

К.И. Поздняков

Париж

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).¹ Наши материалы показывали, что в подавляющем большинстве языков атлантической группы (а в эту группу входят приблизительно сорок языков, которые в генетическом плане отстают друг от друга дальше, чем, например, русский и французский языки Поздняков 1993), 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**, **fap**, **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 фоноло-

¹ Считаю приятным долгом поблагодарить Г. Сежерера за помощь в работе над статьей, а также Г.А. Левинтона за ценные замечания и комментарии.

гии (см., например: Frisch, Pierrehumbert, Broe 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 **lvr** в составе слова допускается, а сочетание **rvl** — не допускается), а также в группе сибилантов — 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 избежать 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 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 посівав 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 у ajo; ajo, caracol у col».

Хорошо известно и ограничение на сочетание близких сибилантов, различающихся лишь одним признаком, например, [s] 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 Саша в

«Шел Саша по шоссе и сосал сушку» и кончая Сашей французским в «*Sachez, mon cher Sasha, que Natasha n'attache pas son chat!*».

Ср. также в русском языке: «Шли сорок мышей, несли сорок грошей, две мыши поплотше 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 pâtitse?*», «*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 слове.

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 восточно-атлантических языков не существует) не очевидна — до сегодняшнего дня никому из исследователей не уда-

лось 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 это на русском и 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) — *anticonstitutionnellement* ‘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 реальному количеству 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)²

		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	+			-

² По материалам (N'Diaye-Corréard 1970).

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 огра-

ничения 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 плане общей типологии «запретов» и «разрешений» некоторые аспекты проявления обозначенных универсальных тенденций в русском языке. В указанную статью (Pozdniakov, Segeger 2007) статистика по русскому языку не вошла. Чтобы описать неизвестный факт в языке балант, можно себе позволить ограничиться одной статистической таблицей, поскольку о языке балант мы не знаем практически ничего. Применительно к русскому языку это, к счастью, невозможно — анализ интересующих нас русских данных может быть проведен на качественно другом уровне и безусловно заслуживает отдельной монографии. Поэтому в этой краткой статье, цель которой — скорее обозначить область перспективных исследований, чем защитить первые результаты, полученные в этой области, ограничимся одним контекстом — рассмотрим последовательности CVC- в начале слова по материалам словаря А.А. Зализняка. Всего в электронной версии словаря Зализняка, полученной нами от С.А. Старостина (Зализняк 1977), оказалось 97328 слов. Эти слова дают нам 291630 последовательностей CVC (техника учета последовательностей была проиллюстрирована выше на примере слова *частнопредпринимательский*). В этой статье нас будут интересовать только 58301 последовательность. Ровно столько последовательностей CVC- оказывается в словаре Зализняка в инициали слов, если вычесть из 97328-и все слова, имеющие иные структуры — CCV-, VCV-, VCC- и проч. Так, например, односложное слово *всплеск* в число интересующих нас слов не войдет, поскольку имеет структуру CCCCVC (инициаль — CCCCVC-).

Вначале несколько слов о методике подсчета. 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 из 58301 слова структуры CVC- имеется 1216 слов с начальным **ф-**. Мы знаем, что **-п-** — частый звук. Действительно, в словаре Зализняка — 3069 слов, в которых в сочетаниях CVC- в качестве второго согласного 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 сло-

варе 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*), ***gh > f**.

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, шеголь, хлыщ’ (Мюллер 1990). Это слово вошло в английский язык в середине XV 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 входящих в эту комбинацию согласных, таких слов мы ожидали бы встретить 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 сочетания **f-** и **-p-** не появились).

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 [ш]. 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 *биplane* 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 комбинация 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 **fVp-** или **bVp-**)? 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 **bVп-** практически нет. Рассмотрим обратные сочетания: **пVб-**. В соответствии с позиционными частотами данных согласных в словаре, в котором слов структуры CVC- — 58301, мы ожидали бы встретить приблизительно 382 интересующих нас комбинаций. Реально же имеются *публика*, *павлисти* и *пав* (столь же «исконно русские», как *бантист*, *биplane* 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 бы (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*, *вепс*, идеофон *вопить* с производными и два действительно неправильных русских слова: *вепрь* и *выть*. 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 *монед*. 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**: в начале слова комбинация **рVл** встречается 75 раз при норме 578, а в целом в языке (из 291630 комбинаций **СVC**) **рVл** встречается 1484 раза при норме 3014, то есть в 2 раза реже, чем она должна была бы встречаться в соответствии с частотами согласных. 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 (по крайней мере, в начале слова) лишь в одном слове *раритет* 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-**!

Рассмотрим комбинации согласных в других локальных рядах. Все без исключения русские слова, начинающиеся с **зас-** (397) включают префикс *за-*. Почему же в русском языке есть слова *задний*, *зайчик*, *зал*, *залп*, *замша*, *заря*, *зябь*, *зять* и нет слов **зас-** без префикса? Почему есть слова *зевать*, *зелень*, *земля*, *зерно*, *зима*, *зыбь*, *зола*, *зуб*, *зулус* и многие другие — исконные или заимствованные, и нет ни одного слова с **зес-**, **зис-**, **зос-**, **зус-**, **зыс-**? 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 преемственник ЗИЛ («Завод имени Лихачева»), своей «фонетической неправильностью» воздействуя на носителей так, как призваны воздействовать идеофоны. Не менее экспрессивно в своей неправильности и сочетание СИЗО.³

³ Для В. Набокова сходное сочетание в английском языке — «тернистое»: «Столь же упорно он английским „тоже“ орудовал, как немецким „итак“, и, одолевая тернистое окончание в слове, означавшем „одежды“, неизменно добавлял лишний свистящий слог, как если б человек поскользнулся после взятия препятствия» (В. Набоков. «Дар»).

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 *гукать*. В соответствии с фонемными частотами мы бы ожидали встретить не меньше 80 слов обозначенной структуры. Таким образом, фонетические слова ГКЧП 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. **kogъть*) 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 *кау́б*, *яу́б* (род. п. ед. ч. от *кто*, *он*)» (Фасмер 2004). Переход [г] > [в] в окончании родительного падежа (отсутствие этого перехода 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 /огó/ в /овó/ 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: 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* (< **gda* < **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, энто, 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 *т-лексем* и *с-лексем*, где *т-* означает — →, а *с-* — ← <...>. Ясно, что никакой „естественной“ связи понятий „тамошности, тудашности“ с *т-* и понятий „здешности, сюдашности“ с *с-* нет; связь эта чисто конвенциональная, но она неоспоримо указывает на то, что *т-* слова — это →, а *с-* 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 **тVдж-**, а не **тVд-**.

На фоне рассмотренных выше комбинаций (практически табуированных в русском языке — по крайней мере, в начале слова), комбинация **дVт-** 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 (**тVд-**), 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-**, **тVт-**, **шVш-**, **гVг-** и т. д. Сочетания одинаковых **дентальных** согласных (**дVд-**, **рVр-**, **лVл-** **нVн-**, **сVс-**, **зVз-**) встречаются существенно реже ожидаемого (только частота сочетаний **тVт-** 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] [x] 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*, *шишига*, *шишка*, *шишимора*, *шиш*, *шушура*, *шушукаться*, *шушпан*, *шушун*, *шушваль* и др. М. Фасмер, наряду с *шушукаться*, приводит и другое звукоподражательное слово — *шишикать*. У 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 отмечать 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* / *шишить* / *шишкать* / *шишлять* ‘копаться, возиться, мешкать’ — «**неясно**», *шушера* ‘хлам, старье’ / *шушульки* ‘старое тряпье’ — «**неясно**», *шушулькаться* / *шушкаться* ‘мешкать’ — «**неясно**», *шиш* ‘разбойник, бродяга’, ‘островерхая кладь, копна сена’, ‘шишка’, ‘кукиш’, диал. ‘черт’ — «**ЭТИМОЛОГИЯ ЗАТРУДНИТЕЛЬНА**», *шашмура* / *шашмура* ‘головной убор замужней женщины’ — «**темное слово**», *шунун* — «**по-видимому, иноязычное**», *шашни* / *шашень* ‘осторожный шаг назад’ / *шашал* ‘дрянной человек’ — «**недостоверно сближение с хахаль**», *шишимола* ‘большая шишка’ — «**образование представляется необычным**» (Фасмер 2004, 4: 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 **xVx-**. Об экспрессивности русской фонемы /x/ написано очень много (см., в частности, один из последних обзоров проблемы [x] в исторической перспективе: Маслова 2004: 196–244). Слова типа *хохотать*, *хихикать*, *хрустеть*, *хранеть*, многократно рассматривались в контексте дискуссии об источниках происхождения [x] 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 ([x] выводится в таких словах из *k^h). Сбои в других словах с неясным [x] объясняются экспрессивным характером этого звука в русском языке. Однако, по нашим статистическим данным, повышенной частотностью характеризуется в словаре не сам звук, а именно комбинация **xVx-**.

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 комбина-

цией. 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 проявляется 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 **рVз-** в начале слова определяется тем, что 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, в котором 2936 корней, что приблизительно соответствует, по мнению С.А. Старостина (устное сообщение), числу корней в любом языке. Здесь нет префикса *под-*, определяющего повышенную частоту сочетания **п** и **д** в начале слова, здесь нет суффикса *-тель*, распространенность которого

определяет высокую частоту сочетания **т** и **л/ль** в конце слова, здесь нейтрализуется высокочастотное сочетание **д** и **л**, которое обнаруживает резкое положительное отклонение от нормы благодаря огромному числу слов, производных от корня **дел**. Это означает, что одинаковый интерес представляют как комбинации в начале слова, так и все не-начальные комбинации. Ниже приводятся обобщенные данные по «запретам» и «разрешениям» сочетаний согласных в структуре CVC корневых морфем русского языка во всех позициях. Напомню, что плюсами отмечаются положительные отклонения от «нормы», минусами — отрицательные; отсутствие знака обозначает приблизительно точное соответствие ожидаемому числу. 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 локальным 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 **С** 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 веларный, 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 веларный К (к, г, х): **ТVP, TVK**. 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 признаков в русском и в баскском языках (статистика по баскскому из: Pozdniakov, Segerer 2007):

Русский язык (N=58301)

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

Баскский язык (N=3140)

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

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 проявляется всюду, 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 **-С-ок**. При этом, сочетаясь с конечными корневыми велярными, 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 велярными 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 велярных?

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. Здесь рассматривались в основном примеры сочетаний согласных в русском языке в начале слова, но наши статистические данные говорят о том, что, по крайней мере в русском языке, позиция согласных в слове практически не имеет значения.

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 ком-

бинирования согласных, составляющих минимальные пары, такие комбинации просто избегаются.

Любой профессиональный фонетист, если его убедят статистические данные, легко найдет причину (и наверняка не одну), которая могла бы лежать в основе ограничений в сочетании согласных одного ряда, а, тем более, сочетаний самых близких согласных одного ряда, различающихся лишь одним или двумя признаками, например, **з** и **с** или **в** и **п**. Но вернемся к примерам с морфемными границами. Если слова *за-столье* и *вы-пить* оказываются «хорошими» словами, а *ЗИС* и *выть* — плохими, это значит, что в иерархии фонетических и знаковых признаков физиология занимает подчиненное положение, а знаковость — главное. Последовательности **зV-** и **вV-** могут сочетаться с любыми согласными (**с**, **ш**, **в**, **ф** и проч.) в том случае, если они что-то значат, то есть являются морфемами. Если же знак отсутствует, в силе остаются артикуляционные ограничения. Но именно там, где позиции не-знакового мира кажутся вполне прочными, этот мир оказывается особенно незащищен перед экспансией культуры. «Природная ниша» становится точкой повышенной «культурной активности», катализатором семиотических процессов: звуковые комбинации, воспринимаемые как труднопроизносимые, активно используются языком для формирования новых смыслов. Так, в некоторых случаях сочетания одинаковых согласных не только не запрещаются, но и регулярно «поставляют материал» для формирования иконических знаков с экспрессивным и, в частности, с модальным значением. А если это так, использование статистики, настроенной на универсальные тенденции звуковых «запретов» и «разрешений», открывает, в частности, перспективу системного исследования иконических слов.

Ссылки

- Зализняк 1977 — А.А. Зализняк. Грамматический словарь русского языка. Словоизменение. М., 1977.
- Кузнецова, Ефремова 1986 — А.И. Кузнецова, Т.Ф. Ефремова. Словарь морфем русского языка. М., 1986.
- Маслова 2004 — В.А. Маслова. Истоки праславянской фонологии. М., 2004.
- Мюллер 1990 — В.К. Мюллер. Англо-русский словарь, 23-е изд. М., 1990.
- Поздняков 1993 — К.И. Поздняков. Сравнительная грамматика атлантических языков. М., 1993.
- Реформатский 1979 — А.А. Реформатский. Местоимения // Очерки по фонологии, морфонологии и морфологии. М., 1979, с. 64–93.
- Фасмер 2004 — М. Фасмер. Этимологический словарь русского языка: В 4 т. [1950–1958], 4-е изд. М., 2004.
- Яacobсон 1985 — Р. Яacobсон. Морфологические наблюдения над славянским склонением // Р. Яacobсон. Избранные работы. М., 1985, с. 176–197.

- Frisch, Pierrehumbert, Broe 2004 — *S.A. Frisch, J. Pierrehumbert, M. Broe. Similarity avoidance and the OCP // Natural Language and Linguistic Theory, vol. 22, 2004, p. 179–228.*
- N'Diaye-Corréard 1970 — *G. N'Diaye-Corréard. Etudes fca ou Balanta (dialecte ganja) (Bibliothèque de la SELAF). Paris, 1970.*
- Pozdniakov, Segerer 2007 — *K. Pozdniakov, G. Segerer. Similar Place Avoidance: A Statistical Universal // Linguistic Typology, vol. 11, No. 2, 2007 (в печати).*

Ссылки на источники в Интернете

Баскский язык:

<http://weblandarbaso.miarroba.com>

Зализняк 1977 — база данных: по STARLING Database:

<http://starling.rinet.ru>

Индоевропейский язык — STARLING Database:

<http://starling.rinet.ru>

Монгольский язык:

<http://membres.lycos.fr/brunogml/sub/corps.htm>

Скороговорки:

<http://orator.kiev.ua>