

Е. В. Коровина

(Институт языкознания РАН, Москва)

РАНЖИРОВАНИЕ БАЗИСНОЙ ЛЕКСИКИ В ТРУДАХ С. А. СТАРОСТИНА: МАТЕРИАЛЫ К УЛУЧШЕНИЮ¹

В статье рассматривается методика для получения ранжированного списка базисной лексики С. А. Старостина. Показывается, что в некоторых случаях она может быть улучшена. Так, во-первых, кажется, что в случае неполноты данных более точный результат дает деление не на общее число языков, а на число языков, где это слово представлено. Во-вторых, при собственно ранжировании присвоение разных рангов словам с одинаковой устойчивостью в рамках одной семьи при небольшом числе языков в семье может существенно изменить порядок ранжирования.

Ключевые слова: сравнительно-историческое языкознание, базисная лексика, ранжирование, лексикостатистика.

E. V. Korovina

(Institute of Linguistics, RAS, Moscow)

Method of obtaining a ranked list of the basic vocabulary developed by Sergei A. Starostin

The article discusses the method of obtaining a ranked list of the basic vocabulary elaborated by Sergei A. Starostin. Unfortunately, as it seems, the original procedure cannot be completely reproduced. This is mainly due to the fact that some languages do not preserve the original data in the form suggested in Starostin's work. In particular, the Austro-Asiatic base has undergone significant changes. However, the reconstruction of Starostin's procedure shows that in some cases it can be improved. So, firstly, it seems that a more accurate result is given not by the total number of languages, but by the number of languages where a given word is present, since otherwise, if the word is not attested in all the languages, significant distortions are possible. E.g. in the Austronesian database, the word *warm* has a very low stability index because it was attested only in 5 out of 94 languages. Secondly, if one follows the actual ranking method, assigning different ranks to words with the same stability within the same family, in case the latter comprises a small number of languages, can significantly change the ranking order. It seems that when using ranks it is better to assign the same ranks – in one way or another – to such words.

Keywords: historical linguistics, basic lexicon, ranked lists, lexicostatistics.

¹ Работа выполнена при поддержке гранта РНФ № 18-18-00501 «Создание электронного диалектологического атласа тюркских языков России» (рук. А. В. Дыбо).

0. Ранжирование базисной лексики

Одним из классических вопросов, который встает перед специалистами по лексикостатистике, является вопрос о том, «равны» ли слова в пределах используемого для лексикостатистических подсчетов списка, действительно ли вероятность замены одного слова в пределах такого списка совпадает с вероятностью замены любого другого слова в нем, а также насколько универсальны подобного рода вероятности. Первые наблюдения подобного рода были, по-видимому, сделаны самим М. Сводешем (Swadesh 1955), который для целей повышения точности вычислений отобрал из 215 слов своего списка (Swadesh 1952) 92 наиболее «устойчивые» единицы. Его метод был основан на сравнении пар «предок-потомок» для разных языков, то есть на непосредственном наблюдении за сохранностью слова, однако подобного рода наблюдения доступны лишь для очень ограниченного набора языковых групп. Поэтому возникает вопрос, что делать, если предковое состояние недоступно для наблюдения. Основным методом решения этой проблемы в работах компаративистов является интерпретация «устойчивости» к заменам в ходе развития языка через тем или иным способом посчитанную дистрибуцию корней, представленных в языках-потомках (например, Thomas 1960, Dolgorolsky 1964 и др.). В редких случаях для этой цели используют и другие параметры, например, морфологическую непроецируемость или склонность к заимствованию (ср. Tadmor 2009). Наиболее крупной работой подобного рода на сегодняшний день является посмертно опубликованная статья С. А. Старостина (Starostin 2007), где определение «индекса стабильности» производилось на материале 14 языковых групп и 737 языков. Этот результат обсуждается, в частности, в работе Г. С. Старостина (Starostin 2010), где ранжирование С. А. Старостина предлагается использовать для получения более короткого 50-словного списка, служащего в первую очередь для первичной демонстрации родства, что, например, сделано в (Kassian, Starostin, Zhivlov 2015). Однако как замечает К. И. Поздняков (Pozdniakov 2014), предлагающий на тех же данных несколько иной способ ранжирования, использованный С. А. Старостиным «алгоритм не является прозрачным», поэтому экспликация, а до некоторой степени и реконструкция использованного им алгоритма, представляется необходимой.

1. Реконструкция метода С. А. Старостина

Материалом для реконструкции служит как сам текст статьи (Kassian, Starostin, Zhivlov 2015), так и лексикостатистические данные, предоставленные автору А. В. Дыбо, М. А. Живловым и И. И. Пейросом. К сожалению, удалось обнаружить только 4 базы данных (австралийскую, австронезийскую, енисейскую, северокавказскую), еще для 8 групп были найдены базы, очевидно, с тем же составом и примерно того же времени, хотя в некоторых деталях и не до конца тождественные. Найденная база дравидийских языков содержит 29 языков, но для ранжирования может быть установлен только 21 язык. В случае австроазиатской базы, наиболее близкая из найденных баз состоит из 110 языков, но установить, какие 99 из них были использованы при ранжировании, не представляется возможным.

Для определения устойчивости базисной лексики С. А. Старостин применяет двухуровневую процедуру. Вначале для каждого слова в каждой из 14 семей (синотибетской, австроазиатской, алтайской, австронезийской, австралийской, койсанской, северокавказской, дравидийской, индоевропейской, картвельской, афразийской, тайской, уральской и енисейской) он вычисляет «индекс стабильности» (ИС) – «отношение максимального числа языков, использующих для данного значения один и тот же корень, к общему числу языков в семье» (Starostin 2007: 827), а затем на основании полученных данных устанавливалось место того или иного слова в иерархии устойчивости.

Первая часть данной процедуры представляется значительно более прозрачной по сравнению с собственно процедурой ранжирования. Однако, как кажется, требуется сделать ряд уточнений. Во-первых, под «общим числом языков в семье» в данном случае понимается общее количество языков, представленное в той или иной лексикостатистической базе, при этом в базе могут иметься как действительно все или почти все языки семьи (как, например, в картвельской или енисейской), так и существенно меньшая по объему выборка (как в случае, например, австралийской, австронезийской или афразийской семьи). Во-вторых, в случае если по каким-либо причинам слово было зафиксировано не во всех языках базы, его ИС может быть занижен по сравнению со словом, зафиксированным во всех языках. Так, например, крайне низкий ИС слова «теплый» в австронезийских языках обусловлен, как

кажется, именно этой причиной – в базе данных этот вход заполнен лишь для 5 языков из 94, и очевидно, что, даже если бы эти 5 слов были бы рефлексами одного праязыкового слова, ИС в принципе не мог бы быть высоким.

Вторая часть процедуры, а именно переход от ИС слова в каждой семье к собственно рангу слова, более комплексна. С. А. Старостин так описывает ее: сначала ранжируются «все значения (от 1 до 110) по индексу стабильности в каждой из рассмотренных семей», а затем «из значений, занимающих позицию n (начиная с $n = 1$ и кончая $n = 11(0)$)» выбирают «такие, которые занимают эту или более высокую ($n-1$) позицию как минимум в двух семьях. При этом при одинаковом ранге (позиции n) “удельный вес” выбранного значения» увеличивается «в соответствии с числом семей, в котором он эту позицию занимает» (Starostin 2007: 833, 838). Проблема здесь состоит главным образом в том, что неясно, как следует приписывать ранги в случае совпадения ИС у двух и более слов к той или иной семье, что является нередким случаем – так в енисейских языках 28 слов имеют наивысший ИС. Также неясно, что делать, если полученный «удельный вес» одинаков для двух и более слов.

Так, можно видеть из таблицы (Starostin 2007: 830–833), что слова «глаз» и «я» имеют наивысший из возможных ИС в 5 семьях, а слово «мы» и «два» только в 4. Соответственно гипотеза о том, что при совпадении ИС словам каждой семьи присваивался наивысший из возможных рангов – в данном случае 1 – является неверной т. к. в этом случае наиболее стабильным словом не могло бы стать «мы». Как представляется, наиболее простым способом объяснить полученный результат является предположение, что словам с одинаковым ИС присваивался разный ранг в зависимости от их следования в списках ранжированных значений (обычно соответствующего обратному порядку следования английских эквивалентов слов списка) (Starostin 2007: 833–837). В этом случае единственным словом, которое стоит на первом месте в двух списках, является «мы» (афразийский и уральский), слова же «глаз», «я» и «два» начинают список только в одной из семей. При переходе к рангу 2 слово «два» занимает первую или вторую позицию в 4 семьях, поэтому оказывается вторым по устойчивости и т.д. Как представляется, если данное наблюдение верно, то это ставит результат ранжирования в зависимость от довольно произвольного порядка слов в списках. Представляется, что лучшим

решением могло бы быть при ранжировании присваивать словам с одинаковым ИС одинаковые ранги, например, наименьший из возможных.

2. Вычисление ИС для отдельных семей с учетом неполноты данных

Ниже представлен результат вычисления ИС стабильности для отдельных семей при использовании исходных материалов С. А. Старостина или максимально близких к ним и несколько измененной методики вычисления. Так, деление производилось не на общее количество языков в базе данных, а на количество языков, для которых данная позиция была заполнена. В случае, если позиция была заполнена менее чем в половине языков в базе ставился прочерк. Обоснованием для применения именно такой методики служит то, что как было показано выше, во многих случаях списки базисной лексики были заполнены не полностью, при этом очевидно, что лакуны в данных занижают ИС в тех случаях, где они присутствуют и могут исказить картину стабильности. Так, в случае енисейской базы, например, слово «кровь» не было засвидетельствовано в пумпокольском языке, а в четырех остальных языках зафиксированные слова восходили к одному слову. При подсчетах без учета количества лакун ИС слова «кровь» в енисейских равен 0,8, как в случае если бы отсутствующее слово восходило бы к другому корню. Как кажется, в случае же учета только присутствующих значений данные получаются более точными, аналогично тому, что в стандартной процедуре лексикостатистики процент сходства между двумя списками лексики вычисляется для слов, зафиксированных в обоих списках (в случае применения методики С. А. Старостина также исключаются заимствования), а не от общей потенциальной длины списка. Следует отметить, что существуют и другие методики работы с неполными данными.

Таблица 1.

Индекс стабильности для отдельных семей

	YEN	IE	AF	AA*	Alt	AUS	NC	AN	UR	TH	ST	KS	KT	DR
1	all	0,67	0,21	0,09	0,44	0,13	0,17	-	0,18	0,3	0,14	0,67	0,57	0,41
2	ashes	0,33	0,12	0,2	0,36	0,17	0,42	0,7	0,24	0,89	0,25	0,57	0,71	0,35
3	bark	0,67	0,3	0,12	0,13	0,7	0,16	0,41	0,41	0,41	0,16	0,36	0,25	0,21
4	belly	0,5	0,19	0,13	0,15	0,2	0,19	0,36	0,18	0,5	0,22	0,52	0,4	0,43
5	big	0,6	0,16	0,1	0,17	0,11	0,24	0,2	0,29	0,34	0,24	0,24	0,57	0,67
6	bird	0,75	0,29	0,13	0,78	0,23	0,15	0,49	0,24	0,91	0,34	0,4	0,86	0,29
7	bite	-	0,15	0,11	0,45	0,31	0,72	0,39	0,41	0,45	0,15	0,48	0,5	0,47
8	black	1	0,25	0,09	0,13	0,69	0,25	0,35	0,29	0,85	0,35	0,64	0,43	0,74
9	blood	1	0,18	0,28	0,92	0,36	0,23	0,45	0,47	0,96	0,9	0,56	1	0,68
10	bone	0,67	0,41	0,22	0,85	0,48	0,58	0,43	0,82	0,89	0,62	0,43	0,71	0,38
11	breast	0,6	0,22	0,16	0,52	0,36	0,32	0,67	0,29	0,23	0,5	0,48	0,86	0,22
12	burn tr.	0,67	0,24	0,08	0,15	0,63	0,34	0,19	0,29	0,24	0,15	0,46	0,57	0,44
13	claw(nail)	0,67	0,95	0,21	0,4	0,38	0,19	0,76	0,94	0,85	0,69	0,75	1	0,84
14	cloud	1	0,25	0,12	0,2	0,37	0,21	0,24	0,53	0,7	0,14	0,26	0,43	0,43
15	cold	0,4	0,34	0,18	0,17	0,45	0,17	0,3	0,35	0,32	0,24	0,28	0,33	0,57
16	come	0,67	0,49	0,19	0,14	0,62	0,31	0,21	0,48	0,68	0,27	0,5	0,67	1
17	die	1	0,76	0,49	0,42	0,43	0,15	0,58	0,76	0,98	0,92	0,5	0,33	0,76
18	dog	1	0,45	0,17	0,88	0,62	0,23	0,82	0,45	0,98	0,9	0,38	1	0,8
19	drink	0,4	0,6	0,35	0,15	0,43	0,26	0,21	0,59	0,49	0,39	0,92	0,67	0,71
20	dry	1	0,62	0,15	0,12	0,67	0,32	0,97	0,24	0,35	0,23	0,61	0,5	0,35
21	ear	0,8	0,44	0,25	0,39	0,54	0,28	0,61	0,71	0,65	0,79	0,42	0,57	0,81
22	earth	1	0,35	0,13	0,84	0,37	0,34	0,21	0,33	0,55	0,3	0,38	0,43	0,33
23	eat	0,8	0,36	0,22	0,55	0,53	0,26	0,42	0,83	0,91	0,67	0,5	1	0,81
24	egg	0,8	0,59	0,15	0,21	0,52	0,17	0,58	0,53	0,89	0,29	0,41	0,43	0,37
25	eye	1	0,71	0,3	0,83	0,69	0,42	0,93	1	0,96	0,93	0,36	0,57	1
26	fat n.	1	0,12	0,1	0,17	0,45	0,24	0,15	0,24	0,51	0,48	0,52	0,5	0,79

	YEN	IE	AF	AA*	Alt	AUS	NC	AN	UR	TH	ST	KS	KT	DR
27	feather	1	0,3	0,19	0,67	0,34	0,12	0,09	0,51	0,58	0,23	0,5	0,43	0,17
28	fire	0,6	0,29	0,1	0,61	0,33	0,19	1	0,8	0,71	0,89	0,73	0,57	0,65
29	fish	0,6	0,28	0,11	0,75	0,35	0,22	0,28	0,36	0,88	0,67	0,75	0,43	0,75
30	fly v.	0,67	0,16	0,23	0,74	0,42	0,33	0,13	0,25	0,35	0,36	0,25	1	0,47
31	foot	1	0,62	0,17	0,89	0,37	0,36	0,3	0,12	0,29	0,25	0,28	0,71	0,6
32	full	0,67	0,82	0,21	0,5	0,49	0,23	1	0,46	0,5	0,79	0,28	1	0,8
33	give	0,75	0,76	0,16	0,13	0,57	0,72	0,67	0,41	0,47	0,87	0,62	0,3	0,62
34	good	1	0,13	0,1	0,2	0,28	0,14	0,26	0,1	0,24	0,83	0,12	0,5	0,37
35	green	0,25	0,32	0,11	0,08	0,34	-	0,19	0,2	0,38	0,81	0,09	0,5	0,74
36	hair	1	0,24	0,13	0,83	0,58	0,15	0,3	0,41	0,71	0,94	0,36	0,88	0,21
37	hand	0,5	0,53	0,19	0,91	0,31	0,78	0,36	0,5	0,65	0,55	0,48	0,6	0,95
38	head	0,8	0,43	0,19	0,17	0,5	0,28	0,42	0,59	0,29	0,4	0,33	0,42	0,79
39	hear	1	0,4	0,2	0,18	0,35	0,26	0,3	0,68	0,59	0,64	0,21	0,42	0,62
40	heart	0,6	0,76	0,26	0,25	0,52	0,34	1	0,42	0,88	0,85	0,27	0,48	0,64
41	horn	1	0,67	0,23	0,29	0,35	-	0,21	0,26	0,47	0,57	0,35	0,79	0,6
42	I	1	0,71	0,76	0,35	0,82	1	0,94	0,85	1	0,81	0,77	0,54	1
43	kill	0,67	0,38	0,13	0,43	0,43	0,37	0,42	0,57	0,41	0,88	0,63	0,61	0,28
44	knee	0,75	0,53	0,19	0,35	0,34	0,43	0,21	0,42	0,62	0,7	0,26	0,52	0,45
45	know	0,67	0,88	0,22	0,11	0,44	0,29	0,36	0,24	0,29	0,87	0,49	0,48	0,45
46	leaf	1	0,22	0,11	0,87	0,61	0,17	0,18	0,55	0,29	0,67	0,31	0,52	0,45
47	lie	-	0,31	0,08	0,25	0,44	0,42	0,25	0,16	0,29	0,91	0,24	0,48	0,2
48	liver	0,5	0,26	0,17	0,72	0,49	0,28	0,5	0,81	0,88	0,57	0,66	0,77	0,61
49	long	1	0,66	0,12	0,28	0,6	0,23	0,82	0,13	0,29	0,91	0,28	0,39	0,5
50	louse	0,75	0,4	0,12	0,88	0,39	0,31	0,85	0,96	0,59	0,11	0,62	0,35	0,9
51	man	1	0,34	0,09	0,18	0,35	0,19	0,39	0,26	0,18	0,57	0,2	0,46	0,27
52	many	0,67	0,19	0,13	0,12	0,32	0,15	0,12	0,08	0,29	0,57	0,13	0,37	0,38
53	meat	0,8	0,42	0,23	0,35	0,33	0,23	0,73	0,15	0,24	0,77	0,58	0,46	0,19
54	moon	0,6	0,49	0,13	0,5	0,37	0,22	0,94	0,8	0,47	0,4	0,83	0,5	0,58

	YEN	IE	AF	AA*	Alt	AUS	NC	AN	UR	TH	ST	KS	KT	DR
55	mountain	0,8	0,19	0,12	0,14	0,38	0,27	0,16	0,29	0,49	0,21	0,47	0,43	0,29
56	mouth	0,6	0,25	0,36	0,11	0,35	0,18	0,2	0,59	0,14	0,15	0,48	0,8	0,57
57	name	1	0,96	0,66	0,22	0,36	0,94	0,54	1	0,72	0,97	0,91	1	1
58	neck	0,5	0,24	0,13	0,28	0,64	0,15	0,17	0,24	0,91	0,26	0,67	0,43	0,32
59	new	0,67	0,92	0,14	0,48	0,67	0,21	0,75	0,53	0,51	0,43	0,25	0,33	0,95
60	night	1	0,48	0,16	0,24	0,3	0,19	0,38	0,4	0,59	0,87	0,2	0,36	0,43
61	nose	0,6	0,74	0,32	0,91	0,39	0,45	0,78	0,24	0,34	0,75	0,46	0,57	0,76
62	not	1	0,88	0,25	0,15	0,41	0,38	0,31	0,12	0,94	0,55	0,79	0,4	0,75
63	one	1	0,92	0,25	0,85	0,54	0,11	0,94	0,82	0,65	0,74	0,37	0,5	0,67
64	person	1	0,3	0,12	0,16	0,26	0,22	0,51	0,18	1	0,54	0,52	0,25	0,35
65	rain	1	0,34	0,19	0,5	0,53	0,36	0,72	0,35	0,79	0,48	0,5	0,57	0,35
66	red	1	0,47	0,19	0,16	0,36	0,17	0,3	0,22	0,24	0,55	0,2	0,41	0,56
67	road	0,6	0,28	0,13	0,25	0,4	0,17	0,61	0,79	0,68	0,53	0,65	0,57	0,35
68	root	0,6	0,31	0,26	0,89	0,2	0,15	0,24	0,44	0,38	0,2	0,29	0,43	0,63
69	round	0,33	0,2	0,07	0,25	0,47	-	0,56	0,12	0,04	0,18	-	0,8	0,55
70	sand	0,8	0,15	0,09	0,29	0,33	0,14	0,09	0,19	0,26	0,38	0,43	0,29	0,58
71	say	1	0,14	0,19	0,06	0,34	0,4	0,36	0,06	0,62	0,17	0,45	0,5	0,81
72	see	0,5	0,21	0,09	0,12	0,43	0,67	0,42	0,34	0,67	0,15	0,5	0,43	0,45
73	seed	-	0,4	0,11	0,24	0,4	0,21	0,34	0,29	0,83	0,34	0,4	1	0,9
74	sit	0,6	0,73	0,12	0,14	0,4	0,64	0,38	0,11	0,41	0,85	0,18	0,48	0,26
75	skin	-	0,35	0,17	0,22	0,35	0,18	0,15	0,46	0,83	0,18	0,52	0,8	0,89
76	sleep	1	0,59	0,09	0,23	0,34	0,37	0,45	0,28	0,32	0,53	0,43	0,5	0,33
77	small	0,6	0,2	0,09	0,19	0,3	0,12	0,16	0,13	0,29	0,83	0,27	0,3	0,43
78	smoke	1	0,71	0,13	0,17	0,36	0,37	0,44	0,5	0,29	0,64	0,78	0,65	0,74
79	stand	0,4	0,75	0,13	0,31	0,6	0,56	0,52	0,22	0,35	0,81	0,39	0,4	0,95
80	star	1	0,77	0,18	0,36	0,85	0,14	0,58	0,62	0,83	0,57	0,52	0,4	0,65
81	stone	1	0,28	0,13	0,54	0,94	0,11	0,18	0,89	0,5	0,89	0,68	0,32	0,68
82	sun	1	0,75	0,12	0,52	0,37	0,17	0,79	0,42	0,37	0,54	0,5	0,86	0,57
83	swim	1	0,24	0,08	0,34	0,24	0,15	0,25	0,45	0,83	0,16	0,4	0,8	0,56

	YEN	IE	AF	AA*	Alt	AUS	NC	AN	UR	TH	ST	KS	KT	DR
84	tail	1	0,33	0,25	0,72	0,24	0,39	0,46	0,29	0,55	0,82	0,52	1	0,28
85	that	0,33	0,34	0,34	0,1	0,43	0,32	-	0,56	0,89	0,17	0,2	0,4	1
86	this	0,67	0,36	0,4	0,24	0,53	0,38	-	0,79	0,66	0,31	0,5	0,4	0,95
87	thou	1	0,98	0,45	0,71	0,97	0,82	0,84	0,82	0,85	0,88	0,92	0,71	1
88	tongue	0,6	0,94	0,43	0,45	0,67	1	0,63	0,65	0,66	0,93	0,88	1	0,57
89	tooth	1	0,77	0,36	0,32	0,4	0,88	0,51	0,41	0,91	0,55	0,72	0,57	0,95
90	tree	0,8	0,24	0,1	0,32	0,28	0,12	0,73	0,71	0,79	0,71	0,46	0,43	0,9
91	two	1	1	0,24	0,93	0,37	0,91	0,99	1	0,79	0,92	0,62	1	1
92	walk (go)	0,75	0,43	0,1	0,12	0,24	0,24	0,19	0,29	0,57	0,22	0,5	0,6	0,33
93	warm	0,8	0,46	0,08	0,18	0,5	-	-	0,18	1	0,45	0,22	1	0,6
94	water	1	0,37	0,26	0,72	0,52	0,61	0,45	0,88	0,33	0,33	0,44	0,57	0,57
95	we	1	0,67	0,85	0,29	1	0,91	0,8	1	0,4	0,61	0,85	0,8	0,81
96	what	1	0,98	0,67	0,15	0,7	0,43	0,28	0,81	0,53	0,31	0,25	0,71	1
97	white	1	0,32	0,13	0,28	0,32	0,29	0,18	0,32	0,39	0,15	0,4	0,67	0,79
98	who	0,67	1	0,43	0,13	0,58	0,56	0,28	1	0,28	0,42	0,29	0,57	1
99	woman	0,5	0,44	0,15	0,17	0,2	0,11	0,27	0,44	0,66	0,48	0,42	0,33	0,28
100	yellow	0,75	0,28	0,19	0,11	0,59	-	0,31	0,3	0,2	0,2	-	0,4	0,33
101	far	1	0,46		0,71	0,21	0,39		0,29		0,21	-	0,4	0,45
102	heavy	1	0,33		0,27	0,35	0,62		0,24		0,52	-	0,33	0,45
103	near	0,67	0,19		0,21	0,45	0,27		0,24		0,48	-	0,43	0,54
104	salt	1	0,49		0,29	0,46	0,55		0,35		0,61	-	0,71	0,57
105	short	0,67	0,4		0,23	0,33	0,29		0,29		0,36	-	0,57	-
106	snake	0,5	0,25		0,39	0,39	0,36		0,38		0,51	-	1	0,52
107	thin	1	0,49		0,25	0,4	0,69		0,25		0,69	-	1	-
108	wind	1	0,79		0,44	0,38	0,15		0,29		0,25	-	0,43	0,44
109	worm	-	0,39		0,1	0,5	0,35		0,19		0,27	-	0,71	1
110	year	1	0,14		0,65	0,62	0,48		0,35		0,29	-	0,57	-

4. Предварительные результаты

Для предварительной демонстрации результатов можно вычислить среднее значение для полученных ИС каждой группы и расположить слова в порядке убывания этого среднего значения. Этот метод аналогичен использованному в работе (Pozdniakov 2014). Результат представлен ниже:

Таблица 2.

Ранжирование ИС

1–10	I, thou, two, eye, we, name, tongue, one, claw (nail), tooth
11–20	dog, die, blood, fire, full, thin, eat, what, heart, louse
21–30	liver, who, water, hand, salt, star, smoke, ear, give, stone
31–40	new, bone, nose, moon, stand, sun, not, hair, fish, year
41–50	tail, this, drink, rain, horn, tree, come, long, foot, snake
51–60	hear, warm, kill, swim, wind, dry, far, heavy, leaf, black
61–70	road, bird, earth, worm, skin, know, egg, fly v., knee, head
71–80	night, seed, meat, that, sleep, sit, fat n., breast, person, feather
81–90	white, ashes, short, say, root, near, red, mouth, bite, see
91–100	neck, man, woman, cloud, walk (go), burn tr., lie, green, good, mountain
101–110	yellow, all, bark, cold, round, belly, small, sand, big, many

Полученное ранжирование во многом совпадает как с исходным ранжированием С. А. Старостина, так и с модификацией К. И. Позднякова. Хотя ряд слов и значительно изменили свою позицию, так слово «зуб», ранжированное С. А. Старостиным 21-ым, согласно нашей работе и данным К. И. Позднякова, оказывается 8-ым, т. е. значительно более устойчивым. В случае слова «полный», которое, согласно исследованию С. А. Старостина, ранжируется 34-ым, а согласно К. И. Позднякову – 32-ым, после исключения лакун, имевшихся в списках почти для каждой семьи, переместилось на 15-ое место.

Следует также отметить, что средняя часть графика представляет собой практически линейную убывающую зависимость. Как представляется, это может говорить о том, что базисная лексика делится на более и менее устойчивую не в отношении 50 – 60, как сделано, например, в (Starostin 2010), а в несколько ином отношении, позволяющем, например, выделять более устойчивую часть из первых 15–20 слов.



Литература

- Dolgopolsky, A. B. 1964: [A probabilistic hypothesis concerning the oldest relationships among the language families of Northern Eurasia]. *Voprosy Jazykoznanija* 2. 53–63. Долгопольский, А.Б. Гипотеза древнейшего родства языковых семей Северной Евразии с вероятностной точки зрения. *Вопросы Языкознания* 2. 53–63.
- Kassian, A., Zhivlov, M., Starostin, G. 2015: Proto-Indo-European-Uralic comparison from the probabilistic point of view. *The Journal of Indo-European Studies* 43. 301–347.
- Pozdniakov, K. 2014: [On the threshold of relationship and the “stability index” of basic lexicon in mass comparison: Atlantic languages]. *Journal of Language Relationship* 11. 187–237. Поздняков К.И. О пороге родства и индексе стабильности в базисной лексике при массовом сравнении: атлантические языки. *Вопросы языкового родства* 11. 187–237.
- Starostin, G. 2010: Preliminary lexicostatistics as a basis for language classification: A new approach. *Journal of Language Relationship* 3. 79–116.
- Starostin, S. 2007: [S. A. Starostin: Works on linguistics. Moscow: Languages of Slavic Cultures, 580–590. Старостин С. А. 2007: Определение устойчивости базисной лексики. В С.А. Старостин: *Труды по языкознанию*. М.: Языки славянских культур, 580–590.
- Swadesh, M. 1952: Lexico-statistic dating of prehistoric ethnic contacts. With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96. 452–463.
- Swadesh, M. 1955: Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics* 21. 121–137.
- Tadmor, U. 2009: Loanwords in the world's languages. Findings and results. In Haspelmath, Martin and Tadmor, Uri (eds.), *Loanwords in the world's languages. A comparative handbook*, 55–75. Berlin and New York: de Gruyter.
- Thomas, D. 1960: Basic vocabulary in some Mon-Khmer languages. *Anthropological Linguistics* 2. 7–11.