

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ
ФЕДЕРАЦИИ»

Перестановочный тест говорит в пользу гипотезы алтайской языковой
макросемьи

Научный доклад (препринт)
Лаборатории востоковедения и компаративистики
по материалам отчета НИР
«Составление лексических списков для языков Евразии и Океании: генетическое
родство языков в междисциплинарных исследованиях»

А. С. Касьян

kasyan-als@ranepa.ru

И.М.Егоров

egorov-ilm@ranepa.ru

А. А. Трофимов

atrofimov@ranepa.ru

Москва 2020

Аннотация

В настоящем препринте применяется метод статистической обработки лексических списков, разработанный авторами: транскрипция консонантных классов и взвешенный перестановочный тест. В эксперименте участвуют пять языковых семей, предположительно образующих алтайскую макросемью: тюркская, монгольская, тунгусская, корейская, японская. Выявлен сильный сигнал между так ядерными алтайскими таксонами: тюркским, монгольским, тунгусским, что может объясняться как генеалогическое родство, частично затемненное контактами. Выявлен сигнал родства в японско-тюркской и японско-тунгусской парах (контактный сценарий для этих пар подходит плохо).

Список исполнителей:

Руководитель НИР, Старший
научный сотрудник, Научно-
исследовательская лаборатория
востоковедения и
компаративистики Школа
актуальных гуманитарных
исследований ИОН, д.ф.н.

А.С. Касьян

подпись, дата

Научный сотрудник, Научно-
исследовательская лаборатория
востоковедения и
компаративистики Школа
актуальных гуманитарных
исследований, ИОН

И.М. Егоров

подпись, дата

Старший научный сотрудник,
Научно-исследовательская
лаборатория востоковедения и
компаративистики, Школа
актуальных гуманитарных
исследований, ИОН, к.ф.н.

А.А. Трофимов

подпись, дата

Содержание

1 Введение	5
2 Материалы и методы	6
3 Результаты	10
4 Обсуждение	10
5 Выводы.....	12
ЛИТЕРАТУРА	12

1 Введение

Так называемая алтайская гипотеза касается нескольких консенсусно выделяемых языковых семей, на которых говорят в Азии и частично в Европе, это тюркские, монгольские, тунгусские, корейские и японские семьи. Гипотеза имеет две версии: узкую и широкую. Узкая версия утверждает, что тюрки, монголы и тунгусы генеалогически связаны друг с другом и образуют отдельную кладу. Широкая версия расширяет эту кладу с корейским и японским языками. Получившуюся кладу традиционно называют алтайской макросемьей. Некоторые авторы используют ярлык “алтайская (макро)семья” только для клады [тюркский, монгольский, тунгусский], тогда как это же плюс корейский и японский называют макро-алтайской или (недавно введенный термин) трансевразийской макросемьей. Очень подробный и обширный обзор алтайской гипотезы, ее истории и современного состояния предлагается в [1], недавний том [2] прекрасно дополняет его. Пока единственная систематическая реконструкция протоалтайской (в широком смысле) фонологии и лексики предложена в работе [3].

Все или почти все эксперты сходятся во мнении, что лексические сходства между ядерными алтайскими таксонами — тюркскими, монгольскими, тунгусскими — слишком многочисленны, чтобы быть случайными, но предлагаемые исторические сценарии различаются: одни ученые утверждают генеалогическое родство, другие приписывают наблюдаемые лексические соответствия (доисторическим) контактам. То же самое, хотя и в меньшей степени, касается связей между корейским и японским языками и дальнейшего их родства с тюрко-монголо-тунгусским.

Авторы настоящей статьи придерживаются гибридного сценария, предполагающего генеалогическое родство тюркских, монгольских, тунгусских, корейских и японских языков, которое впоследствии перекрывалось и частично затемнялось различными контактами большей или меньшей интенсивности и продолжительности. Природа алтайского родства, однако, находится за пределами настоящей статьи.

Цель настоящего проекта — оценить статистическую поддержку алтайской гипотезы с помощью алгоритма, который бы не был абстрактной математической процедурой, оторванной от реальности, а наоборот как можно точнее моделировал бы работу традиционного лингвиста-компаративиста. Мы считаем, что близость автоматического алгоритма к реальному научному подходу и методологии должна быть решающим критерием при выборе из широкого спектра доступных вычислительных методов, по крайней мере в лингвистике.

2 Материалы и методы

Мы проанализировали пять 110-словных список Сводеша для пяти ветвей гипотетической алтайской макросемьи: тюркский, монгольский, тунгусский, корейский, японский. См. [4] о принципах составления сводешевских списков и семантических спецификаций концептов.

Из этих пяти списков четыре реконструированы: прото-тюркский, прото-монгольский, прото-тунгусский, прото-японский. Пятый список слов отражает синхронную базовую лексику среднекорейского языка (из-за несовершенства описаний корейских диалектов и скудости данных доханкульского периода трудно реконструировать протокорейский список Сводеша, но среднекорейский список слов является разумным приближением).

Мы используем стандартные фонологические и морфологические методы реконструкции для данных протоязыков. Мы придерживались принципов семантической (семасиологической) реконструкции, предложенных в работе [5]: 305-306. Это: топологический принцип, принцип внешней этимологии, принцип внутренней этимологии, принцип семантической правдоподобности, принцип исключения ареального эффекта.

Мы закодировали все задействованные словоформы в транскрипции консонантных классов, сведя их к двуконсонантному виду *СС*. Процедура транскрипции полностью аналогична ранее описанной для индоевропейско-уральского сравнения [5]: 307-309. Используются следующие классы согласных: Таблица 1.

Таблица 1. Консонантные классы

P-class (губные)	p b β f v...
T-class (зубные)	t d ɗ θ ð ɬ ...
S-class (сибилянты фриктивы)	s z š ž...
ʒ-class (сибилянты аффрикаты)	c ʒ č ž...
Y-class (палатальные глайды)	y...
W-class (губные глайды)	w м...
M-class (губные носовые)	m ɱ...
N-class (негубные носовые)	n ɳ ɲ...
Q-class (латеральные аффрикаты)	ɭ ...
R-class	ɾ ɽ...

L-class	l ʎ ʎʎ...
K-class (велярные и увулярные)	k g x ɣ q χ ʁ...
zero-class или H-class	h ʕ н ʕ ʔ h ʕ ʔ и все гласные

Взвешенный перестановочный тест был применен к каждой паре в пяти задействованных списках слов, всего десять пар. Алгоритм проверки помечает пару форм с идентичной СС-транскрипцией как совпадение (СС-match), а все другие пары - как несовпадающие. Таким образом, метод консонантных классов является, с одной стороны, грубой вариацией измерения расстояний Левенштейна, а с другой стороны, близок к моделированию реального сравнительно-исторического исследования, по крайней мере в том, что касается критериев выявления возможных этимологических лексических соответствий между двумя языками. Взвешенный критерий перестановки был недавно предложен нами в работе [6]. Его отличие от стандартного (невзвешенного) теста (описанного, например, в работе [5]: 309-310) заключается в том, что учитывается типологическая устойчивость отдельных сводешевских концептов: совпадение в более устойчивом концепте трактуется как более дорогое, чем совпадение в менее устойчивом концепте.

Наш алгоритм выявил следующие потенциальные этимоны.

Тюркский-монгольский, $p = 0.000071$

1. black: kara - kara
2. dry: ku:r - kahuray
3. I: bi - bi
4. kill: öl - ala
5. long: urʏ - urtu
6. man: e:r - ere
7. seed: ur - hüre
8. that: ti - te
9. we: bi - ba
10. yellow: sia:rig - sira
11. worm: ku:rt - kora

Тюркский-тунгусский, $p = 0.000899$

1. I: bi - bi
2. liver: biagır - pa:ki
3. many: ö:k - egdi
4. sleep: u: - a:w
5. that: ti - ta
6. thou: si - si
7. we: bi - bu:

Тюркский-корейский, $p = 0.574548$

1. bark: ka:p - kəp
2. that: ti - tyé

Тюркский-японский, $p = 0.000131$

1. bark: ka:p - kapa
2. black: kara - kùruà
3. I: bi - bà
4. sleep: u: - úi
5. this: kö - kó
6. we: bi - bà
7. what: ne: - nà

Монгольский-тунгусский, $p < 0.000001$

1. egg: emdüge - umu:
2. green: nogoha - ɲog
3. I: bi - bi
4. mouth: ama - amɲa
5. see: üže - iče
6. stone: čila - žolo
7. that: te - ta
8. this: e - e
9. tongue: kele - xilɲü
10. we: ba - bu:
11. snake: mogay - mü:ki
12. thin: nim - nem

13. year: hon - аᠢᠨᠭᠠ

Монгольский-корейский, $p = 0.082684$

1. I: na - nà
2. leaf: nabčī - ní᠑^h
3. that: te - tyó
4. this: e - í

Монгольский-японский, $p = 0.072733$

1. black: kara - kùruà
2. I: bi - bà
3. we: ba - bà

Тунгусский-корейский, $p = 0.080664$

1. heart: miawan - т᠋ᠠᠷᠠᠨ
2. that: ta - tyó
3. this: e - í
4. who: ᠡᠭᠡᠨ - nú

Тунгусский-японский, $p = 0.004814$

1. burn tr.: deg - dák
2. I: bi - bà
3. knee: pe᠑ - pínsá
4. sleep: a:w - úi
5. water: mu: - mí
6. we: bu: - bà
7. near: daga - tìkà

Корейский-японский, $p = 0.015218$

1. bark: kə᠑ - kapa
2. good: tyǒ:h - dè
3. many: mǎ:n - mana
4. not: àní - an
5. swim: həy - əyè

6. that: kì - ká
7. thou: nè - ná
8. snake: p'lyám - p'àimV

3 Результаты

Вероятностные результаты парного сравнения пяти алтайских списков Сводеша представлены в таблице 1. Мы применили поправку Холма-Бонферрони [7] с $\alpha = 0,05$ для решения проблемы множественного тестирования.

Таблица 1 - Вероятности фонетических совпадений между алтайскими языками, полученные с помощью взвешенного перестановочного теста. Статистически значимые значения затенены ($\alpha = 0,05$).

	Mongolic	Tungusic	Korean	Japonic
Turkic	7.1×10^{-5}	9×10^{-4}	0.575	1.3×10^{-4}
Mongolic	-	$< 10^{-6}$	0.083	0.073
Tungusic		-	0.081	0.005
Korean			-	0.015

Пары между ядерными алтайскими таксонами — тюркский, монгольский, тунгусский — имеют либо высоко значимые ($\alpha=0,001$), либо очень значимые ($\alpha=0,01$) p -значения.

Японский язык демонстрирует более слабый результат. японско-тюркский: очень значимое p ($\alpha=0,01$). Японско-тунгусский: значимое p ($\alpha=0,05$). Японско-монгольский: не значимое p .

4 Обсуждение

Может ли генеалогическая связь между языками быть доказана с помощью таких статистических методов, как перестановочный тест? Строгий ответ - нет, поскольку в случае обнаруживаемой генеалогической связи следует ожидать не только значительного числа фонологических совпадений по классам согласных в базисном словаре данных языков, но и регулярности, т. е. рекурсивности этих звуковых соответствий.

Статистически значимое значение p , полученное такими методами, является скорее эвристическим признаком того, что данный язык может быть связан друг с другом либо генеалогически, либо через интенсивные контакты.

Доказывает ли незначимое p -значение неродственность данных языков? Определенно, это не так. Хотя мы ожидаем, что большинство наблюдаемых звуковых соответствий попадет в классы согласных, некоторые соответствия, которые часто встречаются в данной паре языков, могут быть слишком далеки для наших классов согласных (например, $\check{s}$ / s или g / d). Особенно этот эффект имеет решающее значение в ситуации множественных сравнений, когда применяемая коррекция множественного тестирования резко снижает уровень статистической значимости (например, корейско-японская пара имеет $0,01 \leq p < 0,05$, но этот результат становится не значимым в более широком контексте алтайского сравнения).

Ядерные алтайские таксоны — тюркские, монгольские, тунгусские — демонстрируют сильный сигнал родства и/или контактов. Природа этого сигнала выходит за рамки настоящей статьи, но, вкратце, мы полагаем, что большинство совпадений между прототюркским, протомонгольским и прототунгусским базисным словарем является унаследованными, т.е. восходят к единому праязыку, хотя действительно некоторые совпадения, особенно в тюркско-монгольской паре, могут быть объяснены как древние заимствования. Низкие p -значения ядерных алтайских таксонов неудивительны. Двусторонние лексические сходства между тюркским, монгольским, тунгусским видны невооруженным глазом и давно привлекают внимание лингвистов.

Более важным результатом являются (очень) значимые p -значения для японско-тюркской и японско-тунгусской пар. Если хотя бы некоторые лексические сходства между протояпонским и прототунгусским языками можно отнести к гипотетическим доисторическим контактам, то для протояпонского и прототюркского языков такой сценарий маловероятен из-за их географической удаленности. Так что в случае японско-тюркской пары мы имеем дело либо с серией случайных созвучий, либо с генеалогическим родством.

Корейский язык не показывает положительных результатов ни с одним из своих потенциальных алтайских родственников. P -значение для корейско-японской пары само по себе значимо ($0,01 \leq p < 0,05$), но его недостаточно для прохождения поправки на множественный эксперимент. Корейско-японский результат несколько удивителен, поскольку значительное число (если не большинство) экспертов в этой области признают корейско-японскую связь, будь она генеалогическая или контактная ([1]: 48-54). Однако общий отрицательный результат корейского языка не является неожиданным, поскольку сторонники алтайской гипотезы или, по крайней мере, корейско-японской генеалогической связи (например, [3; 8; 9]) вынуждены предполагать различные процессы падения неначальных согласных в пракорейском языке, с одной стороны, и необъяснимое начальное $*s$ - в некоторых корейских корнях (например, *srua* 'кость'), с другой. Оба фактора влияют на

СС-транскрипцию корейских форм и, таким образом, влияют на результаты перестановочного теста.

5 Выводы

Подводя итог, можно сказать, что четыре таксона — прототюркский, протомонгольский, прототунгусский, протояпонский — демонстрируют статистическую поддержку генеалогического и/или древних контактов с различным уровнем значимости. В некоторых случаях географическое распределение позволяет делать предположения о доисторических контактах и обмене заимствованиями, но в таких парах, как тюркско-тунгусский и тюркско-японский, сценарий контакта, естественно, маловероятен. Одним из возможных объяснений статистически незначимых результатов корейского языка является предполагаемое массовое падение неначальных согласных в пракорейском.

ЛИТЕРАТУРА

1. Blažek V. Altaic languages: history of research, survey, classification and a sketch of comparative grammar. Altaic languages / V. Blažek. – Brno: Masaryk University Press, 2019.
2. The Oxford guide to the Transeurasian languages / eds. M. Robbeets, A. Savelyev. – New York: Oxford University Press, 2020.
3. Starostin S.A. Etymological dictionary of the Altaic languages : Handbook of Oriental studies = Handbuch der Orientalistik. Section eight, Central Asia : в 3 vols. EDAL / S.A. Starostin, A.V. Dybo, O.A. Mudrak. – Leiden: Brill, 2003. – 8. – Vol. 1-3.
4. The Swadesh wordlist. An attempt at semantic specification. / A.S. Kassian [et al.] // Journal of Language Relationship. – 2010. – Vol. 4. – P. 46–89.
5. Kassian A.S. Proto-Indo-European-Uralic comparison from the probabilistic point of view / A.S. Kassian, G.S. Starostin, M.A. Zhivlov // Journal of Indo-European Studies. – 2015. – Vol. 43. – № 3-4. – P. 301-347.
6. Circumpolar peoples and their languages: lexical and genomic data suggest ancient Chukotko-Kamchatkan–Nivkh and Yeniseian–Burushaski–Na-Dene connections / P. Changmai [et al.]. – forthcoming.
7. Holm S. A simple sequentially rejective multiple test procedure / S. Holm // Scandinavian Journal of Statistics. – 1979. – Vol. 6. – № 2. – P. 65-70.
8. Martin S.E. Lexical evidence relating Korean to Japanese / S.E. Martin // Language. – 1966. – Vol. 42. – № 2. – P. 185.

9. Robbeets M.I. Is Japanese related to Korean, Tungusic, Mongolic and Turkic? : *Turcologica* / M.I. Robbeets. – Wiesbaden: Harrassowitz, 2005. – 64. – 975 p.