

ЮРИЙ А. ТАМБОВЦЕВ, АЛИНА Ю. ТАМБОВЦЕВА,
ЛЮДМИЛА А. ТАМБОВЦЕВА

**Частота встречаемости смычных согласных
в некоторых языках мира как показатель степени
смычности их звуковой цепочки**

Abstract

It is still an enigma why some languages produce many occlusive consonants in their speech sound chains while others do not. Thus, some world languages may have many occlusive consonants in their sound speech chains, e.g. Nenets of the Samoyedic taxon – 22.67%. By occlusive consonants we mean the speech sounds which are produced by some sort of constriction in the human sound tract. The complete closure produces stop consonants. The occurrence of occlusive consonants were considered in different language subgroups, groups and families, that is language taxa. The taxon of the Samodian (Samoyedic) languages has a very high degree of the occurrence of stop consonants among the world languages under our study. The mean of occlusive consonants in Samoedic taxon is 26.74% while in the taxon of the Australian aboriginals the mean is only 19.49%. The least concentration is in the sound chain of the Samoan languages of the Austronesian language taxon. It is nearly twice less than the mean of occlusives in the other world languages. The occlusive data in every language taxon are compared to the other language taxa (groups and families) by the mean values.

Keywords: linguistics, occlusive consonants, language taxon

Введение

Каждый человеческий язык имеет свое материальное воплощение в виде звуков речи. В процессе речевой коммуникации эти звуки образуют цепочки.

Таким образом, способ существования любого языка человека – это звуки, которые производятся речевым аппаратом человека. Древнегреческий философ Демокрит утверждал, что в мире ни одна вещь не происходит беспричинно, т.е. все имеет свою закономерность (Асмус, 1965: 98). Из этого следует, что звуки речи употребляются в потоке речи на каких-то языках с какой-то своей закономерностью. Эти закономерности в языках мира необходимо изучать, чтобы выявлять как общие, так и индивидуальные тенденции. Эти тенденции зависят от артикуляционной базы языка. В каждом этносе артикуляционные базы могут быть различны. На различие артикуляционных баз могут указывать особенности некоторых антропологических характеристик (Тамбовцев, 2011).

Звуки речи конечны в своем наборе, т.е. номенклатуре. Как же тогда при их помощи передаются бесконечные потоки информации? Дело в том, что они имеют способность повторяться в соответствии с системой правил каждого языка. Эта система правил индивидуальна для каждого языка, хотя в родственных языках номенклатуры звуков речи похожи. Именно частота повторения этих фонетических элементов выстраивают определенную фонетическую картину определенного языка. Эти повторяющиеся фонетические элементы могут классифицироваться по их артикуляторным характеристикам, таким образом создавая те или иные группы. Мозаика этих групп может быть совершенно отличать один язык от другого, но может быть и в какой-то степени похожа (Гамкрелидзе, 1986).

Целью нашего исследования является нахождение устойчивых отношений сходства между звуковыми цепочками языка на основе выделенных нами признаков, имеющих определенную частоту встречаемости. Мы анализируем каждый язык как материальный объект, который имеет набор определенных признаков, имеющих количественные характеристики. Сопоставление языков по их количественным значениям повышает степень точности и объективности классификации языков по различным языковым таксонам (Brainerd, 1974: 248; Wray et. al., 1998: 255–264).

Нельзя не отметить, что речевой аппарат человека у разных народов мира с анатомической точки зрения одинаков. Это прослеживается на многих антропологических характеристиках (Тамбовцев, 2012). В то же время результаты работы артикуляторных органов различны. Мы изучаем разброс этих результатов по разным языкам, объединенным в различные подгруппы, группы, семьи и другие таксоны. Анализируются интервалы, в пределах которых могут распределяться как отдельные признаки (смычность, фрикативность, губность, гортанность и т.д.), так и их совокупности, например, эфония, т.е. благозвучие. Под эфонией языка мы понимаем сумму сонорных согласных и гласных в звуковой цепочке.

Некоторые языки звучат благозвучно, т.е. мелодично, вследствие большой концентрации сонорных согласных, а другие – прерывисто, как следствие большого количества смычных согласных в их звуковой картине. Мы уже проанализировали языки мира с точки зрения величины их эфонии, т.е. благозвучия. Под этим

признаком мы понимали совокупность частоты встречаемости гласных и сонорных согласных в звуковой цепочке языка [Тамбовцев, 2009].

В данной работе нашей задачей является анализ степени смычности, т.е. прерывистости, в звуковой цепочке языков мира. Смычные согласные являются хорошей маркирующей характеристикой звуковой цепочки потому, что они хорошо слышны в потоке речи. Фактически они прерывают звуковой поток речи (Halle et al., 1957). В нашей интерпретации, смычными согласными будут быть как губные (типа [p, b], переднеязычные типа [t, d,], так и велярные и гуттуральные согласные типа [k, g и гортанный смык]. Эти типы согласных объединены в одну группу потому, что не во всех языках представлены все указанные типы. В одном языке есть смычные заднеязычные, увулярные и ларингальные, – в другом смычные заднеязычные и фарингальные, а в третьем – только смычные заднеязычные. Таким образом, эти языки невозможно сравнивать потому, что они несоизмеримы. Необходимость соблюдения принципа соизмеримости заставляет нас объединить смычные заднеязычные, увулярные, фарингальные и ларингальные согласные в одну группу – группу смычных гуттуральных.

Фактически смычные гуттуральные согласные взяты во всей своей совокупности как одна группа с тем, чтобы противопоставить им каждую другую группу: группу фонем, например, частоту встречаемости губных, переднеязычных и гуттуральных согласных, которые артикулируются в передней и задней областях ротовой полости близ твёрдых пассивных органов зубов и нёба. Губные согласные производятся при помощи губ, а переднеязычные – кончика и передней части языка. Гуттуральные смычные согласные образуются в задней части речевого аппарата человека. Мы не анализируем смычные среднеязычные согласные типа [j]. Они обычно являются щелевыми (fricative), но никак не смычными (Широков 1985: 45).

Довольно интересны взгляды известного фонетиста О. С. Широкова на некоторые дополнительные противопоставления фонематичных групп в звуковой цепочке некоторых стилей языка. Так, кроме указанных выше фонемных групп, он предлагает создать новые фонемные группы, куда бы вошли такие согласные, которые артикулируются в маргинальных (т.е. крайних, пограничных) областях. Так можно выделить крайние передние, то есть те согласные, которые артикулируются в передней маргинальной позиции. Этими согласными являются губные. Они противопоставляются крайним задним, то есть тем, которые артикулируются в районе мягкого нёба и далее. При артикуляции заднеязычных (велярных) выгибается задняя часть спинки языка, приближаясь к твёрдому нёбу. Впереди фокуса оказывается большое резонирующее пространство ротовой полости, которое и придаёт заднеязычным низко тональную окраску (Широков 1985: 45). Обсуждая более детально производство смычных заднеязычных согласных как таковых (то есть в классическом их понимании без увулярных, фарингальных и ларингальных), Л.Р. Зиндер различает два их вида: а) одни образуются самой задней частью языка, касающейся мягкого нёба (и которые, следовательно, трудно отграничить от фарингальных, увулярных и ларингальных); другие задней

частью языка, но в которых присутствует смычка. Реже смычка наблюдается тех смычных, которые граничат со средней частью языка и соответственно касаются более передней части нёба (Зиндер, 1979: 162). Заметим, что этот вид заднеязычных согласных намного реже встречается в языках мира. В то же время, следует заметить, что их трудно отграничить от среднеязычных согласных. Это, например, в тюркских, финно-угорских и тунгусо-маньчжурских языках часто приводит к путанице, так как в одном языке один и тот же согласный считается среднеязычным, а в другом он же считается заднеязычным, продвинутым вперёд. Заднеязычные, продвинутые вперёд, характеризуются более высоким тембром, который на слух воспринимается как некоторая «мягкость», что особенно характерно для тюркских, славянских, и уральских языков, где они обычно идут перед гласными переднего ряда. В тюркских языках они называются гласными «мягкого ряда». Тем не менее, эти согласные, на наш взгляд, все равно остаются заднеязычными смычными. Это же характерно и для некоторых славянских языков. Так, например, в русском языке такие смычные гуттуральные как [g, k] перед гласными переднего ряда сильно продвигаются вперёд. Тем не менее, на наш взгляд, они не переходят в группу среднеязычных согласных, поэтому при подсчетах их нужно учитывать в группе смычных гуттуральных согласных.

Заметим, что наличие смычки хорошо показывает палатография, но она проводилась только по небольшому количеству языков мира (Зиндер 1979: 163). К сожалению, следует заметить, что фонетическая классификация большинства языков сделана на слух, т.е. без привлечения палатографии.

Обсуждая некоторую специфическую акустическую окраску некоторых гуттуральных смычных согласных, можно отметить, что она воспринимается как «гортанный смык». Л.Р. Зиндер отмечает, что гортанное произношение часто наблюдаются в кавказских, палеоазиатских и арабских языках (Зиндер, 1979: 164). Это подтверждается нашими конкретными цифровыми данными (Тамбовцев, 2008; 1985) некоторых других наших работах дифференциация языков по степени маргинальности, нёбности, губности, переднеязычности, сонорности, эвфонии и т.д. рассмотрены подробнее (Тамбовцев, 1983; 1985; 2008).

По мнению многих фонетистов, артикуляция глоточных (фарингальных) и ларингальных смычных принадлежит к числу наиболее неясных, хотя они имеют довольно широкое распространение, особенно в восточных языках.

В отношении смычного гортанного существует ещё больше разногласий, вплоть до того, куда его относить, к оттенкам гласных или к согласным. Во многих языках до сих пор не установлено, является ли этот звук фонемой (например, в кетском, чукотском, языках американских индейцев и т.д.) или же он фонематически связан с гласными как их особый колорит. Хорошо известно явление «кнак-ляут» в немецком языке, где этот звук всегда чётко относят к артикуляции гласных. Это так называемый «сильный приступ» гласного, который фактически является гортанным смыком. Он всегда находится в начале немецкого слова, которое начинается с гласного. Имея все характеристики согласного, кнак-ляут, тем не

менее, не считается согласным. Примерно такое же фонетическое явление можно найти и в датском языке. Там тоже гортанный смык фонематически не отделим от гласного. Считается, что в состав гласного входит гортанная смычка в удэгейском, корякском, кавказских и языках американских индейцев. Может быть поэтому, Л.В. Щерба в своей таблице рядом со значком гортанного смыка поместил значок, который обозначает голос как таковой, а не тип согласного. Л.Р. Зиндер ставит под сомнение уместность такого знака в таблице типов гласных. Мы думаем, что подобного рода значок может обозначать какой-то тип согласного и его нужно помещать в таблицу согласных, но не в таблицу гласных. Такой разницей говорит о неразработанности данной проблемы.

Таблица Л.В. Щербы показывает, что теоретически возможных заднеязычных, увулярных, фарингальных и гортанных может быть 45. Однако, фактически чаще всего встречаются заднеязычные шумные смычные чистые типа [k, g,]; Увулярные согласные в различных взятых нами языках представлены чистыми шумными смычными типа [q]. Гортанные смычные согласные представлены только чистыми шумными типа [ʔ]. Часто в орфографии гортанный смык обозначается знаком апострофа [']. Таким образом, во взятых нами языках употребляются далеко не все из теоретически возможных смычных гуттуральных согласных описанных в таблице Л.В. Щербы (Тамбовцев, 2001-а; 2001-б; 2001-в).

Интересно отметить, что в языках мира смычные согласные функционируют в определенном интервале частоты встречаемости, не выходя за эти пределы. В данной работе мы анализируем эти пределы (Таб. 1–27).

Материалом работы послужили тексты в фонематической транскрипции. Более детальное описание материала даны в наших монографиях (Тамбовцев 1994-а; Тамбовцев 1994-б; Тамбовцев 2001-а; Тамбовцев 2001-б; Тамбовцев, 2001-в).

Применяемые методы статистических расчетов

В данной работе применяется несколько методов статистических расчетов, которые основаны на некоторых критериях математической статистики. Некоторые из этих методов мы несколько модифицировали для того, чтобы сделать более понятными процедуры расчета для получения лингвистически значимого результата. Другие критерии математической статистики были взяты в классическом виде. Но и в том, и в другом случае мы старались описать их настолько просто, насколько это было возможно без потери правильности и точности.

Среднее квадратическое отклонение

Язык подвержен изменениям во всех его проявлениях, т.е. языку свойственна вариативность. Велика степень вариативности и на фонетическом уровне.

В лингвистических работах значение средних употребляется довольно часто, но не отмечается, как эти средние варьируют. Это, на наш взгляд, является значительным упущением. Между тем без учета степени варьирования встречаемости отдельных фонем или групп фонем нельзя составить полную картину функционирования этих лингвистических единиц в звуковой цепочке на каком-либо языке. Отсюда следует, что наряду с использованием средних величин нужны еще и показатели их изменчивости, т.е. степень их вариативности. Иногда степень изменчивости называют не степенью вариативности, а степенью вариабельности или вариации. Необходимо применить метод, который бы учитывал равномерность употребления какой-либо фонемы (или группы фонем) в тексте, что является крайне важным в лингвистических исследованиях. Одним из самых удобных методов определения рассеяния (разброса) является расчет среднего квадратического отклонения (СКО), величина которого обычно обозначается латинской буквой *S*. Вычисление среднего квадратического отклонения (СКО) важно еще и потому, что он является базой для расчета многих критериев математической статистики, которые используются в математической лингвистике, например, коэффициента вариации (*V*), доверительного интервала (*CI*), критерия Фишера (*F*) и критерия Стьюдента (*t*), о чем подробнее сказано в нашей монографии. В то же время, в этой работе мы достаточно подробно рассмотрим коэффициент вариации далее. Обычно его обозначают символом “*V*” В то же время заметим, что СКО (*S*) и коэффициент вариации (*V*) являются самыми простыми показателями меры разброса значений вокруг их среднего.

Перед расчетом среднего квадратического отклонения (*S*), необходимо решить проблему разбиения фонологической цепочки на порции (выборки). В принципе текст можно разбить на порции любой величины. Важно, чтобы объемы этих порций были равны во всех языках. В фонологических исследованиях часто величина одной порции равна 1000 фонем. Мы тоже придерживались этой величины. В принципе, можно брать порции большего или меньшего размера. В то же время, нельзя не учитывать, что, если размеры порции очень маленькие, то значения частоты встречаемости будут слишком сильно колебаться. Это может искусственно завязать степень вариативности. Слишком большие порции могут быть неподходящи потому, что их будет слишком мало, если текст короткий, а длинного текста в распоряжении исследователя часто просто нет. В математической статистике количество порций часто подчеркивают, что количество порций не должно быть слишком маленьким. Решив проблему величины объема порции, необходимо определить, сколько таких порций необходимо и достаточно для исследования. В принципе, чем больше количество порций, тем более надежные результаты должны получиться. При этом, объем каждой порции должен быть достаточно большим. Это приведет к тому, что объем общей выборки тоже будет достаточно большим. Минимальное число таких порций не должно быть меньше 30, как рекомендует математическая статистика (Вентцель, 1964: 319; Тамбовцев, 2003). В наших исследованиях мы часто брали

число выборок $30 + 1$, т.е. 31 тысячу. Во многих наших выборках объем был больше этого, т.е. величины выборок были (где это было возможно) по 31 000 и более фонем.

Весь имеющийся материал по какому-либо языку будет считаться генеральной совокупностью звуковых последовательностей (цепочек). Эту генеральную совокупность нужно разбить на выборки по 1000 фонем, при этом количество этих выборок, по возможности, не должно быть меньше 31 тысячи. Нами взята уточненная формула расчета СКО (S), в которой от общей суммы выборок отнимается единица ($n - 1$), т.е. $31 - 1 = 30$. Как мы помним, 30 порций – это минимальный объем выборки, который обеспечивает необходимую достаточную статистическую надежность. Выборки менее 30 порций считаются малыми выборками, со всеми вытекающими отсюда последствиями. Довольно часто, статистические критерии на малых выборках показывают иные результаты, чем на выборках большого и среднего объема. Лингвисту не нужно вдаваться в математические подробности, почему это именно так. Чтобы избавиться от проблем особых приемов обработки малых выборок, проще брать выборки более 30 порций.

Следует помнить, что иногда среднее квадратическое отклонение называют стандартным отклонением или просто стандартом. Подробный расчет СКО (S) рассмотрен нами ранее (Тамбовцев, 1994-а; 1994-б: 36–37), поэтому здесь рассмотрим его кратко. В большинстве работ по математической лингвистике рекомендуется следующая формула:

$$S = \sqrt{\sum (X_i - X)^2 / n - 1}$$

где X_i – частота встречаемости фонемы (или группы фонем) в каждой порции;

X – средняя частота встречаемости фонемы во всей сумме порций;

n – количество порций в выборке из генеральной совокупности.

Обратим еще раз внимание на знаменатель в данной формуле, где стоит число ($n - 1$). Это несколько повышает точность вычислений, особенно при небольшом количестве порций и в связи с этим рекомендуется к применению в математической лингвистике и фоностаистике (Головин, 1971: 25; Носенко, 1981: 40; Пиотровский и др., 1977: 245, 272; Тамбовцев, 1994-б: 36). В принципе, на малых выборках можно применять только ($n - 1$). В то же время, на выборках большого объема можно применять упрощенную формулу, где в знаменателе используется просто число выборок (n). Величина среднего квадратического отклонения (S) может служить в качестве показателя степени вариации частоты встречаемости фонем (или фонемных) групп в каких-либо двух текстах, только при условии, если их объемы одинаковы. В противном случае для измерения изменчивости этих текстов нужно применять другие методы. Такими методами могут быть, например, коэффициент вариации, доверительный интервал, и модифицированный критерий “хи-квадрат” в виде коэффициента T . Они

позволяют сравнивать вариативность текстов разного объема. Мы рекомендуем брать тексты по возможности одинакового объема, так как вычисления при этом значительно упрощаются, а вероятность ошибки понижается. При этом нельзя забывать, что количество этих порций должно быть достаточно велико при применении любых статистических методов, если мы хотим получить статистически устойчивые результаты. Таким образом, нужно по возможности стремиться к тому, чтобы при применении статистических критериев выборки были примерно одного объема.

На наш взгляд, необходимо привести пример расчета величины среднего квадратического отклонения (СКО). В качестве примера могут послужить тексты на древнеанглийском языке ("Beowulf, Ohthere's and Wulfstan's Story; The Description of Britain; Julius Caesar, etc.). В этих текстах случайным образом выбирались фонемные цепочки по 500 фонем. Далее случайным образом они складывались в порцию по 1 000 фонем, затем случайным образом отбиралась каждая тысяча древнеанглийских фонем. Следовательно, это был уже не связный текст, а квази-текст. Может быть, именно поэтому, как мы увидим далее, у нас получились довольно большие значения СКО. Приведем значения частоты встречаемости губных согласных в выборке объемом 31 000 фонем, по мере того, как они встречались:

124, 111, 103, 96, 100, 93, 112, 96, 106, 78, 117, 122, 125, 94, 111, 96,
140, 95, 100, 123, 110, 116, 106, 92, 119, 120, 101, 135, 103, 123, 128.

Прежде всего, требуется вычислить среднее, для чего нужно сумму данного ряда разделить на количество его членов, т.е. 31. Здесь среднее (M) равно 109,52. Далее от каждого значения нужно отнять это среднее, возвести в квадрат и сложить: например, $(124 - 109,52)^2 + (111 - 109,52)^2 + (103 - 109,52)^2 + \dots$ и т.д. Полученную сумму нужно разделить на 30, так как в знаменателе стоит число на единицу меньше всего объема выборки. В результате имеем квадрат СКО, который здесь равен 204,46. Далее из этого числа нужно извлечь квадратный корень. Получаем 14,30. Много это или мало познается в сравнении с другими порциями древнеанглийского текста. Так СКО второй порции равно 37,06; третьей – 14,16; четвертой – 15,42.

Попробуем сравнить данные по сплошному тексту. В качестве примера возьмем текст библии английского короля Джеймса. Интересно, что на разных частях этой библии получены разные величины СКО. Третья книга Старого Завета «Левит» дает значение СКО частоты встречаемости губных согласных равное 11,92. Книга «Нового Завета» от Матфея дает СКО по губным равное 8,57; от Марка – 8,59; от Луки – 10,09; от Иоана – 12,30. Как мы видим, на материале книг Ветхого и Нового Завета колебания значений СКО значительно меньше, чем на материале древнеанглийских текстов. Вероятно, так происходит потому, что в последнем случае целостность текста (т.е. связность) не нарушена. Мы видим, что СКО реагирует на то, является ли текст семантически целостным, или же разорванным на кусочки. На таком же объеме драмы Б. Шоу СКО = 9,99.

В прозе Дж. Г. Байрона $СКО = 10,42$. В поэзии Байрона она ещё выше (14,57). Величина $СКО$ на материале газеты повышается ещё больше (14,79). Ещё выше она в диалогах на бытовые темы (15,93).

Причины повышения или понижения значения $СКО$ будут обсуждаться отдельно. В данном случае нам было важно понять, реагирует ли $СКО$ на изменение стиля и в какой степени. Приведённые выше примеры значений $СКО$ в разных стилях говорит о том, что $СКО$ реагирует на изменение стиля, т.е. структуру материала исследования.

Следовательно, $СКО$ является достаточно тонким инструментом для исследования звуковой цепочки. Из этого можно сделать общий вывод, что все критерии, базирующиеся на $СКО$, должны реагировать на изменение стиля. Это могут быть, прежде всего, коэффициент вариации (V) и критерий Стьюдента (t). Как мы увидим далее, наши результаты подтвердят наше предположение.

Коэффициент вариации

В лингвистических исследованиях, также как и в других гуманитарных и точных науках, часто случается так, что исследователю необходимо сравнивать между собой параметры разных величин. Нужно соблюдать принцип соизмеримости. В качестве отвлеченной меры изменчивости в этом случае рекомендуется применять коэффициент вариации, который другими словами можно назвать коэффициентом изменчивости (Зайцев, 1990, 37–39; Носенко, 1981: 49; Пиотровский и др., 1977: 180; Сиськов, 1975: 32; Хьетсо и др., 1989: 116; и др.).

Как обсуждалось выше, среднее квадратическое отклонение характеризует степень отклонения частоты встречаемости фонемных групп от среднего арифметического значения в абсолютных числах. При этом, $СКО$ не учитывает, что, например, в группу губных согласных входит больше фонем, чем в группу среднеязычных, поэтому величина среднего по губным намного больше в абсолютных единицах, чем величина среднего по среднеязычным. В то же время частота встречаемости переднеязычных согласных, в свою очередь, намного выше, частоты встречаемости губных и т.д. Строго говоря, чтобы сравнивать их между собою, необходимо их каким-то образом привести в соизмеримый вид, так как статистически сравнивать можно только соизмеримые величины. В своих работах мы придерживались этого правила о сопоставимости величин (Тамбовцев, 1992: 27–43; 1993-а: 344; 1993-б: 348; Тамбовцев и др., 1980: 323–324). Против прямого сравнения изменчивости, т.е. вариаций, которые распределены по-разному и принадлежат разным лингвистическим признакам, предостерегают опытные специалисты в области математической лингвистики (Пиотровский и др., 1977: 242–243). На наш взгляд, лингвисту не всегда обязательно проверять форму распределения признака, если величина выборки больше 30, так как по

закону больших чисел распределение должно стремиться к нормальному. Наше исследование подтвердило предположение о том, что частоты фонем и фонемных групп распределены по нормальному закону, если выборка достаточно велика. Более подробные результаты изложены в следующей главе данной работы.

В отличие от среднего квадратического отклонения, коэффициент вариации (КВ) позволяет сравнивать варьирование частоты встречаемости фонем и групп фонем, имеющих разное значение средних. Обычно его обозначают латинским символом *V*. Коэффициент вариации (*V*) используют в качестве меры вариации (показателя устойчивости) лингвистических элементов в выборке или генеральной совокупности. При этом желательно, чтобы сравнивались величины коэффициента вариации, полученные на выборках одного объема, величина которого не меньше 30.

Коэффициент вариации приводит вариации к общему знаменателю и фактически представляет собою средний процент рассеяния случайной величины по отношению к средней арифметической (Пиотровский и др., 1977: 243). Коэффициент вариации равен:

$$V = S/M \times 100\%$$

где *V* – коэффициент вариации;

S – среднее квадратическое отклонение;

M – среднее арифметическое.

Обычно в любом вариационном ряду (или как мы говорили ранее “распределении”) существует некоторая изменчивость (неоднородность) или вариация, которая может проявляться сильнее или слабее через количественную изменчивость. Обычно неоднородность проявляется через чрезмерную вариацию признака (Сиськов, 1975: 32). В то же время считается, что большая и однородная выборка характеризуется малой дисперсией, что приводит к небольшим величинам коэффициента вариации (Тамбовцев, 2003). Некоторые лингвисты считают, что традиционный коэффициент вариации лучше использовать в виде его модификации, называемой коэффициентом стабильности Жийяна, который “работает” более эффективно (Тамбовцев, 2003: 13–18). В известной работе по определению авторства “Тихого Дона” Г. Хьетсо применяет коэффициент вариации, называя его “коэффициент варьирования” (Тамбовцев, 2003: 19). Лингвистическую устойчивость генеральной совокупности можно понимать как устойчивость в кибернетическом смысле, т.е. когда устойчивость – это способность системы стремиться из различных начальных состояний к некоторому равновесному (стационарному) состоянию. Состоянию покоя системы соответствует точка равновесия (ЭК, 1975: 468). На наш взгляд, коэффициент вариации позволяет несколько лучше измерять устойчивость, чем СКО. Так происходит потому, что в нём задействовано не только отклонение от среднего, но и само среднее, через которое и проявляется устойчивость. Известно, что устойчивость – это свойство,

состоящее в том, что отклонение её реальных выходных сигналов от идеальных не превышает допустимо малые величины, если сигналы возмущающих воздействий находятся в заданных пределах (ЭК, 1975: 478). Следовательно, закономерно возникает вопрос о допустимых пределах коэффициента вариации. В лингвистике, так же как и в других науках, связанных с человеком и его деятельностью, такие допустимые пределы выведены эмпирически, хотя разные исследователи считают приемлемым колебания до 50% (Мартыненко, 1988: 62). На наш взгляд, наиболее убедительные доводы приводит В.Н. Сиськов, который считает, что однородность группы, через которую и проявляется устойчивость, не может давать величину коэффициента вариации более 33% (Сиськов, 1971: 10). Так, Г.Я. Мартыненко считает совокупность текстов писателей конца 19 века – начала 20 века устойчивым и целостным образованием по некоторым синтаксическим признакам потому, что их коэффициент вариации намного меньше 33% (Мартыненко, 1988: 150–154). Р.Г. Пиотровский приводит пример устойчивости употребления английского определённого артикля и немецкого слова “мощь” в текстах ФРГ и ГДР, измеренной с помощью коэффициента вариации, который равен соответственно: 7,12% – 90,00% – 160,00%. Р.Г. Пиотровский отмечает, что “коэффициент вариации английского определённого артикля заметно меньше, чем коэффициент вариации у немецкого существительного. Это неудивительно: служебные формы обычно имеют во всех языках менее рассеянное употребление, чем знаменательные слова” (Пиотровский и др., 1977: 243). Следовательно, с точки зрения принятого нами порога величины коэффициента вариации английский определённый артикль распределяется устойчиво (ср. 7,12% и 33%), в то время как распределение немецкого существительного “мощь” как в текстах ФРГ, так и ГДР явно неоднородно. Как верно указывает Г.Я. Мартыненко, чем больше коэффициент вариации, тем больше шансов на то, что совокупность неоднородна. Он называет его критерием колеблемости (Мартыненко, 1988: 62). Такого же мнения придерживается В.И. Сиськов, который считает, что высокая величина коэффициента вариации может указывать на два распределения с резко отличными математическими ожиданиями (Сиськов, 1975: 101). Он считает также, что в таких случаях следует проверять части этого распределения с помощью критерия Фишера. Мы полностью согласны с этим, поэтому проверим некоторые наши распределения (или части распределений) по критерию Фишера ниже. Неоднородность распределения также проявляется в величине асимметрии (косости) и эксцесса (крутости), о чём более подробно также пойдёт речь ниже. Здесь же укажем, что коэффициент вариации чутко реагирует на асимметрию, эксцесс и многовершинность распределений. В частности, на наш взгляд, причиной большой величины коэффициента вариации у немецкого существительного “мощь”, скорее всего, могла быть многовершинность распределения. Это могло быть то, что в математической статистике называют трансгрессией. Многовершинность распределений часто связывается с явлением трансгрессии, заключающейся в том, что при слиянии двух совокупностей максимум распределений значений

признака одной из них заходит за минимум другой. Нужно избегать таких ситуаций и работать с одновершинными распределениями. Многовершинность распределения легко увидеть, если построить график распределения. Многовершинность распределения можно также увидеть по большому значению коэффициента вариации. Если между средними арифметическими рядов различия статистически достоверны, то такие ряды однозначно показывают различия между звуковыми картинами языков (Тамбовцев, 1982). Для того, чтобы избежать объединения в одно совокупность двух разнородных частей, необходимо количественному анализу предпосылать качественный анализ. Таким образом, чтобы ответить на вопрос о пороге коэффициента вариации в наших исследованиях нужно принять во внимание величины порогов, полученные другими авторами. Перейдём к анализу коэффициента вариации, полученного на двух разновидностях совокупности. В первом случае – это величины коэффициента вариации, полученные на таких совокупностях как подгруппы, группы, языковые семьи и общности, а во втором – это какой-либо язык.

Прежде, чем анализировать величину коэффициента вариации, полученную на материале фонемных групп, ещё раз обратим внимание на его величину в других исследованиях. Так, например, обсуждая коэффициент вариации веса и жирности коров с одной стороны, и жирности их молока, с другой стороны, Романовский вычисляет соответственно 12,0% и 5,0%, что хорошо вписывается в критический порог Сиськова, т.е. 33% (Тамбовцев, 2003). Флуктуации частоты пульса (13,30%), потоотделения (11,0%) и кровяного давления (5,0%) тоже не выходят за эти пределы (Венчиков и др., 1974: 20), при этом Венчиков полагает, что можно сравнивать изменчивости только тех процессов, коэффициент вариации которых не выходит за пределы 50% (Венчиков и др., 1974: 21). Интересно заметить, что измерение коэффициента вариации веса зерна главного колоса (30,6%), веса зерна всего растения (38,2%), количества колосков в главном колосе (11,1%) и количества семян в главном колосе (19,0%) также не выходят за пределы устойчивости по Сиськову. Довольно интересно анализирует порог коэффициента вариации Г.Н. Зайцев. Несмотря на то, что Г.Н. Зайцев признаёт критическими величины коэффициента вариации более 100%, его данные по продолжительности вегетационного периода у ячменя по сортам (8,1%), по высоте растений сои (18,6%) и роста листа жимолости (19%) не превышает критических 33%.

Изучив 165 различных признаков растений, Зайцев предлагает шкалу в процентах (%) следующего типа:

1. небольшое варьирование – от 0 до 4;
2. нормальное варьирование – от 5 до 44;
3. значительное – от 45 до 64;
4. большое – от 65 до 84;
5. очень большое – от 85 до 104;
6. аномальное – от 105 и более.

В пределах нормы Зайцев выделяет, так называемую, нижнюю норму от 5 до 24 (Зайцев, 1990: 39). Таким образом, нормальное варьирование по Зайцеву поднимается почти до 50%, что для других исследователей является запредельным вариантом. Следовательно, взгляды на предельный порог коэффициента вариации у разных ученых различны. Они, по-видимому, зависят от материала исследования. Нашей задачей будет установить критический порог коэффициента вариации для функционирования фонемных групп. Прежде всего, проанализируем, как в общем и целом ведёт себя коэффициент вариации в сфере, связанной с порождением звука речи. Нельзя утверждать, что частота употребления звука речи и его порождение в речевом тракте не связаны. Начнём наш небольшой обзор с колебаний коэффициента вариации при измерении длительности звуков. Нирей и Хоган вычисляли флуктуацию длительности звонкости в английских согласных. Они представляют среднюю и её дисперсию, из которых легко получить значение коэффициента вариации. Он колеблется в пределах 26,56–27,49%. Коэффициент вариации характеристики темпа речи остаётся в пределах 7,69–16,04%. Орфофоническую вариативность современной произносительной нормы показывает, что величина коэффициента вариации изменяется в интервале 14,97–27,83% при измерении долготы русских гласных в ударном и безударном положениях. Значение первой форманты (5,81–12,11%) и второй форманты (3,79–9,92%) при этом изменялось ещё меньше. Несколько большую величину имеет коэффициент вариации при измерении долготы фраз и строк в поэзии (23,50–46,25%) и прозе (24,62–38,23%), при этом поэзия намного более вариабельна, чем проза. В работе Нирей и Дервин приводятся данные по ошибкам восприятия смычных согласных, при этом коэффициент вариации изменяется в пределах 16,57–25,21%. Коэффициент вариации ошибки при реакции на фонему “к” составил 13,16%, а на аффрикаты – 47,83%. Как мы видим, коэффициент вариации колеблется вокруг 33% и довольно часто не выходит за пределы 50%, что как указывалось выше, очень важно. Эти значения коэффициента вариации указывают на нормальное распределение измеряемых величин, что позволяет нам без опаски применять методы математической лингвистики к распределению фонем в потоке речи. Здесь же отметим мнение Е.Н. Львовского, который считает, что при коэффициенте более 33% есть большая вероятность, что распределение не подчиняется нормальному закону (Львовский, 1982: 38; Тамбовцев, 2003).

Перейдём к собственно анализу частоты встречаемости фонемных групп в языке. В частности, прежде, чем рассматривать встречаемость смычных согласных, начнём с самой крупной группы: согласные и гласные, каждая из которых в свою очередь разбивается на более мелкие подгруппы. Функционирование частоты встречаемости группы согласных и группы гласных проанализируем в виде величин консонантного коэффициента, то есть соотношения согласных и гласных в потоке речи. С точки зрения коэффициента вариации консонантный коэффициент ещё не анализировался, хотя по другим параметрам (например, величине доверительного

интервала, СКО и др.) детальный анализ его распределений был сделан в нашей работе ранее (Тамбовцев, 1986). Ниже приведём таблицу, в которой распределим языки по их генетическим семьям, в которых они будут упорядочены по величине коэффициента вариации (Таб. 27). Обзор данной таблицы показывает, что внутри финно-угорской языковой семьи величина коэффициента вариации (КВ) изменяется в интервале 2,96–5,15%, в тюркских языках (2,94–5,24%) колебания этого коэффициента немного больше, в тунгусо-маньчжурских языках, где он колеблется больше, чем в финно-угорских (ср. 2,82–5,18), а в индо-европейских языках меньше, чем в финно-угорских языках (ср. 3,48–5,81). В то же время словари показывают большую величину КВ (6,36–9,32%). Таким образом, величина КВ при измерении соотношения согласных и гласных в потоке речи колеблется в довольно узких пределах: 2,82–5,81%. Немного больше она на материале словарей, но ей далеко до 50% или даже 33%, которые считаются критическими точками для коэффициента вариации. Как было видно выше, характеристики звуков речи при речепроизводстве также находятся в пределах критических точек. Несколько иную картину мы наблюдаем при измерении частотности фонемных групп согласных, которую рассмотрим несколько подробнее. Мы приводим величину коэффициента вариации смычных в различных языках мира (Таб. 27). По величине коэффициента вариации распределения групп согласных в различных языках, можно сделать вывод о том, что у губных этот коэффициент обычно не превышает 11–18%. В персидском языке этот процент значительно выше, но это, вероятно, потому, что персидский материал состоит из смеси текста и словаря (42%). Величина этого коэффициента у переднеязычных согласных не выходит за пределы 4–12%, но опять у персидского языка он больше – 16,70%. По величине коэффициента вариации среднеязычные согласные распределены в интервале 15,91–69,39%, хотя в персидском языке эта величина неожиданно меньше – 8,85%. Границы коэффициента вариации у заднеязычных согласных составляют 11–22%, в то же время опять выпадает из этого ряда персидский с его 46,51%.

Интервал распределения у сонорных – 4,13%, – а у персидских сонорных – 19,79%. Интервал колебания смычных согласных также невелик (4–15%), не учитывая персидские смычные – 28,76%. Разрыв флуктуации щелевых согласных ещё меньше: 10–16%, при этом даже персидские щелевые сохраняют норму – 33,99%. Из таблицы также ясно, что в узких пределах изменяется коэффициент вариации у звонких согласных (11–22%), у мягких согласных (18–21%), у твёрдых согласных (2,95–5,54%) и у всех согласных (2–7%), при этом даже персидские согласные сохраняют эту норму (ср. 4,03%).

Интересно сопоставить некоторые данные по функционированию групп фонем в маргинальных (начало и конец) позициях слова с аналогичными данными во всех позициях слова. Так, например, глухие согласные в японском языке в начале слова имеют разные коэффициент вариации (ср. 11,35%) по сравнению с общей фонемной цепочкой (7,01%). В то же время в цыганском языке эти коэффициенты

в этих двух позициях различаются довольно незначительно (ср. 10,89% и 14,19%), но резко в финальной позиции (39,99%). Также резко различаются коэффициенты вариации глухих согласных.

Мы и ранее применяли величину коэффициента вариации (V) для измерения схожести языков в каких-либо языковых таксонах (Тамбовцев, 1990; 1992).

Рекомендация измерять величину коэффициента вариации важна для лингвистических исследований. На наш взгляд, если эта величина больше 33%, то распределение смычных одномодальное. Если эта величина больше 33%, то распределение может быть двумодальное. Наши исследования показали, что все таксоны имеют величину коэффициента меньше 33%, что показано в Таб. 27.

Обсуждение результатов исследования

Палео-азиатские языки имеют значительную смычную гортанную окраску в своей звуковой цепочке. Минимум степени гортанной смычности мы находим в чукотском языке (21,21%), а максимум – в имаclinском диалекте эскимосского языка (26.8.21%) при средней частоте встречаемости 24,03% в звуковой цепочке палеоазиатских языков. Мы можем отметить, что чукотский употребляет меньше смычных, чем ему следует, а имаclinский диалект эскимосского – больше. Сравнивая палео-азиатские языки с другими языками, например, с тюркскими, мы видим, что они имеют большую степень смычности. В тюркских языках (Таб. 2) в среднем употребляется 23,38% смычных. При этом, некоторые языки, например, саларский употребляют значительно меньше смычных, чем в среднем по тюркским языкам, а другие, например, кумандинский – значительно больше смычных, чем в среднем. Это превышает степень смычности у других языковых таксонов, например, у монгольских (Таб. 3) или тунгусо-маньчжурских языков (Таб. 4). В тоже время, следует обратить внимание, насколько схожи по параметру смычности языки внутри групп, т.е. таксонов. Схожесть или несхожесть языков проявляется через величину коэффициента вариации. Чем больше величина коэффициента вариации, тем более непохожи друг на друга языки. Наоборот, чем меньше величина данного коэффициента, тем более похожи языки (Тамбовцев, 2010). Мы можем говорить о компактности какого-либо таксона по данному параметру. В частности, величина компактности измерена у 17 таксонов языков мира (Таб. 27). Было выявлено, что наиболее схожи – это балтийские языки, которые имеют наименьший разброс $V\% = 4,83\%$. Наибольший коэффициент вариации показали аустронезийские языки – 26,07%. Процедура вычисления коэффициента вариации более детально показана в нашей монографии (Тамбовцев, 2003).

Прежде чем перейти к анализу функционирования смычных согласных в потоке речи, проанализируем функционирование в звуковой цепочке распределение каких-то других групп фонем, например, гуттуральных согласных.

В финно-угорских языках частота гуттуральных согласных равна в среднем 9,09 при коэффициенте вариации $V = 19,59$. Значительная величина коэффициента вариации (V) вызвана значительной разницей в степени гортанности в финно-угорских языках. Так, в своем минимальном значении гортанность обнаруживается у коми-зырянского языка (5,94%), а в максимальном у восточного диалекта хантыйского языка (13,53%). В самодийских языках в среднем гуттуральные встречаются больше – 14,09%, их разброс тоже больше – $V = 21,15$. В тюркских языках употребление гуттуральных более равномерное, на что указывает более низкий показатель коэффициента вариации, хотя в среднем в звуковой цепочке больше гуттуральных 11,39 при $V = 15,54$. О месте финно-угорских языков (9,09) среди других языков мира можно судить по упорядоченному ряду употребления гуттуральных в среднем.

Величина смычности в различных языковых таксонах

Для того, чтобы понять место палео-азиатских языков по типологии распределения гуттуральных согласных, необходимо построить упорядоченные ряды по величине минимальной, максимальной и средней гуттуральности во всех изученных нами таксонах языков мира. Отметим, что степень гуттуральности рассчитывается в процентах по отношению ко всем элементам звуковой цепочки языка.

Упорядоченный ряд средней величины смычности по языковым таксонам представлен в Таб. 26. Эта таблица показывает, что в среднем смычные согласные в языках мира распределяются в пределах от 19,49% в языках австралийских аборигенов до 22,67% в самодийских языках.

Можно рассчитать среднюю теоретическую величину встречаемости какого-либо фонологического признака в звуковой цепочке языка (Тамбовцев 2009). Для такого расчета требуется несколько вероятностных предположений. Звуковая цепочка языка состоит из согласных и гласных. Допустим, что они употребляются с одинаковой вероятностью в 50%. Далее нужно учесть, что по работе активного органа речеобразования все согласные делятся на 4 группы: 1) губные; 2) переднеязычные; 3) среднеязычные (палатальные) и 4) гуттуральные (Тамбовцев 2008). Допустим, что все эти группы употребляются равновероятно. Следовательно, вероятность появления гуттуральных согласных в звуковой цепочке языка равна $50\% : 4 = 12,50\%$. Таким образом, теоретическая величина среднего значения для гуттуральных согласных для любого языка мира составит 12,50%. Если исходить из теории экономии усилий, то в любом языке частота встречаемости гуттуральных должна быть значительно ниже. Так должно происходить потому, что артикулировать звуки с помощью кончика языка намного легче, чем с помощью задней части или корня языка.

Однако, теоретическая величина для точных исследований не подходит потому, что мы сделали «грубые» вероятностные допуски. На самом деле, нужно вычислять

эмпирическую среднюю величину, которая зависит от конкретной величины встречаемости гуттуральных в звуковой цепочке различных языков мира. В нашем исследовании задействовано более 200 естественных и 1 искусственный язык (эсперанто). За исключением эсперанто, средняя величина гуттуральных по 200 языкам составила 12,21%. Интересно отметить, что эта величина близка к средней теоретической (ср. 12,21% и 12,50%). Если это не простое совпадение, то здесь проявилась какая-то универсальная закономерность речеобразования.

Все языки, которые имеют в своей звуковой цепочке концентрацию гуттуральных меньше 12,21% не могут считаться гортанными. Из двадцати финно-угорских языков только один язык – язык восточных хантов в какой-то мере может считаться гортанным, т.к. его значение 13,53%. В то же время, все палеоазиатские языки превышают 12,21%. В этом отношении палео-азиатские языки похожи на многие языки американских индейцев, которые тоже имеют большую частоту встречаемости гуттуральных в своей звуковой цепочке.

Такой же алгоритм расчетов можно применить и к смычным согласным; 50% гласных и 50% согласных, но здесь уже согласные имеют три группы (сонорные, смычные и щелевые). Из этого следует, что вероятность употребления смычной согласной в звуковой цепочке составляет 16,67%. Известный американский математик Д. Поля провёл этим методом статистический анализ 10 европейских языков по частоте встречаемости отдельных звуков и обнаружил, что венгерский язык непохож на них на фонетическом уровне (Polya, 1954). Из этого следует, в нашем случае, что те языки, которые имеют концентрацию смычных больше теоретического числа, употребляют слишком много смычных в своей звуковой цепочке, и наоборот. Таб. 1–24 показывают, что языков, у которых идет превышение частоты встречаемости смычных, достаточно много.

Известный английский математик Хердан предлагает использование метода «хи-квадрат», чтобы проанализировать те случаи, когда в каком-нибудь языке употребляется слишком много или слишком мало лингвистических единиц (Herdan, 1966). В частности, мы можем точно указать на те языки, которые в своей звуковой цепочке употребляют слишком много или слишком мало каких-либо выделенных нами групп фонем, например, губных, гуттуральных, сонорных и других (Тамбовцев, 2010; Tambovtsev, 2010). На основе частоты встречаемости групп согласных в звуковой цепочке языка можно построить модель, которая покажет типологические расстояния между языками (Tambovtsev, 2009).

В качестве вывода по употреблению смычных согласных в звуковых цепочках различных языков мира можно отметить что:

1. Наименьшее значение употребления смычных согласных (9,27%) наблюдается в самоанском языке (аустронезийский таксон), а наибольшее (33,30%) – в языке американских индейцев тотонак. (Таб. 24 и Таб. 25).
2. Упорядоченный ряд средней величины смычности по языковым таксонам представлен в Таб. 26. Эта таблица показывает, что в среднем смычные

согласные в языках мира распределяются в пределах от 19,49% в языках австралийских аборигенов до 22,67% в самодийских языках.

3. Распределение смычных согласных по всем взятым языкам (Таб. 1–26) менее 33%, что однозначно говорит об одномодальности. Это показывает то, что величину средних по смычным во всех взятых языках можно сравнивать без опаски.

Литература

- Асмус, 1965 – В.Ф. Асмус, *История античной философии*, Высшая школа, Москва 1965.
- Вентцель, 1964 – Е.С. Вентцель, *Теория вероятностей*, Наука, Москва 1964.
- Венчиков и др., 1974 – А.И. Венчиков, В.А. Венчиков, *Основные приемы статистической обработки результатов наблюдений в области физиологии*, Медицина, Москва 1974.
- Гамкрелидзе, 1986 – Т.В. Гамкрелидзе, *Языковое развитие и предпосылки сравнительно-генетического языкознания*, в: *Литература, язык, культура*, Москва 1986, с. 201–208.
- Головин, 1971 – Б.Н. Головин, *Язык и статистика*, Просвещение, Москва 1971.
- Зайцев, 1990 – Г.Н. Зайцев, *Математика в экспериментальной ботанике*, Наука, Москва 1990.
- Зиндер, 1979 – Л.Р. Зиндер, *Общая фонетика*, Высшая школа, Москва 1979.
- Львовский, 1982 – Е.Н. Львовский, *Статистические методы построения эмпирических формул*, Высшая школа, Москва 1982.
- Сиськов, 1975 – В.И. Сиськов, *Корреляционный анализ в экономических исследованиях*, Статистика, Москва 1975.
- Тамбовцев, 1982 – Ю.А. Тамбовцев, *Эмпирическое распределение частотности фонем в казымском диалекте хантыйского языка*, „Ученые записки Тартусского ГУ”, 1982, № 628, с. 121–134.
- Тамбовцев, 1990 – Ю.А. Тамбовцев, *Компактность финно-угорской языковой семьи по данным консонантного коэффициента*, „Linguistica Uralica” 1990, Том 26, № 1, с. 13–20.
- Тамбовцев, 1992 – Ю.А. Тамбовцев, *Оцінка компактності груп мов на фонетичному рівні*, „Іноземна філологія” 1992, № 103, с. 79–84.
- Тамбовцев, 1994-а – Ю.А. Тамбовцев, *Типологія упорядоченності звукових ланцюгів в мові*, НГУ, Новосибірськ 1994.
- Тамбовцев, 1994-б – Ю.А. Тамбовцев, *Динаміка функціонування фонем в звукових ланцюжках мов різного строю*, НГУ, Новосибірськ 1994.
- Тамбовцев 2001-а – Ю.А. Тамбовцев, *Компендіум основних статистических характеристик функціонування согласных фонем в звуковій ланцюжці англійського, німецького, французького і других індоевропейських мов*, Новосибирский классический институт, Новосибирск 2001.
- Тамбовцев, 2008 – Ю.А. Тамбовцев, *Звуковая картина языка как совокупность цепочек звуков в речи. Часть 1.*, „Acta Linguistica” 2008, № 2, с. 3–13.
- Тамбовцев 2009 – Ю.А. Тамбовцев, *Звуковая картина языка как совокупность цепочек звуков речи. Часть 2: гласные и сонорные согласные как составляющие мелодичности звуковой картины языка*, „Acta Linguistica” 2009, № 1.
- Тамбовцев, 2010-а – Ю.А. Тамбовцев, *Степень гортанности в палео-азиатских и других языках Азии*, „Rocznik Orientalistyczny” 2010, Т. LXIII, z. 2, с. 133-150.
- Тамбовцев, 2010-б – Ю.А. Тамбовцев, *Типологические расстояния между казахским языком и некоторыми тюркскими языками на фонологическом уровне*, „Сходознавство” 2010, № 49–50, с. 170–176.
- Тамбовцев, 2011 – Ю.А. Тамбовцев, *Фоно-типологическая близость по характеристикам распределения групп фонем в звуковых цепочках (на материале эрзя-мордовского и мокша-мордовского языков)*, „Финно-угорский мир” 2011, № 1, с. 38–41.

Тамбовцев, 2012-а – Ю.А. Тамбовцев, *Распределение групп фонем в звуковых цепочках языков и диалектов*, „Сибирский филологический журнал” 2012, № 3, с. 146-152.

Тамбовцев, 2012-б – Ю.А. Тамбовцев, *Этнографическая криминалистика: к какому народу ближе всего венгры по данным дактилоскопии?*, „Journal of Eurasian Studies”, July-September 2012, Volume 4, Issue 3, с. 22–37.

Широков 1985 – О.С. Широков, *Введение в языкознание*, МГУ, Москва 1985.

ЭК, 1975 – *Энциклопедия кибернетики*, Укр. Сов. Энцикл., Киев 1975.

Brainerd, 1974 – B. Brainerd, *Weighing Evidence in Language and Literature: A statistical Approach*, University of Toronto Press, Toronto 1974.

Butler, 1985 – C. Butler, *Statistics in Linguistics*, Basil Blackwell, Oxford 1985.

Halle et al., 1957 – M. Halle, G.W. Hughes, J.P.A. Radley, *Acoustic Properties of Stop Consonants*, „Journal of Acoustic Society of America” 1957, No. 29, p. 107–116.

Herdan, 1966 – G. Herdan, *The Advanced Theory of Language as Choice and Chance*, Springer Verlag, Berlin 1966.

Polya, 1954 – G. Polya, *Patterns of Plausible Inference*, Volume 2, Princeton University Press, Princeton 1954.

Tambovtsev 1983 – Yuri Alex. Tambovtsev, *Selected consonantal characteristics of some Finno-Ugric language from a phonostatistical point of view*, „Finnish-Ugrische Forschungen” 1983, XLV, H. 1–3, pp. 154–163.

Tambovtsev 1985 – Yu. A. Tambovtsev, *The Consonantal Coefficient in Selected Languages*, „Canadian Journal of Linguistics”, Summer 1985, 30(2).

Tambovtsev 2006 – Yuri Tambovtsev, *Degrees of Dispersion of Various Language Taxa as the Index of Their Typological Similarity*, in: *Favete Linguis. Studies in Honour of Viktor Krupa*, Institute of Oriental Studies of Slovak Academy of Sciences, Bratislava 2006, pp. 219–260.

Tambovtsev, 2009 – Yuri Tambovtsev, *Is the taxon of Altaic languages natural or artificial?*, „Lingua Posnaniensis” 2009, Vol. 51, pp. 133–154

Tambovtsev, 2010 – Yuri Tambovtsev, *Labial consonant distribution in Niger-Congo and other language taxa: A basis for determining typological distance*, „California Linguistic Notes” Spring 2010, Vol. XXXV, 2, pp. 1–35.

Wray et.al., 1998 – Alison Wray, Aileen Bloomer, *Projects in Linguistics*, Arnold Publishers, New York 1998.

Таблица 1.

Частота встречаемости смычных согласных в звуковой цепочке палео-азиатских языков, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Чукотский	21,21	4.	Ительменский	24,27
2.	Корякский	23,14	5.	Эскимосский (имаклинский)	28,21
3.	Эскимосский (науканский)	23,33			
	статистика:				
	среднее	24,03		S ²	6,69
	S	2,59		V%	10,76

Таблица 2.

Частота встречаемости смычных согласных в звуковой цепочке тюркских языков, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Саларский	16,58	14.	Уйгурский	23,17
2.	Чувашский	18,30	15.	Татарский (казанский)	23,17
3.	Башкирский	18,75	16.	Турецкий	23,22
4.	Азербайджанский	18,94	17.	Татарский (Барабинский)	23,54
5.	Туркменский	20,12	18.	Тофаларский	23,87
6.	Хакасский	20,42	19.	Шорский	24,69
7.	Каракалпакский	20,97	20.	Тувинский	24,79
8.	Татарский (Крымский)	21,55	21.	Татарский (Чулымский)	25,35
9.	Карачаево-Балкарский	21,98	22.	Киргизский	25,76
10.	Сары-уйгурский	22,08	23.	Алтайский (Чалканский)	27,63
11.	Якутский	22,14	24.	Долганский	28,48
12.	Узбекский	22,16	25.	Алтайский (Кизи)	29,12
13.	Казахский	23,12	26.	Кумандинский	29,67
	статистика:				
	среднее	23,38		S ²	10,05
	S	3,17		V%	13,54

Таблица 3.

Частота встречаемости смычных согласных в звуковой цепочке монгольских языков, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Калмыцкий	21,25	3.	Бурятский	25,33
2.	Монгольский (Халха)	22,66			
	статистика:				
	среднее	23,08		S ²	4,29
	S	2,07		V%	8,97

Таблица 4.

Частота встречаемости смычных согласных в звуковой цепочке тунгусо-маньчжурских языков, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Маньчжурский	18,71	6.	Нанайский	23,58
2.	Удэгейский	19,98	7.	Эвенский (ламутский)	24,14
3.	Негидальский	21,72	8.	Эвенкийский (тунгусский)	25,08
4.	Ульчский	23,71	9.	Орочский	25,46
5.	Орокский	23,10			
	статистика:				
	среднее	22,81		S ²	5,38
	S	2,32		V%	10,17

Таблица 5.

Частота встречаемости смычных согласных в звуковой цепочке языков афразийской семьи, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Иврит	15,13	5.	Сомалийский	20,56
2.	Хауса	19,89	6.	Нео-Арамейский	21,73
3.	Арабский	20,34	7.	Сокотрийский	22,95
4.	Ассирийский	20,53			
	статистика:				
	среднее	20,16		S ²	6,00
	S	2,45		V%	12,15

Таблица 6.

Частота встречаемости смычных согласных в звуковой цепочке сино-тибетских языков, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Бирманский	16,10	4.	Тайский	25,51
2.	Дунганский	19,97	5.	Тибетский	31,74
3.	Китайский	21,51			
	статистика:				
	среднее	22,97		S ²	35,40
	S	5,95		V%	25,90

Таблица 7.

Частота встречаемости смычных согласных в звуковой цепочке самодийских языков, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Ненец (тундр.)	22,67	5.	Камасинский	27,28
2.	Энецкий (Бай)	23,14	6.	Сельк. (ср.обск)	30,83
3.	Энецк. (Маду)	23,95	7.	Сельк. (обской)	32,89
4.	Нганасанский	26,44			
	статистика:				
	среднее	26,74		S ²	15,37
	S	3,92		V%	14,66

Таблица 8.

Частота встречаемости смычных согласных в звуковой цепочке финно-угорских языков, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Марийский (Горный)	16,35	11.	Коми-зырянский	20,65
2.	Мансийский (Кондинский)	16,56	12.	Мокша-мордовский	20,86
3.	Мансийский (Северный)	17,00	13.	Эрзя-мордовский	21,36
4.	Хантыйский (северный)	17,19	14.	Карел(людик)	21,67
5.	Финский	18,00	15.	Водский	21,93
6.	Мари-луговой	18,22	16.	Удмуртский	21,98
7.	Карел(ливви)	20,22	17.	Венгерский	22,62
8.	Коми-пермяцкий	20,33	18.	Саамский	23,55
9.	Карел(тихвинский)	20,36	19.	Ханты (восточный)	24,20
10.	Эстонский	20,45	20.	Вепсский	24,71
	статистика:				
	среднее	20,41		S ²	6,27
	S	2,50		V%	12,27

Таблица 9.

Частота встречаемости смычных согласных в звуковой цепочке языков аустронезийской семьи, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Самоанский	9,27	11.	Кебуано	20,39
2.	Гавайский	11,70	12.	Капампанган	21,24
3.	Малагаши	12,55	13.	Тонган (Малай)	22,37
4.	Фиджи	15,53	14.	Чаморро	22,51
5.	Рукаи (Тайван.)	16,43	15.	Льокано	24,68
6.	Таити	16,76	16.	Ума (Сулавеси)	25,20
7.	Тагалог (Фили.)	19,00	17.	Даякский	25,78
8.	Маори	19,10	18.	Индонезийский	26,78
9.	Тсоу	19,29	19.	Маркизский	27,35
10.	Хилигайнон	21,04			
	статистика:				
	среднее:	19,84		S ²	26,76
	S	5,17		V%	26,07

Таблица 10.

Частота встречаемости смычных согласных в звуковой цепочке языков Африки (банту, малави, нигер-конго, нило-сахарские) в % ко всем фонемам в потоке речи

№	Язык	%	№	Язык	%
1.	Суахили	16,44	8.	Цевана	21,13
2.	Ханга	17,89	9.	Муре	22,13
3.	Бемба	18,56	10.	Тонга	22,14
4.	Х'оса	18,69	11.	Луганда	22,76
5.	Зарма	19,66	12.	Фульфульде	23,38
6.	Киньяранда	19,85	13.	Волоф	23,65
7.	Чичева	20,90	14.	Луо	23,72

Таблица 11.

Частота встречаемости смычных согласных в звуковой цепочке языков австралийских аборигенов, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Нунгубуйу	14,86	4.	Ньангумада	19,20
2.	Мангарайи	18,07	5.	Нгааньатджара	22,90
3.	Дьингили	18,73	6.	Нганди	23,16
	статистика:				
	среднее	19,49		S ²	9,83
	S	3,13		V%	16,09

Таблица 12.

Частота встречаемости смычных согласных в звуковой цепочке языков кавказской семьи, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Кабардинский	16,52	4.	Грузинский	23,00
2.	Адыгейский	19,13	5.	Абхазский	24,46
3.	Чеченский	21,00	6.	Аварский	24,76
	статистика:				
	среднее	21,48		S ²	10,37
	S	3,22		V%	14,99

Таблица 13.

Частота встречаемости смычных согласных в звуковой цепочке языков индо-арийской группы индо-европейской семьи, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Гуджарати	20,88	4.	Хинди	26,04
2.	Цыганский	21,62	5.	Бенгальский	26,28
3.	Маратхи	24,13			
	статистика:				
	среднее	23,79		S ²	6,14
	S	2,48		V%	10,41

Таблица 14.

Частота встречаемости смычных согласных в звуковой цепочке языков иранской группы индо-европейской семьи, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Пашто	19,90	5.	Осетинский	22,09
2.	Иранский	20,61	6.	Дари (Афганистана)	22,67
3.	Тальшский	20,68	7.	Курдский	23,86
4.	Гилянский	21,29	8.	Таджикский	24,28
	статистика:				
	среднее	21,92		S ²	2,52
	S	1,59		V%	7,24

Таблица 15.

Частота встречаемости смычных согласных в звуковой цепочке языков славянской группы индо-европейской семьи, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Болгарский	14,96	7.	Словацкий	20,82
2.	Сербо-лужицкий	18,79	8.	Словенский	21,08
3.	Русский	20,14	9.	Польский	21,16
4.	Древнерусский	20,17	10.	Украинский	21,18
5.	Сербско-хорватский	20,37	11.	Белорусский	21,91
6.	Чешский	20,47	12.	Македонский	24,12
	статистика:				
	среднее	20,43		S ²	4,58
	S	2,14		V%	10,47

Таблица 16.

Частота встречаемости смычных согласных в звуковой цепочке языков балтийской группы индо-европейской семьи, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Латышский	20,78	2.	Литовский	22,25
	статистика:				
	среднее	21,52		S ²	1,08
	S	1,04		V%	4,83

Таблица 17.

Частота встречаемости смычных согласных в звуковой цепочке языков романской группы индо-европейской семьи, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Португальский	17,99	4.	Итальянский	20,60
2.	Французский	19,14	5.	Молдавский	21,21
3.	Испанский	19,44	6.	Румынский	21,54
	статистика:				
	среднее	19,99		S ²	1,85
	S	1,36		V%	6,80

Таблица 18.

Частота встречаемости смычных согласных в звуковой цепочке языков германской группы индо-европейской семьи, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Датский	17,61	4.	Английский	20,33
2.	Норвежский	18,80	5.	Нидерландский	21,24
3.	Немецкий	19,44	6.	Шведский	21,90
	статистика:				
	среднее	19,89		S ²	2,53
	S	1,59		V%	8,00

Таблица 19.

Частота встречаемости смычных согласных в звуковой цепочке языков италийской группы индо-европейской семьи, в % ко всем фонемам в потоке речи

№	Язык	%
1.	Латинский	22,92

Таблица 20.

Частота встречаемости смычных согласных в звуковой цепочке языков американских индейцев, в % ко всем фонемам в потоке речи Взятые различные таксоны, т.е. группы и семьи

№	Язык	%	№	Язык	%
1.	Гуарани	14,59	19.	Покомчи	25,52
2.	Икито	17,13	20.	Апинайэ	25,70
3.	Капанауа	19,49	21.	Каваскар	25,72
4.	Сириано	19,60	22.	Навахо	26,60
5.	Кечуа	20,81	23.	Сайула популука	27,34
6.	Суит грасс кри	21,25	24.	Вичита	27,45
7.	Джакалтек	21,73	25.	Онеида	27,98
8.	Науатль (Ацтекский)	22,19	26.	Тонкава	27,98
9.	Секойа	22,41	27.	Оджибве	28,39
10.	Арикара	22,57	28.	Кадивеу	28,80
11.	Гуамбиано	23,26	29.	Кокопа	28,97
12.	Инга	23,29	30.	Овекено	28,98
13.	Кекчи	23,39	31.	Мам	29,56
14.	Тзутуджил	23,68	32.	Пиратапуйо	29,85
15.	Кофан	24,02	33.	Хайда	29,88
16.	Кайва	24,64	34.	Хуастек	30,56
17.	Акатеко	25,19	35.	Тотонак	33,30
18.	Киче	25,24			

Таблица 21.

Частота встречаемости смычных согласных в звуковой цепочке языков балканского языкового союза, в % ко всем фонемам в потоке речи (S – среднее квадратическое отклонение; V% – коэффициент вариации)

№	Язык	%	№	Язык	%
1.	Болгарский	14,96	4.	Румынский	21,17
2.	Греческий	17,09	5.	Албанский	28,61
3.	Сербско-хорватский	20,37			
	статистика:				
	среднее	20,44		S ²	27,14
	S	5,21		V%	25,48

Таблица 22.

Частота встречаемости смычных согласных в звуковой цепочке изолированных языков, в % ко всем фонемам в потоке речи

№	Язык	%	№	Язык	%
1.	Греческий	17,09	6.	Корейский	23,88
2.	Айнский	19,11	7.	Армянский	24,66
3.	Нивхский	21,24	8.	Японский	24,73
4.	Баскский	21,92	9.	Кетский (Югский)	26,94
5.	Юкагирский	22,31	10.	Албанский	28,61

Таблица 23.

Частота встречаемости смычных согласных в звуковой цепочке искусственного языка эсперанто, в % ко всем фонемам в потоке речи

№	Язык	%
1.	Эсперанто	18,28

Таблица 24.

Минимальные значения частоты встречаемости смычных согласных в звуковых цепочках языков мира, в % ко всем фонемам в потоке речи

№	Язык	%
1.	Самоанский. Аустронезийский таксон.	9,27
2.	Гуарани. Языки американских индейцев.	14,59
3.	Нунгубуйу. Языки австралийских абориг.	14,86
4.	Болгарский. Славянский таксон.	14,96
5.	Иврит. Афразийский таксон.	15,13
6.	Бирманский. Сино-тибетский таксон.	16,10

№	Язык	%
7.	Горно-марийский. Финно-угорский такс.	16,35
8.	Суахили. Языки Африки.	16,44
9.	Кабардинский. Кавказский таксон.	16,52
10.	Саларский. Тюркский таксон.	16,58
11.	Греческий. Изолированные языки.	17,09
12.	Датский. Германский таксон.	17,61
13.	Португальский. Романский таксон.	17,99
14.	Маньчжурский. Тунгусо-маньчжурский	18,71
15.	Пашто. Иранский таксон.	19,90
16.	Латышский. Балтийский таксон.	20,78
17.	Гуджарати. Индо-арийский таксон.	20,88
18.	Чукотский. Палеоазиатский таксон.	21,21
19.	Калмыцкий. Монгольский таксон.	21,25
20.	Ненецкий (Тундровый). Самодийский.	22,67

Таблица 25.

Максимальные значения частоты встречаемости смычных согласных в звуковых цепочках языков мира, в % ко всем фонемам в потоке речи

№	Язык	%
1.	Румынский. Романский таксон.	21,54
2.	Шведский. Германский таксон.	21,90
3.	Литовский. Балтийский таксон.	22,25
4.	Сокотрийский. Афразийские.	22,95
5.	Нганди. Таксон языков австрал. абориген.	23,16
6.	Луо. Языки Африки.	23,72
7.	Македонский. Славянский таксон.	24,12
8.	Таджикский. Иранский таксон.	24,28
9.	Вепский. Финно-угорский таксон.	24,71
10.	Аварский. Кавказский таксон.	24,76
11.	Бурятский. Монгольский таксон.	25,33
12.	Орочский. Тунгусо-маньчжурский таксон.	25,46
13.	Бенгальский. Индо-арийский таксон.	26,28
14.	Маркизский. Аустронезийский таксон.	27,35
15.	Эскимосский (Имаклинский). Палеоазиат.	28,21
16.	Албанский. Изолированный.	28,61
17.	Кумандинский. Тюркский таксон.	29,67
18.	Тибетский. Сино-тибетский таксон.	31,74
19.	Селькупский (обской). Самодийский такс.	32,89
20.	Тотонак. Языки американских индейцев.	33,30

Таблица 26.

Средние значения частоты встречаемости смычных согласных в звуковых цепочках языков мира, в % ко всем фонемам в потоке речи. Различные языковые таксоны (т.е. подгруппы, группы, семьи)

№	Языковой таксон. Среднее значение частоты встречаемости по всем языкам	%
1.	Таксон языков австралийских аборигенов.	19,49
2.	Аустронезийский таксон.	19,84
3.	Германский таксон.	19,89
4.	Романский таксон.	19,99
5.	Афразийский таксон.	20,16
6.	Финно-угорский таксон.	20,41
7.	Славянский таксон.	20,43
8.	Кавказский таксон.	21,48
9.	Балтийский таксон.	21,52
10.	Иранский таксон.	21,92
11.	Сино-тибетский таксон.	22,97
12.	Монгольский таксон.	23,08
13.	Индо-арийский таксон.	23,79
14.	Тюркский таксон.	23,38
15.	Палеоазиатский таксон.	24,03
16.	Самодийский таксон.	26,74

Таблица 27.

Значения коэффициента вариации как показатель схожести языков в языковом таксоне. Это проявляется в плотности (т.е. компактности) языкового таксона, %

№	Плотность (компактность) языков. таксона	%
1.	Балтийский таксон.	4,83
2.	Романский таксон.	6,80
3.	Иранский таксон.	7,24
4.	Германский таксон.	8,00
5.	Монгольский таксон.	8,97
6.	Тунгусо-маньчжурский таксон.	10,17
7.	Индо-арийский таксон.	10,41
8.	Славянский таксон.	10,47
9.	Палеоазиатский таксон.	10,76
10.	Афразийский таксон.	12,15
11.	Финно-угорский таксон.	12,27
12.	Тюркский таксон.	13,54
13.	Самодийский таксон.	14,66
14.	Кавказский таксон.	14,99
15.	Таксон австралийских аборигенов.	16,09
16.	Сино-тибетский таксон.	25,90
17.	Аустронезийский таксон.	26,07