**) и *меш-ок* (**шVк**) и нет слов **бокок* (**кVк**), **богкок* (**гVк**) и **мехкок* (**хVк**). Можно возразить, что чередования основы происходят не только перед суффиксом *-ок*, но почему они происходят здесь только с веллярными согласными, но не с дентальными (ср. *год-ок*, но не **гож-ок*, *рот-ок*, но не **роч-ок*)? Не потому ли (наряду с другими причинами, известными русистам), что чередование позволяет в данном случае избежать комбинации двух веллярных?

Подведем итоги.

Статистические данные однозначно говорят о том, что в русском языке, как и в большинстве других языков мира, резко ограничиваются комбинации согласных одинакового места образования с гласным между ними.

При этом согласные чутко реагируют на фактор «опасной» близости по способу образования. В наибольшей степени ограничиваются комбинации согласных одинакового ряда, различающихся только одним дифференциальным признаком (глухой / звонкий смычный / фрикативный, носовой / ртовый - **сVз, кVх** и т.д.). В парадигматике – оппозиции согласных, составляющих минимальные пары, часто нейтрализуются (мы

различаем в русском или в волоф [б] и [п] в начале слова, но не в конце). В синтагматике в случае непосредственного контакта согласных происходит то же самое (*вос-парить*, но *воз-горание*). В случае же дистантного комбинирования согласных, составляющих минимальные пары, такие комбинации просто избегаются.

Какие бы ни были артикуляционные, чисто физиологические, причины этого универсального запрета, наиболее интересно здесь другое. «Природная» ниша становится точкой повышенной культурной активности, катализатором семиотических процессов: труднопроизносимые звуковые комбинации активно используются языком для формирования новых смыслов. В частности, сочетание одинаковых согласных в некоторых случаях не только не запрещается, но и системно привлекается для выражения категорий модальности и экспрессивности в целом.

Ссылки:

Зализняк, А.А. (1977) Грамматический словарь русского языка. Словоизменение. М.: «Русский язык» (источник).

Кузнецова, А.И., Т.Ф.Ефремова. (1986). Словарь морфем русского языка. М.: «Русский язык» (источник).

Маслова, В.А. (2004). Истоки праславянской фонологии. М.: «Прогресс-Традиция».

Мюллер, В.К. (1990). Англо-русский словарь. М.: «Русский язык». Издание 23-е.

Поздняков К.И. (1993). Сравнительная грамматика атлантических языков. М.: «Наука».

Реформатский, А.А. (1979). *Местоимения* // Очерки по фонологии, морфонологии и морфологии, М.

Фасмер М. (2004). Этимологический словарь русского языка. [1950-1958]. Т.2, 4-ое изд., М.: «Астрель-АСТ».

Якобсон, Р. *Морфологические наблюдения над славянским склонением* // Р.Якобсон. Избранные работы. М.: «Прогресс».

Frisch, S.A., J.Pierrehumbert & M.Broe (2004). *Similarity avoidance and the OCP* // Natural Language and Linguistic Theory 22:179-228.

N'Diaye-Corréard, G. (1970). Etudes fca ou Balanta (dialecte ganja) (Bibliothèque de la SELAF). Paris: SELAF 17 (источник).

Pozdniakov K., G.Segerer (2007). *Similar Place Avoidance: A Statistical Universal* // Linguistic Typology, 11, 2, p. 307-348.

Ссылки на электронные ресурсы:

Баскский язык: <http://weblandarbaso.miarroba.com> (источник)

Зализняк, 1977, база данных: по STARLING Database: <http://starling.rinet.ru> (источник)

Индоевропейский язык: STARLING Database: <http://starling.rinet.ru> (источник)

Монгольский язык: <http://membres.lycos.fr/brunogml/sub/corps.htm> (источник)

Скороговорки: <http://orator.kiev.ua>