
**ТАМБОВЦЕВ, Юрий Алексеевич & ТАМБОВЦЕВА, Алина
Юрьевна & ТАМБОВЦЕВА, Людмила Алексеевна**

**БАЛТО-СЛАВЯНСКОЕ ЕДИНСТВО: РЕАЛЬНОСТЬ ИЛИ
ФИКЦИЯ?**

Аннотация

Применение критерия “хи-квадрат” позволяет объективно измерить расстояние между языками по звуковой картине. Было измерено фоно-типологическое расстояние между балтийскими и славянскими языками. Фоно-типологическое расстояние между литовским и латышским намного больше, чем между литовским и древнерусским. Литовский язык по его звуковой картине более похож на современный русский (6,07), чем на латышский.

В свою очередь, латышский язык по звучанию еще больше похож на древнерусский язык (2,47), чем литовский. Латышский язык также ближе по его звучанию к современному русскому (3,65), чем литовский.

Близость балтийских языков по их звуковой картине к славянским языкам можно объяснить как былой балто-славянской общностью, так и интенсивными языковыми контактами.

В статье предлагается метод исследования, который решит давний спор, является ли балто-славянское единство реальностью или же это фикция. Этот метод основан на измерении типологической близости языков. Он применяет некоторые критерии фонологической статистики, используемые в языкознании. Фоно-метрические, т.е. типологические характеристики звуковых цепочек языка проявляются через их частотные величины в звуковом потоке любого языка, поэтому они могут быть измерены [Тамбовцев 2010].

Целью данной работы является установление фоно-типологических расстояний между славянскими и балтийскими языками, чтобы установить их единство. Можно применять различные критерии математической лингвистики. Нами был выбран критерий «хи-квадрат» как один из самых простых, но достаточно мощных. Так, по величине критерия “хи-квадрат” мы установили схожесть между балтийскими языками (латышским и литовским) с одной стороны, а также древнерусским, русским и некоторыми другими славянскими языками, с другой стороны.

Многие лингвисты и этнографы [С.Б. Берштейн, С. В. Бромлей, В.Н. Топоров и другие] отмечают, что в прошлом существовал длительный и тесный контакт славян с древними балтами, что дает основание предположить существование общего балто-славянского праязыка [Бромлей, 1985; БСЭ, 1976: 552 – 553; Топоров, 1966: 455]. Исследователи балтийских (литовского и

латышского) языков также отмечают их близость к славянским языкам. К. Ульвидас, В. Амбразас, В. Вайткявичюте, А. Валяцкене, К. Моркунас, А. Сабалаяускас и другие отмечают, что фонетическая, морфологическая, лексическая и синтаксическая структура литовского языка больше всего похожа на структуру славянских языков. Это может происходить от того, что балтийские и славянские языки имели общий балто-славянский праязык, как результат былого балто-славянского единства [Амбразас и др., 1966: 500–501]. Целью нашей работы является проверка балто-славянского единства на фонологическом уровне.

Было протранскрибировано большое количество текстов. Далее, на компьютере была вычислена схожесть этих лингвистических объектов с помощью методов математической статистики и распознавания образов. Анализируется типология строения звуковых цепочек в языках и их диалектах в качестве образа [Тамбовцев 1994а; Тамбовцев 1994: б]. На основании доминантных фонетических признаков звуковых картин построены фоно-типологические расстояния между лингвистическими объектами. Сравняются между собою только звуковые картины лингвистических объектов, хотя в дальнейшем эти же методы можно применить как на лексическом, так и синтаксическом уровне. Под обобщенным термином «лингвистический объект» подразумевается язык или диалект. Мы вводим понятие «лингвистический объект» именно потому, что часто разделить диалект и язык невозможно. Эдельман Д.И. отмечает, что формальные признаки отделения диалекта от языка не выработаны [Эдельман 1980].

Для построения расстояний используются значения критерия «хи-квадрат». На его базе вычисляется коэффициент ТМВ (Тамбовцева), детальное описание вычисления которого дается ниже. Самым главным для наших исследований является то, что коэффициент ТМВ служит показателем расстояния между лингвистическими объектами [Тамбовцев 2010б]. В научной литературе по классификации языков, взятые нами лингвистические объекты зарегистрированы по-разному: то как отдельные языки, то как диалекты какого-либо языка. С учетом порогового значения коэффициента ТМВ эти лингвистические объекты классифицируются на диалекты и отдельные языки уже по новому, т.е. на основании величины показателей типологических расстояний по фонетическим признакам. Лексические, синтаксические и семантические характеристики в данном случае не учитываются. Таким образом, построенная нами модель языка является чисто фонетической.

Заметим, что один язык может выступать как один лингвистический объект, если мы берем его в литературном варианте. Язык может выступать и как несколько лингвистических объектов, т.е. как класс объектов, если мы берем несколько его диалектов. Класс объектов может называться таксоном или образом. Следовательно, диалект – это единичный объект, а язык может быть классом объектов, т.е. образом или таксоном, если он имеет целый набор диалектов.

Введение

Языкознание в настоящее время находится на таком этапе развития, когда без применения точных количественных методов исследования продвижение вперед невозможно. Числовые характеристики хорошо отражают лингвистические закономерности [Tambovtsev, 2006]. На компьютере нами были обработаны большие массивы текстов, что позволило получить надежные

статистические результаты в численном виде по звуковым картинам различных языков. На основе применения методов математической лингвистики получена возможность решения лингвистической задачи как близости различных языков, так и их диалектов, которая не могла быть решена в старых рамках классического языкознания.

Язык и диалект в данном случае трактуются как фонетические объекты. Таким образом, язык (как и любой другой объект) можно интерпретировать в терминах распознавания образов, где объекты понимаются как отдельные представители образов в пространстве признаков [Загоруйко 1972: 7]. В связи с тем, что выбранные нами фонетические признаки имеют количественное выражение, мы можем сказать, что к изучению языков в данном случае применена самая сильная, т.е. абсолютная шкала. Точная нумерическая информация позволяет измерять точные расстояния как между отдельными объектами, так и между таксонами [Загоруйко 1999: 61]. Под таксонами мы будем понимать подгруппы, группы, семьи или общности (супер-семьи) языков. В то же время, таксоном может считаться один язык, если он имеет несколько диалектов. В том случае, когда мы имеем какой-либо лингвистический объект (т.е. язык или диалект), который мы должны поместить в какой-либо лингвистический таксон (т.е. подгруппу, группу, семью и т.д.), мы можем это сделать на основе типологического расстояния, измеренного по выбранным признакам. По величине расстояния между лингвистическими объектами можно определить величину функции принадлежности объекта к образу [Загоруйко 1999: 167-169]. Если расстояние от лингвистического объекта до некоторого образа (таксона) меньше, чем до других образов, то можно считать, что объект принадлежит к данному образу. Велико или невелико расстояние можно судить по величине коэффициента ТМВ [Tambovtsev 2006; 2007]. Вычислить этот коэффициент достаточно просто [Тамбовцев 2003].

В данной работе осуществлен типологический подход, в основе которого лежит выделение каких-либо признаков систем лингвистических объектов и их дальнейшая группировка с помощью обобщенной модели на основе этих выделенных признаков. Таким образом, лингвистический объект уже представлен не целостно, а в виде некоторого количества признаков, которые имеют нумерические значения. Этот подход обычно применяется в распознавании образов, но мы также используем его и в языкознании, где он дает интересные лингвистические результаты [Tambovtsev 2009b]. Мы можем назвать этот подход численно-типологическим. Следует отметить, что применение типологического подхода на основе определенных выбранных признаков, но без нумерического значения признаков, осуществлялось в языкознании и ранее. Наиболее последовательным представителем типологического подхода в языкознании является В.Д. Аракин. Таким образом, типологический подход позволяет сопоставлять не только генетически родственные языки (например, русский, украинский и белорусский), но и достаточно далекие языки, например, русский и английский [Аракин 2000: 5-8]. Концепция типологического подхода в языкознании позволяет сравнивать любые языки мира с разными структурами не только на фонетическом, но и на других уровнях языка. В то же время важность сопоставления различных языков в типологическом плане на фонетическом уровне подчеркивается в современном языкознании, например, [Tambovtsev, 2007, 2008, 2009]. Метод фундаментального типологического сопоставления позволил В.Д. Аракину сравнивать между собой такие разные

языки как славянские, германские, тюркские, малайско-полинезийские и другие языки с разным генетическим происхождением и разных систем и структур [Аракин, 2000].

Еще раз отметим, что в общем виде, язык можно определять как лингвистический объект. При определении степени самостоятельности лингвистических объектов, т.е. языков и их диалектов, мы используем фоностатистический критерий близости звуковых картин. Этот же фонотипологический метод может быть использован при измерении близости звуковых картин любых других протранскрибированных текстов [Тамбовцев 2009]. Под звуковой картиной языка мы подразумеваем совокупность звуковых цепочек анализируемых языковых объектов. Исследование проводилось на материале самодийских, финно-угорских, монгольских, индо-европейских и тюркских языков. С целью сопоставления привлекались данные, полученные нами ранее при изучении языков некоторых других семей.

Проблемой, насколько один язык отличается от другого, интересовались еще в древности. Уже тогда на арену истории вышли различные народы с различными языками. Проблему языка и диалекта поставили еще древнегреческие философы, которые пытались решить ее в общем, т.е. как проблему целого и частного. Известно, что частное заключает в себе характеристики целого. Действительно, можно ли диалект считать частью определенного языка или же этот языковой объект следует считать самостоятельным языком? Иными словами, служат ли отличия, которые характеризуют диалект, подтверждением общих тенденций функционирования языка и его диалекта? В связи с тем, что любой язык достаточно вариативен, это является трудной задачей. Диалект от языка может отделять зона неопределенности, которая будет тем шире, чем больше их схожесть. Являются ли близкородственные языки действительно разными языками или же это диалекты, которые называют отдельными языками? Это касается, например, таких близкородственных языков как русский, белорусский и украинский или немецкий и голландский.

Материал исследования.

Материалом исследования послужили тексты на различных славянских и балтийских языках, которые были представлены в фонетической транскрипции. Детальное описание материала по всем языкам, которые были взяты для фонологических и статистических исследований приведено в монографиях Ю.А. Тамбовцева [Тамбовцев, 2001а; 2001б; 2001в].

По латышскому языку были использованы фоно-статистические данные З.М. Байкова, которые мы пересчитали по 9 фонетическим признакам. Мы сравниваем их в деталях с аналогичными фоностатистическими данными Е.Ш. Пиель, А. Лоренц, З. Несауле, З.М. В. А. Кузиной и Т.А. Якубайтис [Тамбовцев, 2001а: 65-66].

Литовский язык представляют пересчитанные по 9 фонетическим признакам данные А. Гирденеса [Girdenis, 1981]. Мы сравниваем их в деталях с результатами фоностатистического подсчета Б.И. Свецевичуса [Тамбовцев, 2001: 67-69].

В качестве материала по древнерусскому языку послужили некоторые тексты из Христоматии по истории русского литературного языка [М.:Высшая школа, 1974, которую составил для студентов университетов и педагогических институтов А.Н. Кожин. Памятники древнерусской

письменности 10-13 веков включали следующие тексты: 1) Русская правда; 2) Повесть временных лет; 3) Киевская летопись; и 4) Поучение Владимира Мономаха. Эти древнерусские тексты сравнивались по частоте встречаемости в их звуковых цепочках выделенных фонетических признаков.

Сколько и каких языков конкретно было привлечено для исследования в данной работе и их объемы в фонемах показано ниже. Кроме того, Таб.1 показывает частоту встречаемости 9 групп фонем в потоке речи на этих языках. Выборки фонем по лингвистическим объектам в данном исследовании имеют следующий объем:

Балтийские языки: латышский (ЛТШ) — 25410; литовский (ЛИТ) — 100000.

Славянские языки:

Восточно-славянские: древнерусский (ДРС) — 68234; русский (РУС) — 20189475; белорусский (БЛР) — 37487; украинский (УКР) — 300000;

Западно-славянские: польский (ПОЛ) — 104603; чешский (ЧЕШ) — 186641; словацкий (СЛО) — 20000; сербско-лужицкий (ЛУЖ) — 93110.

Южно-славянские: болгарский (БОЛ) — 18240; сербско-хорватский (СРБ) — 45486; македонский (МАК) — 62795; словенский (СЛН) — 108363.

Методика исследования

Расстояние между лингвистическими объектами измеряется при помощи величины коэффициента ТМВ, вычисленного на основе критерия хи-квадрат. Известно, что критерий «хи-квадрат» используется в статистических исследованиях для сравнения распределений объектов двух совокупностей на основе измерений в числовом виде признаков в двух независимых выборках. Данный статистический критерий широко известен. В рамках данной статьи невозможно, да и нет необходимости вдаваться в тонкости математических подсчетов с помощью критерия «хи-квадрат». Его описание можно найти в любых учебниках по математической статистике или математической лингвистике [напр. Пиотровский и др., 1977]. Кроме того, они даются в компьютерных приложениях в виде таких программ, как «Статистика», «Математика», «XL» и другие. Критерий хи-квадрат применялся нами ранее для определения типологической близости языков различных семей [Тамбовцев, 2003].

Важность применения критерия хи-квадрат для лингвистических исследований заключается в том, что он позволяет исключить субъективизм суждений путем точного сопоставления любых лингвистических объектов, у которых признаки выражены в числовых величинах. Например, сравнение числовых значений признаков звуковых картин венгерского, хантыйского и мансийского языков подтвердило правомерность классификации венгерского языка в качестве члена угорской группы финно-угорской семьи. В то же время, оно не позволило отнести его к подгруппе обско-угорских языков, куда традиционно относятся хантыйский и мансийский язык. Наш метод позволил выделить его в отдельную, венгерскую, подгруппу, которая до сих пор еще не выделялась внутри угорской группы [Tambovtsev 2009: 123].

При трактовке полученных данных следует учитывать наличие обратной зависимости между близостью языков и фонетическими расстояниями между ними: схожесть языков тем больше, чем меньше расстояния, и наоборот [Tambovtsev, 2008]. Таким образом, величина «хи-квадрат» показывает несхожесть языков и диалектов. Мы воспользовались методом сравнения двух выборок, которые могут иметь разный объем [Тамбовцев, 2003].

В качестве основополагающего критерия мы предложили ввести понятие «лингвистическая компактность» [Тамбовцев, 2002]. Его сущность заключается в том, что два типологически близких друг другу лингвистических объекта образуют один лингвистический таксон, если они близки друг другу. Это покажет типологическое расстояние [Tambovtsev 2006]. Чем меньше расстояние, тем более похожи эти объекты. В тоже время, чем меньше сумма расстояний в каком-либо таксоне, тем более похожи лингвистические объекты в этом таксоне. Под таксоном, как мы уже указывали выше, мы имеем в виду любую группу языков, даже такую, которая состоит только из двух членов, т.е. любая пара языков или диалектов [Тамбовцев 2002: 131]. Так, можно говорить о таксоне мансийского языка, представленного сосьвинским и кондинским диалектами. Карельский язык состоит из трех диалектов, татарский язык представлен также тремя диалектами. Языки и диалекты выступают в виде лингвистических объектов и составляют финно-угорский таксон. Алтайский язык имеет пять диалектов и в совокупности с другими тюркскими языками составляет отдельный таксон. В этом же смысле можно говорить о таксоне балтийских языков или таксоне славянских языков.

Рассчитывая величину критерия «хи-квадрат», мы должны выбрать уровень статистической значимости и количество степеней свободы. Уровень статистической значимости и количество степеней свободы может быть разным. В данном случае, такая мера сходства как «хи-квадрат» может чаще употребляться в лингвистических исследованиях, если мы облегчим для классического лингвиста понимание процедуры его расчета. Под мерой сходства мы подразумеваем величину расстояния между двумя лингвистическими объектами.

Для упрощения понимания величины расстояния нами был введен коэффициент ТМВ, который учитывает количество степеней свободы и уровень статистической значимости. Фактически, коэффициент ТМВ – это отношение эмпирически полученной величины критерия «хи-квадрат» к его теоретической к величине с учетом степеней свободы и уровня значимости. Нами выбран уровень значимости 0,05. Мы имеем 9 признаков, что дает 8 степеней свободы. Теоретическое значение «хи-квадрат» при 8 степенях свободы и уровне значимости 0,05 составляет 15,507. Если количество признаков будет другое, то даже при этом же уровне значимости, теоретическое значение критерия «хи-квадрат» будет другое.

Выбор признаков

В типологических исследованиях важно определить базовые признаки [Тамбовцев, 2003], на которых строится наше сравнение. В данном случае было выбрано 9 доминантных фонетических признаков, к которым мы относим показатели частоты встречаемости восьми групп согласных в звуковых цепочках языка: губных, переднеязычных, среднеязычных, заднеязычных, сонорных, шумных смычных, шумных щелевых, шумных звонких, – а также гласных [Тамбовцев 2001в, 2003].

Фактически данные фонетические признаки являются необходимыми и достаточными для распознавания звуковых картин различных языков. Обычно в распознавании образов признаки должны быть проверены с точки зрения их необходимости и достаточности [Загоруйко 1972: 55]. Нам нет необходимости делать такую проверку в связи с тем, что мы взяли все базовые артикуляционные признаки согласных, которые выделяются в фонетике. Гласные берутся без разделения на группы потому, что данные 9 признаков хорошо различают различные лингвистические объекты на фонетическом уровне. Если два языковых объекта не различаются по 9 признакам, то можно ввести дополнительные фонетические признаки. Мы согласны с Н.Г. Загоруйко, что нужно по возможности минимизировать количество признаков, так как это уменьшает размерность пространства, что, в свою очередь, уменьшает затраты усилий при распознавании объектов [Загоруйко 1972, 54-57; Загоруйко 1999: 102-108].

Под звуковой цепочкой имеется в виду последовательность звуков, т.е. слова, которые в совокупности составляют текст. В языкознании отмечается, что в диалектах звуковые цепочки похожи. Более того, они похожи и в генетически родственных языках. Метод поиска генетической близости языков основан на анализе материального сходства, т.е. сходства звуковых цепочек на уровне слов. Эти материальные сходства приводятся в любом учебнике по введению в языкознание, напр. [Баранникова 1973: 276-279] Действительно, в родственных языках слова имеют сходные звуковые цепочки: русский – «мать»; украинский – «мати»; белорусский – «маці»; болгарский – «майка»; сербский – «мати»; чешский – «матер»; польский – «мац». В то же время, в тюркских языках слово «мать» имеет уже другую звуковую цепочку: азербайджанский – «ана»; башкирский – «ана»; казахский «ана»; киргизский – «энэ»; кумандинский – «анам»; татарский – «ана»; тувинский – «ава»; туркменский – «ана»; узбекский – «она».

Таб. 1

Балтийские языки: ЛТШ – латышский; ЛИТ – литовский.

Славянские языки:

Восточно-славянские: ДРС – древнерусский; РУС – русский; БЛР – белорусский; УКР – украинский;

Западно-славянские: ПОЛ – польский; ЧЕШ – чешский; СЛО – словацкий; ЛУЖ – сербско-лужицкий.

Южно-славянские: БОЛ – болгарский; СРБ – сербско-хорватский; МАК – македонский; СЛН – словенский.

	Группы фонем	ЛТШ	ЛИТ	ДРС	РУС	БЛР	УКР	ПОЛ	ЧЕШ	СЛО	ЛУЖ
1.	Губные	10,86	13,63	13,19	12,63	14,45	13,01	16,66	13,57	12,79	14,83
2.	Переднеязычн	35,89	33,37	33,30	35,18	33,84	34,60	27,42	32,16	30,03	32,12
3.	Среднеязычны	3,07	3,59	2,96	4,27	4,63	3,91	9,31	6,20	7,21	6,47
4.	Заднеязычные	5,55	7,39	5,98	5,74	7,12	6,28	5,75	6,87	6,39	6,68
5.	Сонорные	19,48	19,28	19,44	23,07	20,70	21,42	21,10	21,39	20,25	27,12
6.	Смычные шумные	20,78	22,25	20,17	20,14	21,91	21,18	21,16	20,47	20,82	18,79
7.	Щелевые шумные	15,11	16,45	15,82	14,61	17,43	15,20	16,88	16,94	15,35	14,19
8.	Звонкие шумные	11,10	12,63	11,72	11,69	12,53	15,19	13,90	13,28	9,69	11,51
9.	Сумма гласных	44,63	42,02	44,57	42,18	39,96	42,20	40,86	41,20	43,58	39,90

	Группы фонем	БОЛ	СРБ	МАК	СЛН	
1.	Губные	12,91	11,96	11,67	12,54	
2.	Переднеязычные	30,24	31,75	36,95	35,53	
3.	Среднеязычные	1,45	5,53	0,89	4,40	
4.	Заднеязычные	6,68	5,63	6,20	6,28	
5.	Сонорные	19,54	25,06	18,68	24,04	
6.	Смычные шумные	14,96	20,37	24,12	21,08	
7.	Щелевые шумные	16,78	9,44	12,91	13,63	
8.	Звонкие шумные	12,51	9,89	13,27	13,72	
9.	Сумма гласных	48,72	45,13	44,29	41,25	

В соответствии с целью данной работы мы пытаемся установить близость между языками на фонетическом уровне. Именно поэтому так важно получить надежные данные по частоте встречаемости различных групп фонем в звуковой цепочке языка. Расчет фоно-типологических расстояний по величине критерия “хи-квадрат” между балтийскими языками (латышским и литовским) с одной стороны, а также древнерусским, русским и некоторыми другими славянскими языками, с другой стороны, фактически покажут эту близость. Более того, эти

расстояния исключают субъективизм исследователя. Они показывают, насколько звуковая картина одного языка объективно похожа на звуковую картину другого. Под звуковой картиной языка мы имеем в виду сумму звуковых цепочек языка, которая отражает общую частоту групп фонем в потоке речи. Таким образом, данная работа выполнена в плане лингвистической фонологометрии, или лингвометрии, т.е. науки об измерении межъязыковых расстояний на основе звуковой картины языка [Тамбовцев, 2002].

Применяется объективный типологический метод, который использует информацию о структуре частоты встречаемости фонологических единиц. Известно, что язык имеет две стороны: фонетическое содержание, т.е. материальную оболочку, и семантическое содержание. Мы базируем наше исследование на материальной стороне языка, которая существует в языке вне нас и независимо от нас. В связи с этим, фонетическое содержание хорошо поддается изучению статистическими методами. Это дает нам основание полагать, что мы получаем объективные данные потому, что эти данные могут быть получены любым другим независимым экспертом, кто возьмет данные языки и применит данный статистический метод и выделенные фонетические признаки. Взятые нами фонетические признаки присущи всем языкам мира, поэтому они хорошо показывают место латышского и литовского языка среди других языков древней балтославянской общности. Эти фонетические признаки можно назвать универсальными.

Следует обратить внимание на проблему соизмеримости, которая требует единообразного описания лингвистических единиц. В нашей работе это достигается тем, что мы берем классификацию, т.е. разбиение на группы, с точки зрения работы артикуляционного аппарата человека. Кроме того, получаемые статистические данные должны быть соизмеримы, что достигается определенным способом нормирования через вычисление процентов частоты встречаемости фонетических элементов во всей цепочке языка. Прежде, чем привлекать методы математической лингвистики, например, критерий “хи-квадрат”, можно показать, даже простые данные по частоте встречаемости фонем в звуковых цепочках разных языков, могут дать важную лингвистическую информацию. Конечно, перед тем как сравнивать эти частоты необходимо их пронормировать, чтобы получить соизмеримость. В данном случае – это переводение частоты встречаемости в проценты.

В качестве таких признаков использовалась частота встречаемости групп согласных, которые выделены по работе активного органа артикуляции, характера артикуляционной преграды и работе голосовых связок. Таким образом, сравнение было полным и системным с фонетической точки зрения.

Перед тем, как сравнивать схожесть языков по критерию “хи-квадрат”, приведем упорядоченные ряды пронормированных частот наших 9 признаков по взятым языкам. Эти упорядоченные ряды дают важную лингвистическую информацию потому, что они показывают соседние языки, в которых схожая концентрация определенных фонемных групп в звуковой цепочке. В то же время, следует указать, что шкала порядка дает не такое точное представление о схожести языка, как абсолютная шкала. Именно поэтому мы далее применяем анализ по самой сильной шкале – абсолютной шкале.

Упорядоченный ряд по возрастанию концентрации **губных** согласных в балтийских и славянских языках: 1) латышский – 10,86%; 2) македонский – 11,67%; 3) сербско-хорватский – 11,96%; 4) словенский – 12,54%; 5) русский – 12,63%; 6) словацкий – 12,79%; 7) болгарский – 12,91%; 8) украинский – 13,01%; 9) древнерусский – 13,19%; 10) чешский – 13,57%; 11) литовский – 13,63%; 12) белорусский – 14,45%; 13) сербо-лужицкий – 14,83%; 14) польский – 16,66%.

Данный упорядоченный ряд (шкала порядка) показывает, что по употреблению губных согласных латышский язык больше похож на македонский, а литовский – на чешский и белорусский. Польский выходит из ряда вон, так как концентрация губных у него наивысшая.

Упорядоченный ряд (шкала порядка) по возрастанию концентрации **переднеязычных** согласных в балтийских и славянских языках: 1) польский – 27,42%; 2) словацкий – 30,03%; 3) болгарский – 30,24%; 4) сербско-хорватский – 31,75%; 5) сербо-лужицкий – 32,12%; 6) чешский – 32,16%; 7) древнерусский – 33,30%; 8) литовский – 33,37%; 9) белорусский – 33,84%; 10) украинский – 34,60%; 11) русский – 35,18%; 12) словенский – 35,53%; 13) латышский – 35,89%; 14) македонский – 36,95%.

Данный упорядоченный ряд показывает, что по употреблению переднеязычных согласных латышский язык больше похож на словенский, а литовский – на древнерусский. Македонский выходит из ряда вон, так как концентрация переднеязычных у него наивысшая. Опять польский язык выходит из ряда вон, но уже с наименьшей концентрацией переднеязычных в потоке речи.

Упорядоченный ряд (шкала порядка) по возрастанию концентрации **среднеязычных** согласных в балтийских и славянских языках: 1) македонский – 0,89%; 2) болгарский – 1,45%; 3) древнерусский – 2,96%; 4) латышский – 3,07%; 5) литовский – 3,59%; 6) украинский – 3,91%; 7) русский – 4,27%; 8) словенский – 4,40%; 9) белорусский – 4,63%; 10) сербско-хорватский – 5,53%; 11) чешский – 6,20%; 12) сербо-лужицкий – 6,47%; 13) словацкий – 7,21%; 14) польский – 9,31%.

Данный упорядоченный ряд показывает, что по употреблению среднеязычных согласных латышский язык больше похож на древнерусский, а литовский – на латышский. Македонский выходит из ряда вон, так как концентрация среднеязычных у него наименьшая. Опять польский язык выходит из ряда вон, но уже с наибольшей концентрацией среднеязычных в потоке речи.

Упорядоченный ряд (шкала порядка) по возрастанию концентрации **заднеязычных** согласных в балтийских и славянских языках: 1) латышский – 5,55%; 2) сербско-хорватский – 5,63%; 3) русский – 5,74%; 4) польский – 5,75%; 5) древнерусский – 5,98%; 6) македонский – 6,20%; 7) словенский – 6,28%; 8) украинский – 6,28%; 9) словацкий – 6,39%; 10) сербо-лужицкий – 6,68%; 11) болгарский – 6,68%; 12) чешский – 6,87%; 13) белорусский – 7,12%; 14) литовский – 7,39%.

Данный упорядоченный ряд показывает, что по употреблению заднеязычных согласных латышский язык больше похож на сербско-хорватский, а литовский – на белорусский. Латышский стоит на краю ряда, так как концентрация заднеязычных у него наименьшая. На противоположной стороне ряда оказался литовский язык с наибольшей концентрацией заднеязычных в потоке речи.

Упорядоченный ряд (шкала порядка) по возрастанию концентрации **сонорных** согласных в балтийских и славянских языках: 1) македонский – 18,68%; 2) литовский – 19,28%; 3) древнерусский

– 19,44%; 4) латышский – 19,48%; 5) болгарский – 19,54%; 6) словацкий – 20,25%; 7) белорусский – 20,70%; 8) польский – 21,10%; 9) чешский – 21,39%; 10) украинский – 21,42%; 11) русский – 23,07%; 12) словенский – 24,04%; 13) сербско-хорватский – 25,06%; 14) сербо-лужицкий – 27,12%.

Упорядоченный ряд (шкала порядка) по возрастанию концентрации **смычных** согласных в балтийских и славянских языках: 1) болгарский – 14,96%; 2) сербо-лужицкий – 18,79%; 3) русский – 20,14%; 4) древнерусский – 20,17%; 5) сербско-хорватский – 20,37%; 6) чешский – 20,47%; 7) словацкий – 20,82%; 8) латышский – 20,78%; 9) словенский – 21,08%; 10) украинский – 21,18%; 11) польский – 21,16%; 12) белорусский – 21,91%; 13) литовский – 22,25%; 14) македонский – 24,12%.

Упорядоченный ряд (шкала порядка) по возрастанию концентрации **целевых** согласных в балтийских и славянских языках: 1) сербско-хорватский – 9,44%; 2) македонский – 12,91%; 3) словенский – 13,63%; 4) сербо-лужицкий – 14,19%; 5) русский – 14,61%; 6) латышский – 15,11%; 7) украинский – 15,20%; 8) словацкий – 15,35%; 9) древнерусский – 15,82%; 10) литовский – 16,45%; 11) болгарский – 16,78%; 12) польский – 16,88%; 13) чешский – 16,94%; 14) белорусский – 17,43%.

Упорядоченный ряд (шкала порядка) по возрастанию концентрации **шумных звонких** согласных в балтийских и славянских языках: 1) словацкий – 9,69%; 2) сербско-хорватский – 9,89%; 3) латышский – 11,10%; 4) сербо-лужицкий – 11,51%; 5) русский – 11,69%; 6) древнерусский – 11,72%; 7) болгарский – 12,51%; 8) белорусский – 12,53%; 9) литовский – 12,63%; 10) македонский – 13,27%; 11) чешский – 13,28%; 12) словенский – 13,72%; 13) польский – 13,90%; 14) украинский – 15,19%.

Упорядоченный ряд (шкала порядка) по возрастанию концентрации **гласных** в балтийских и славянских языках: 1) сербо-лужицкий – 39,90%; 2) белорусский – 39,96%; 3) польский – 40,86%; 4) чешский – 41,20%; 5) словенский – 41,25%; 6) литовский – 42,02%; 7) русский – 42,18%; 8) украинский – 42,20%; 9) словацкий – 43,58%; 10) македонский – 44,29%; 11) древнерусский – 44,57%; 12) латышский – 44,63%; 13) сербско-хорватский – 45,13%; 14) болгарский – 48,72%.

Литовский язык по употреблению гласных ближе всего к русскому языку, а латышский – к древнерусскому. Наименьшая концентрация гласных в сербо-лужицком языке. Наибольший процент встречаемости гласных в болгарском языке.

Напомним, что для более точного анализа величины расстояния между языками использовался критерий “хи-квадрат”, который не только учитывает порядок языков в упорядоченном ряду, но и дает точное расстояние по каждому признаку. Свойства аддитивности этого критерия позволяет нам выводить суммарное расстояние по взятым 9 фонетическим признакам. Кроме того, нами был введен коэффициент ТМВ, который упрощает лингвистический анализ. Этот коэффициент учитывает количество степеней свободы и уровень статистической значимости, что позволяет сравнивать выборки с разным количеством признаков. Фактически, коэффициент ТМВ – это отношение эмпирически полученной величины критерия “хи-квадрат” к его теоретической к величине, в данном случае, с учетом 8 степеней свободы и уровня значимости 0,05.

Фоно-типологическое расстояние по 9 признакам между литовским и латышским языком составило 6,45. Мы должны понять, много это или мало. Как говорили древние философы, все

познается в сравнении. Для этого мы должны сравнить расстояние, которое мы вычислили (6,45) с аналогичными расстояниями между другими исследованными языками.

Фоно-типологическое расстояние между литовским и латышским намного больше, чем между литовским и древнерусским. Литовский язык по его звуковой картине более похож на современный русский (6,07), чем на латышский.

В свою очередь латышский язык по звучанию еще больше похож на древнерусский язык (2,47), чем литовский. Латышский язык также ближе по его звучанию к современному русскому (3,65), чем литовский.

Были получен следующий упорядоченный ряд по расстояниям (ТМВ) между **литовским** и некоторыми славянскими языками: 1) белорусский – 1,92; 2) древнерусский – 2,84; 3) украинский – 3,64; 4) русский – 6,07; 5) чешский – 6,14; 6) латышский – 6,45; 7) словенский – 7,46; 8) словацкий – 12,99; 9) македонский – 17,11; 10) сербо-лужицкий – 18,22; 11) болгарский – 19,64; 12) польский – 24,62; 13) сербско-хорватский 25,66.

Таким образом, с учетом всех фонетических характеристик ближе всего к литовскому оказался белорусский язык. Менее всего литовский язык похож на сербско-хорватский.

Был получен следующий упорядоченный ряд по расстояниям (ТМВ) между **латышским** и некоторыми славянскими языками: 1) древнерусский – 2,47; 2) русский – 3,65; 3) литовский – 6,45; 4) украинский – 7,50; 5) словенский – 8,83; 6) белорусский – 10,68; 7) македонский – 12,34; 8) чешский – 14,38; 9) болгарский – 15,89; 10) словацкий – 16,31; 11) сербско-хорватский – 19,97; 12) сербо-лужицкий – 24,46; 13) польский – 39,66.

Прежде всего следует отметить большую схожесть латышского и древнерусского языка, что может говорить о былой балто-славянской группе в индо-европейской языковой семье. Достаточно близок к латышскому и современный русский язык. Менее всего похож по звучанию на латышский – польский язык.

Между **древнерусским** и другими языками были получены следующие расстояния по величине коэффициента ТМВ, которые составили упорядоченный ряд по возрастанию: 1) литовский – 2,84; 2) латышский – 3,65; 3) русский – 4,71; 4) украинский – 5,20; 5) белорусский – 6,42; 6) словенский – 8,60; 7) чешский – 10,29; 8) болгарский – 11,08; 9) македонский – 13,92; 10) словацкий – 14,20; 11) сербско-хорватский – 15,31; 12) сербо-лужицкий – 20,16; 13) польский – 30,54.

Из этих расстояний можно сделать вывод, что по звуковым картинам к древнерусскому ближе всего литовский и латышский, но не русский, украинский или белорусский. Меньше всего на звуковую картину древнерусского языка похожи звуковые картины сербо-лужицкого, греческого и сербско-хорватского языков. Близость литовского языка к белорусскому можно объяснить не только былым балто-славянским единством, но и интенсивными контактами между ними в составе единого литовско-белорусского государства.

СПИСОК ЛИТЕРАТУРЫ

-
- ✚ Амбразас В., Вайткявичюте В., Валяцкене А., Моркунас К., Сабалаяускас А., Ульвидас К., 1966, *Литовский язык. Языки народов СССР. Индо-европейские языки*, том 1., Москва, Наука, с. 500-527.
- ✚ Аракин В. Д., 2000, *Сравнительная типология английского и русского языков*, Москва, сс 256.
- ✚ БСЭ – Большая Советская Энциклопедия, 1976, том 23., Москва.
- ✚ Бромлей С. В., 1985б *Восточнославянские языки как объект лингвогеографии*, [in:] *Восточные славяне: языки, история, культура*, Москва, с. 172–179.
- ✚ Баранникова Л.И., 1973, *Введение в языкознание*, Саратов: Сар. ГУ, сс 384.
- ✚ Загоруйко Н.Г., 1972, *Методы распознавания и их применение*, Москва.
- ✚ Загоруйко Н.Г., 1999, *Прикладные методы анализа данных и знаний*, Новосибирск.
- ✚ Пиотровский Р.Г., Бектаев К.Б., Пиотровская А.А., 1977, *Математическая лингвистика*, Москва, Высшая школа, сс. 383.
- ✚ Тамбовцев Ю.А., 1994а, *Динамика функционирования фонем в звуковых цепочках языков различного строя*, Новосибирск, НГУ, сс 136.
- ✚ Тамбовцев Ю. А., 1994б, *Типология упорядоченности звуковых цепей в языке*, Новоисбирск, НГУ, сс 199.
- ✚ Тамбовцев Ю.А., 2001а, *Компендиум основных статистических характеристик функционирования согласных фонем в звуковой цепочке английского, немецкого, французского и других индоевропейских языков*. Новосибирск, Новосибирский Классический Институт сс 129.
- ✚ Тамбовцев Ю. А., 2001б, *Функционирование согласных фонем в звуковой цепочке урало-алтайских языков*, Новосибирск: Новосибирский классический институт, сс 132.
- ✚ Тамбовцев Ю.А., 2001, *Некоторые теоретические положения типологии упорядоченности фонем в звуковой цепочке языка и компендиум статистических характеристик основных групп согласных фонем*, Новосибирск, Новосибирский классический институт, сс 130.
- ✚ Тамбовцев Ю.А., 2002, *Лингвистическая таксономия: компактность языковых подгрупп, групп и семей*, „Baltistica” XXXII (1), Рига, с 131–161.
- ✚ Тамбовцев Ю.А., 2003, *Типология функционирования фонем в звуковой цепочке индоевропейских, палеоазиатских, урало-алтайских и других языков мира: компактность подгрупп, групп, семей и других языковых таксонов*, Новосибирск, Сибирский Независимый институт, сс 143.
- ✚ Тамбовцев Ю.А., 2010а, *Синергетика звуковых цепочек языка и типологическая близость лингвистических объектов*, „Синергетика образования” №1. (17), с. 48–56.
- ✚ Тамбовцев Ю.А., 2010б, *Фоно-типологическая близость по критерию «хи-квадрат»*, „Вестник Новосибирского государственного университета” Серия: Информационные технологии, том 8, вып.3, с. 46-54.
-

- ✚ Топоров В.Н., 1966, *Балтийские языки. Языки народов СССР. Индо-европейские языки*, том 1. Москва.
- ✚ Girdenis A., 1981, *Siaures Zemaiciu fonemu daznumai*, [in:] „Kalbotyra”, XXXII(1), p. 15-37.
- ✚ Tambovtsev Yuri, 2006, *Degrees of Dispersion of Various Language Taxa as the Index of Their Typological Similarity*, [in:] *Favete Linguis. Studies in Honour of Viktor Krupa*, Bratislava, Institute of Oriental Studies of Slovak Academy of Sciences, p. 219 – 260.
- ✚ Tambovtsev Yuri, 2007, *How can typological distances between Latin and some Indo-European language taxa improve its classification?* [in:] “The Prague Bulletin of Mathematical Linguistics”, Number 88, December, p. 73 – 90.
- ✚ Tambovtsev Yuri, 2008, *The phono-typological distances between Ainu and other world languages as a clue for closeness of languages*, “Asian and African Studies” Vol. 17, № 1. (Bratislava), p. 40–62.
- ✚ Tambovtsev Yuri, 2009, *How should the Hungarian language be correctly classified within the Ugric language taxon?*, “Asian and African Studies”, Vol. 18. № 1., Bratislava., p. 103–137.

TAMBOVTSEV, Yuri & TAMBOVTSEVA, Juliana & TAMBOVTSEVA, Ludmila: The Balto-Slavonic Unity: Reality or Fiction?

There were and still are some heated discussions about the reality or fiction of the Balto-Slavonic unity. The application of “Chi-square” criterion allows us to measure the distances between languages objectively. The phono-typological distance between Baltic and Slavonic languages was measured. The phono-typological distance between Lithuanian and Latvian languages is much greater than between Lithuanian and Old Russian. The sound picture of Lithuanian is more similar to modern Russian (6,07) than to Latvian. In its turn, Latvian's sound picture is more similar to Old Russian (2,47) than Lithuanian. Latvian is more similar to modern Russian (3,65) than Lithuanian by its sound picture. The closeness of Baltic and Slavonic languages may be explained by the former language unity and by intensive language contacts.

Key words: typology, closeness of languages, sound picture, phono-typological picture, “Chi-square” criterion.

