

Список Сводеша

(Лексикостатистическое датирование)

Distanz zwischen slawischen Sprachen, SoSe 2010

9. Juni 2010

Lehner Olga

СВОДЕШ, МОРРИС - Morris Swadesh (1909–1967)

СВОДЕШ, МОРРИС, американский лингвист и антрополог. Родился 22 января 1909 в городке Холиок (шт. Массачусетс) в семье эмигрантов из России. Учился у Э.Сепира в Чикагском университете; начал под его руководством работу над диссертацией по индейскому языку нутка, завершённую в 1933 в Йельском университете. В 1933–1937 занимался фонологической теорией и изучением американского варианта английского языка. После двух лет преподавания в университете Висконсина в Мэдисоне переехал в Мехико, где стал директором Совета по индейским языкам; в 1939–1941 профессор антропологической школы Национального политехнического института. Во время войны состоял на государственной службе в Нью-Йорке, занимался редактированием словарей и подготовкой языковых пособий для нужд военного ведомства. После войны преподавал в Сити-университете в Нью-Йорке, откуда в годы маккартизма был уволен за левые убеждения. В 1949–1953 сотрудник библиотеки Американского философского общества, в 1953–1956 занимался полевыми исследованиями как независимый учёный. В 1956 вернулся в Мехико, где до конца жизни был профессором Института истории Национального автономного университета и Национальной школы антропологии и истории. Умер Сводеш в Мехико 20 июля 1967, не успев завершить свой основной теоретический труд Происхождение и диверсификация языка (The Origin and Diversification of Language), посмертно изданный в 1971. www

Список Сводеша

Swadesh list

Предложенный американским лингвистом М. Сводешем метод для оценки степени родства между различными языками по такому признаку, как схожесть наиболее устойчивого базового словаря, положил начало совершенно новому направлению в лингвистике, названному позднее **глоттохронологией**.

Список Сводеша представляет собой **стандартизированный перечень базовых лексем данного языка**, приблизительно (но не точно) упорядоченный по убыванию их «базовости» или исторической устойчивости. Минимальный набор важнейшей («ядерной») лексики содержится в 100-словном списке. Сам Сводеш использовал менее семантически устойчивые, но зато более подробные 200- и 215-словные списки.

Требования к опытному списку:

1. Универсальность
2. Отношение к культуре
3. Простота
4. Межъязыковая двухсмысленность и неясность значений
5. Потенциальное дублирование
6. Идентичные корни
7. Звукоподражание
8. Служебные слова

Англ.-рус. словарь 1-50

1	all	все (о количестве)	26	to cut	резать
2	and	и (союз)	27	day	день (а не ночь)
3	animal	животное	28	to die	умирать
4	ashes	зола	29	to dig	копать
5	at	в (предлог)	30	dirty	грязный
6	back	спина (человека)	31	dog	собака
7	bad	плохой	32.	to drink	пить
8	bark	кора	33.	dry	сухой (вещество)
9	because	потому что	34.	dull	тупой (нож)
10	belly	живот	35.	dust	пыль
11	berry	ягода	36.	ear	ухо
12	big	большой	37.	earth	земля (почва)
13	bird	птица	38.	to eat	есть (принимать пищу)
14	to bite	кусать	39.	egg	яйцо
15	black	черный	40.	eye	глаз
16	blood	кровь	41.	to fall	падать
17	to blow	дуть	42.	far	далеко
18	bone	кость	43.	fat	жир (органич. вещ-во)
19	breathe	дыхание	44.	father	отец
20	to burn	гореть	45.	to fear	бояться
21	child	ребенок (о возрасте)	46.	feather	перо
22	cloud	облако	47.	few	мало
23	cold	холодный	48.	to fight	сражаться
24	to come	приходить	49.	fire	огонь
25	to count	считать	50.	fish	рыба

Англ.-рус. словарь 51-100

51.	five	пять	76.	to hunt	охотиться
52.	to float	плыть	77.	husband	муж (супруг)
53.	to flow	течь	78.	I	я (местоимение)
54.	flower	цветок	79.	ice	лед
55.	to fly	летать	80.	if	если
56.	fog	туман	81.	in	в (предлог)
57.	foot	ступня	82.	to kill	убивать
58.	four	четыре	83.	to know	знать (факты)
59.	to freeze	замерзать	84.	lake	озеро
60.	to give	давать	85.	to laugh	смеяться
61.	good	хороший	86.	leaf	лист (дерева)
62.	grass	трава	87.	left	левый
63.	green	зеленый	88.	leg	нога
64.	guts	кишки	89.	to lie	лежать
65.	hair	волосы	90.	to live	жить
66.	hand	рука	91.	liver	печень
67.	he	он	92.	long	длинный
68.	head	голова	93.	louse	вошь
69.	to hear	слышать	94.	man	мужчина
70.	heart	сердце	95.	many	много
71.	heavy	тяжелый (по весу)	96.	meat	мясо
72.	here	здесь	97.	mother	мать
73.	to hit	ударить	98.	mountain	гора
74.	to hold	держат (в руке)	99.	mouth	рот
75.	how	как	100.	name	имя

Англ.-рус. словарь 101-150

101.	narrow	узкий	126.	sand	песок
102.	near	близкий	127.	to say	сказать
103.	neck	шея	128.	to scratch	чесать (кожу при зуде)
104.	new	новый	129.	sea	море
105.	night	ночь	130.	to see	видеть
106.	nose	нос	131.	seed	семя (зерно)
107.	not	не	132.	to sew	шить
108.	old	старый	133.	sharp	острый (как нож)
109.	one	один	134.	short	короткий
110.	other	другой	135.	to sing	петь
111.	person	лицо (человек)	136.	to sit	сидеть
112.	to play	играть	137.	skin	кожа (человека)
113.	to pull	тянуть	138.	sky	небо
114.	to push	толкать	139.	to sleep	спать
115.	to rain	идти (о дожде)	140.	small	небольшой
116.	red	красный	141.	to smell	нюхать
117.	right	верный (точный)	142.	smoke	дым
118.	right	правый	143.	smooth	гладкий
119.	river	река	144.	snake	змея
120.	road	дорога	145.	snow	снег
121.	root	корень	146.	some	несколько
122.	rope	веревка	147.	to spit	плевать
123.	rotten	гнилой	148.	to split	расщеплять
124.	to rub	тереть	149.	to squeeze	сжать (сдавить)
125.	salt	соль	150.	to stub	вонзать

Англ.-рус. словарь 101-150

101.	narrow	узкий	126.	sand	песок
102.	near	близкий	127.	to say	сказать
103.	neck	шея	128.	to scratch	чесать (кожу при зуде)
104.	new	новый	129.	sea	море
105.	night	ночь	130.	to see	видеть
106.	nose	нос	131.	seed	семя (зерно)
107.	not	не	132.	to sew	шить
108.	old	старый	133.	sharp	острый (как нож)
109.	one	один	134.	short	короткий
110.	other	другой	135.	to sing	петь
111.	person	лицо (человек)	136.	to sit	сидеть
112.	to play	играть	137.	skin	кожа (человека)
113.	to pull	тянуть	138.	sky	небо
114.	to push	толкать	139.	to sleep	спать
115.	to rain	идти (о дожде)	140.	small	небольшой
116.	red	красный	141.	to smell	нюхать
117.	right	верный (точный)	142.	smoke	дым
118.	right	правый	143.	smooth	гладкий
119.	river	река	144.	snake	змея
120.	road	дорога	145.	snow	снег
121.	root	корень	146.	some	несколько
122.	rope	веревка	147.	to spit	плевать
123.	rotten	гнилой	148.	to split	расщеплять
124.	to rub	тереть	149.	to squeeze	сжать (сдавить)
125.	salt	соль	150.	to stub	вонзать

Англ.-рус. словарь 151-200

151.	to stand	стоять	176.	two	два
152.	star	звезда	177.	to vomit	рвать (тошнить)
153.	stick	палка	178.	to walk	ходить (пешком)
154.	stone	камень	179.	warm	теплый (погода)
155.	straight	прямой	180.	to wash	мыть
156.	to suck	сосать	181.	water	вода
157.	sun	солнце	182.	we	мы
158.	to swell	распухать	183.	wet	мокрый
159.	to swim	плавать	184.	what?	что?
160.	tail	хвост	185.	when?	когда?
161.	that	тот	186.	where?	где?
162.	there	там	187.	white	белый
163.	they	они	188.	who?	кто?
164.	thick	толстый	189.	wide	широкий
165.	thin	тонкий	190.	wife	жена (супруга)
166.	to think	думать	191.	wind	ветер
167.	this	этот	192.	wing	крыло
168.	thou	ты	193.	to wipe	вытирать
169.	three	три	194.	with	с (сопровожд.)
170.	to throw	бросать	195.	woman	женщина
171.	to tie	завязывать	196.	woods	леса
172.	tongue	язык	197.	worm	червяк
173.	tooth	зуб	198.	ye	вы
174.	tree	дерево	199.	year	год
175.	to turn	поворачивать (менять направление)	200.	yellow	желтый

Коэффициент сохраняемости

Коэффициент сохраняемости можно вычислять в его отношении к определенному периоду времени, например к тысячелетиям или к столетиям, для того чтобы было легко сравнивать различные случаи друг с другом. Первые два опыта подобного вычисления были проведены автором на материале современного английского языка в сравнении с древнеанглийским и современного испанского в сравнении с латынью. Метод сравнения проиллюстрирован здесь на ряде элементов списка.

Элемент	Древнеангл.	Совр. англ.	Родств. элем.	Латынь	Испанский	Родств. элем.
"все"	eall	all	+	omnēs	todos	-
"и"	and	and	+	et (que)	y	+
"животное"	dēor	animal	-	animal	animal	+
"зола"	æsk	ashes	+	cinis	cenizos	+
"в"	æt	at	+	in	a	-
"спина"	bæc	back	+	dorsum	espaldo	-
"плохой"	fūl	bad	-	malus	malo	+
"кора"	rind	bark	-	cortex	corteza	+
"живот"	belg	belly	+	venter	vientre	+

Коэффициент сохраняемости

После того как подобный анализ был применен ко всему опытному списку, обнаружилось, что совпадение между **древнеанглийским** и **новоанглийским** составило **85 процентов**, а между **испанским** языком и **латынью** оно равно **70 процентам**. Истекшее время в первом случае составляет **1000 лет**, во втором - **2000 лет**. Так, если после тысячи лет **85** процентов первоначального словаря все еще функционирует, как и прежде, то в течение второго тысячелетия тот же коэффициент сохраняемости дал бы **85** процентов от оставшихся **85** процентов, которые все еще сохранялись в начале этого второго периода.

Иными словами, 2000 лет при коэффициенте сохраняемости, характерном для английского языка, составили бы **72 процента**, т. е. немногим более того, что сохранилось в испанском языке по истечении такого же периода времени. Коэффициент сохраняемости в испанском языке равен менее чем 84 процентам за 1000 лет. Следовательно, он практически одинаков в обоих рассмотренных нами случаях.

Коэффициент сохраняемости

Поскольку интервалы во времени обычно не являются кратными друг другу, для удобства, чтобы свести все случаи к стандартному периоду времени, например в тысячу лет, можно пользоваться логарифмами (wikipedia Logarithmus); математически это можно выразить следующей формулой:

$$\log r = \log C : t$$

т. е. логарифм коэффициента сохраняемости **r** за 1000 лет в процентах равен логарифму процента совпадений **c** между словарями, деленному на количество периодов времени **t**.

Robert R. B. Lees, A method of dating with lexicon statistics

	Проценты за 1000 лет
Среднеегипетский (2100-1700 до н. э.) по сравнению с коптским (300-500 н. э.) (Баер; вычислено как 23 столетия)	76
Классич. латынь по сравнению с совр. румынским (Е. Кросс)	77
Древневерхненемецкий (850 н. э.) по сравнению с совр. немецким (Г. Дж. Мэткальф и Р. Д. Лиз)	78
Классический китайский (950 н. э.) по сравнению с совр. северокитайским (Фанг)	79
Латынь Плавта (200 до н. э.) по сравнению с французским Мольера (1650 н. з.) (Д. А. Гриффин)	79
Доминика кариб (1650) по сравнению с совр. (Д. Тейлор и М. Сводеш)	80
Классическая латынь (50 до н. э.) по сравнению с португальским (Е. Кросс)	82
Койнэ по сравнению с совр. кипрским (Э. Хэмп)	83
Койнэ по сравнению с совр. языком Афин (Э. Хэмп)	84
Классическая латынь (50 до н. э.) по сравнению с итальянским (Е. Кросс)	85

Определение интервала времени

При вычислении **коэффициента сохраняемости** необходимо иметь точно датированные примеры и быть уверенным в том, что оба примера находятся на одной и той же линии языковой традиции. Древнеегипетский язык датируется от 2100 до 1700 г. до н.э., коптский - от 300 г. н. э. до 500 г. н. э. Интервал времени между двумя этими примерами несколько расплывчатый, колеблющийся в пределах от 20 до 26 столетий, коэффициент сохраняемости соответственно должен колебаться от 73 до 78 процентов за 1000 лет. Отмечаемая разница не превышает этих величин, поскольку мы имеем дело с продолжительным периодом времени.

Если бы вопрос касался периода в 400-500 лет, то изменение величины этого периода лишь на одно столетие обусловило бы 4 процента разницы в значении коэффициента для 1000 лет. Таким образом, некоторые колебания в коэффициенте сохраняемости вызываются не различиями в самом процессе языкового изменения, но ошибками в определении промежутка времени между сравниваемыми примерами.

Колебания

Р. Д. Лиз выводит «среднюю константу скорости», равную приблизительно **81 проценту ± 2 процента** за 1000 лет, где предел ошибки составляет 0,9 ошибки при подсчете методом малых промежутков времени.

В добавление к статистическому расхождению константа испытывает влияние культурно-исторических факторов. Международные контакты часто ведут к **заимствованию новых понятий и новых слов**. Например, английский язык воспринял так много слов (особенно из латыни и французского), что почти половина всего словарного запаса этого языка может быть определена как заимствование. Однако внутри основного словарного запаса заимствований гораздо меньше, в нашем опытном словаре - только 6 процентов. Они не оказывают никакого влияния на коэффициент сохраняемости в английском языке. Таким образом, язык, который заимствует некоторые слова из чужого языка, не изменяет свой основной словарный запас быстрее, чем другой язык, где единственным видом языкового изменения является замена одной исконной формы другой.

Дивергенция (расхождениа)

Датирование доисторических событий с помощью словарной статистики возможно благодаря языковой дивергенции. Самым простым случаем дивергенции является такой, когда имеет место миграция или внезапное завоевание, которое раскалывает общество на две части и когда условия препятствуют населению обеих частей поддерживать эффективную связь друг с другом после разделения. Изменения, которые происходят вследствие этого в словаре обоих языковых коллективов, проходят, таким образом, независимо друг от друга; слова основного словаря, которые заменяются другими в одном коллективе, могут утрачиваться, но могут и сохраняться в другом. Спустя, скажем, 1000 лет образуются два различных языка, каждый из которых сохранил определенную часть прежде свойственного им обоим основного словаря. Количество сохранившихся общих элементов, подсчитанное с помощью нашего опытного списка, составило бы приблизительно 81 процент.

Дивергенция (расхождение)

Наиболее общее уравнение, по которому мы можем вычислить время, исходя из процента родственных слов, таково:

$$t1 + t2 = \log C : \log r$$

где $t1$ и $t2$ относятся к двум различным языкам или диалектам, которые развились из первоначально единого языка. В этом уравнении $t1$ — время, выраженное в тысячелетиях (или еще в какой-либо удобной единице времени), которые отделяют один из языков от периода их совместного существования в одном языке, а $t2$ — время самостоятельного существования второго языка; C — процент родственных элементов, найденный при сравнении двух опытных списков, а r — индекс (коэффициент) сохраняемости, т. е. установленный контрольным исследованием предполагаемый процент лексических единиц, сохранившихся в языках по истечении каждой единицы времени.

Дивергенция (расхождения)

Одна линия развития: если сравниваются две различные ступени **одного и того же языка**, то одна из ступеней может совпадать с периодом совместного существования; при этом **t1** будет равно нулю, и, таким образом, $0 + t2 = \log C : \log r$, или

$$t2 = \log C : \log r$$

Одна ступень развития: если оба сравниваемых языка являются более поздними ступенями развития некогда единого языка (дивергенция), то $t1 = t2 = t$, отсюда $2t = \log C : \log r$ или

$$t = \log C : 2 \log r$$

$$t = \log C : \log r^2$$

т. е. время дивергенции равно логарифму от процента совпадающей лексики, деленному на удвоенный логарифм коэффициента сохраняемости.

Время дивергенции выражается в столетиях, минимальным подразделением служат полстолетия:

для $r = 81\%$ ($r^2 = 66\%$)

95	1	70	8	45	19	20	38	10	54
90	2,5	65	10	40	21,5	18	40	9	57
85	4	60	12	35	25	16	43	8	59,5
80	5	55	14,5	30	28,5	14	46,5	7	63
75	6,5	50	17	25	33,5	12	50	6	66,5
								5	70,5

Первый столбец выражает процент совпадающей лексики, второй - время дивергенции, выраженное в столетиях.

$$\log C = 2t \log r$$

$$C = r^{2t} \quad C = 0.81^{2 \cdot 0.1} = 0.95$$

Дивергенция (расхождения)

Различные даты, которые автор вычислял в некоторых ранних исследованиях, основывались на $r = 85\%$, что при вычислении дает несколько более продолжительные периоды времени.

В следующей таблице приводится сравнение для периодов в десятки столетий, чтобы показать разницу в результатах вычисления.

Столетия	10	20	30	40	50	60	70	80	90
$r\% = 81$	66	43	29	18	12	8	5	3	2
$r\% = 85$	72	52	37	27	19	14	10	7	5

Литература

- Swadesh M. Lexicostatistic Dating of Prehistoric Ethnic Contacts with Special Reference to North American Indians and Eskimos // Proceedings of the American Philosophical Society. Vol. 96. 1952.
- Swadesh M. Towards Greater Accuracy in Lexicostatistic Dating // International Journal of American Linguistics. Vol. 21. 1955.
- Дьячок М. Т. Глоттохронология: пятьдесят лет спустя [www \(Сибирский лингвистический](#)
- Сводеш М. Лексикостатистическое датирование доисторических этнических контактов // Новое в лингвистике. Вып. 1. - М., 1960. - С. 23-52; [www](#)
- Сводеш М. К вопросу о повышении точности в лексикостатистическом датировании. - Там же. - С. 53-87. [www](#)
- Старостин С.А. Сравнительно-историческое языкознание и лексикостатистика // Лингвистическая реконструкция и древнейшая история Востока (Материалы к дискуссиям международной конференции). Т. 1. М., 1989. - С. 3-39.