

С.Е. Яхонтов

Глоттохронология и степени вежливости в яванских языках

Представленные ниже заметки являются своего рода комментарием к последним нескольким строкам статьи А.К. Оглоблина о родственных связях австронезийских языков малайско-яванского ареала [Оглоблин 2004]. Материал этих языков позволяет предположить, что процесс замены лексики в этом ареале проходил с относительно высокой скоростью.

Общепризнанной классификации австронезийских языков пока нет. Традиционно различали три ветви этой семьи — индонезийскую, полинезийскую и меланезийскую. Сейчас обычно объединяют последние две под общим названием океанической (или геонезийской) ветви, противопоставляя ее индонезийской (гесперонезийской); кроме того, особо выделяют молуккские языки и три группы языков Тайваня. В составе индонезийской ветви различают ряд групп и подгрупп, важнейшие из которых (по числу входящих в них языков) — малайско-яванская и северо-западная (филиппинская).

Среди существующих исследований по классификации австронезийских или индонезийских языков для нас особый интерес представляет работа И. Дайена [Dyen 1965], так как в основе ее лежит метод глоттохронологии — тот же, что использован в статье А.К. Оглоблина, о которой здесь идет речь.

Глоттохронология, или лексикостатистика, — это метод определения возраста родственных языков (т.е. давности их разделения) по количеству сохранившихся в них слов общего происхождения. Этот метод может быть использован и для других целей, не связанных непосредственно с лингвистическим временем, в частности, для классификации языков внутри семьи.

Общие слова в глоттохронологических исследованиях подсчитываются в пределах заранее составленного «испытательного списка». Существуют стандартные списки из 100 и 200 слов (точнее, понятий), но многие исследователи предлагают собственные наборы единиц, подлежащих сравнению, или, основываясь на одном из стандартных списков, вносят в него некоторые изменения. Так, в статье А.К. Оглоблина приведен вариант списка из 200 слов, предложенный для австронезийских языков Р. Бластом [Оглоблин 2004: 258–259].

Давность разделения двух родственных языков подсчитывается по формуле

$$t = \log C : 2 \log r. \quad (1)$$

Здесь t есть время (давность разделения языков) в тысячах лет; C — доля совпадающих слов (число совпадений, деленное на число слов в списке); r — «коэффициент сохраняемости», т.е. доля слов, сохраняющихся в языке по прошествии тысячи лет.

Коэффициент r определяется эмпирически по материалам языков, история которых хорошо известна. Если для исследования нескольких языков используется один и тот же список слов, то предполагается, что значение r для всех этих языков (или их пар) одинаково. Но для разных списков коэффициент различен.

Лексикостатистические исследования связаны с теорией вероятности, поэтому результаты их всегда бывают неточными, приблизительными. Возраст родственных языков, вычисленный по формуле, не вполне соответствует действительному. Однако величину возможной ошибки можно оценить. Разница между числом совпадающих слов, сохранившихся в двух сравниваемых языках, и средним числом совпадений в языках, имеющих тот же реальный возраст, в $2/3$ случаев не превышает некото-

рой величины σ («стандартного отклонения»). Стандартное отклонение вычисляется по формуле

$$\sigma = \sqrt{Np(1-p)} \quad (2)$$

Здесь σ — стандартное отклонение; N — число слов в списке (обычно 100 или 200); p — доля общих слов (т.е. слов, которые в двух сравниваемых языках имеют общее происхождение). Величина p — правильная дробь, т.е. она всегда меньше единицы.

Значение σ зависит как от общего числа слов в списке (N), так и от числа совпадающих слов в двух сравниваемых языках.

Например, если сравниваются два языка по списку из ста слов и оказывается, что половина их (50) имеет в этих языках общее происхождение, то по обычной формуле давность их разделения будет определена (округленно) в 2300 лет. Но стандартное отклонение при этих условиях составит $\sqrt{100 \times 0.5 \times 0.5} = 5$, что означает ошибку до пяти слов в двух случаях из трех. С учетом этого следует исходить из 45–55 совпадений между языками, или из возраста в 2000–2650 лет.

Но помимо неизбежной математической неточности возможны и ошибки чисто лингвистические. Например, если для обозначения некоторого понятия в языке имеются два синонима, какой из них надо выбрать? Мнения могут разойтись. Наконец, возможны ошибки, связанные с недостаточной компетентностью исследователя (это в полной мере относится к автору настоящей статьи).

А.К. Оглоблин в основном работает с четырьмя языками: малайским, мадурским, сунданским и яванским. Среди них только малайский и мадурский более близки друг к другу. Они имеют 25 общих слов в списке из 35 единиц, который используется А.К. Оглоблиным; в остальных парах языков число совпадений приблизительно одинаково — от 16 до 19, в среднем около половины списка; стандартное отклонение близко к 3. Расхождения в числе совпадающих слов в этих парах не выходят за пределы стандартного отклонения и скорее всего являются случайными (хотя возможно, что несколько более высокое число совпадений между мадурским и яванским — 19 слов —

является результатом длительного контакта между обоими языками, как считает А.К. Оглоблин).

Так или иначе, все, что касается этих языков, представляется вполне убедительным.

Но все меняется, когда исследователь включает в свою работу материал тагальского языка. Тагальский имеет 21 общее слово с малайским языком. Получается, что разделение малайского и мадурского, с одной стороны, и яванского и сунданского — с другой, предшествовало отделению малайского от тагальского, хотя последний принадлежит к другой подгруппе индонезийских языков (филиппинской).

Прежде чем идти дальше, приведем более полные данные по языкам, о которых идет речь. Сравним тагальский язык по списку 35 слов со всеми четырьмя языками малайско-яванского ареала.

Таблица 1

	мадурский	яванский	сунданский	тагальский
малайский	25	17	16	21
мадурский		19	17	16
яванский			17	16
сунданский				14

Малайский и тагальский имеют между собой больше совпадений, чем любая другая пара языков, кроме малайского с мадурским.

Сравним теперь те же языки по списку 100 слов¹.

¹ Я пользуюсь списком, частично отличающимся от стандартного. Из него исключены слова «все», «гореть», «коготь», «кора», «кусать», «лежать», «мы», «перо», «семя», «теплый». Перевод их английских эквивалентов на ряд языков (австронезийских и других) представляет трудности. В частности, некоторые понятия в этих языках выражаются словосочетаниями: «кора» — «кожа дерева», «перо» — «волосы птицы»; слово «мы» представляет собой регулярную форму множественного числа от «я»; «теплый» обозначается так же, как «горячий», «лежать» — так же, как «спать»; английское *burn* может быть понято как «гореть» или «сжигать»; слово со значением «семя» отсутствует

Таблица 2

	мадурский	яванский	сунданский	тагальский
малайский	58	33,5	46	38
мадурский		41	45	36
яванский			42,5	25
сунданский				29

Здесь ближе всего друг к другу по-прежнему стоят языки малайский и мадурский, затем они же с яванским и сунданским; дальше всего от других оказывается тагальский. Есть одно исключение: очень малое число общих слов у малайского и яванского. Во всех других случаях число совпадений каждого языка с малайским оказывается больше, чем с любым другим языком.

Такое же особое место, как малайский, занимает тагальский язык среди филиппинских. Можно подсчитать по данным В.А. Макаренко [Макаренко 1965] число совпадений слов в четырех филиппинских языках. В этих материалах имеются 92 слова из списка 100 (из моего варианта этого списка). Было подсчитано также число слов (в пределах 92), совпадающих в этих четырех языках с языком дусун (кадазан) на Северном Калимантане, который И. Дайен относит к одной группе с филиппинскими языками (северо-западной).

у народов, не знающих земледелия, и т.п. Эти десять слов были заменены другими десятью (взятыми из списка 200 слов): «близко», «ветер», «год», «далеко», «змея», «короткий», «соль», «тонкий», «тяжелый», «червяк». Эта замена может повлиять на подсчеты, но едва ли это влияние может оказаться существенным.

Списки слов языков Индонезии составлены по: [Sugiarto 1999]. Автор пользуется случаем, чтобы выразить свою благодарность А.К. Оглоблину, любезно предоставившему ему возможность пользоваться этой работой.

При подсчетах совпадение считалось за половинное (0,5), если для какого-то элемента списка в языке параллельно существуют два слова, и неясно, какое из них следует выбрать.

Таблица 3

	бисайя	пангасинан	илокано	дусун
тагальский	50	42	42,5	37,5
бисайя		34,5	33	32,5
пангасинан			44,5	30,5
илокано				31

Видно, что особенно близки между собой тагальский и бисайя (50 совпадений), затем пангасинан и илокано (44,5 совпадений). Теоретически тагальский и бисайя, близкие друг к другу, должны бы иметь приблизительно одинаковое число совпадений с каждым из остальных языков; но на самом деле тагальский всегда имеет на несколько совпадающих слов больше, чем бисайя.

Итак, индонезийский и тагальский занимают особое положение среди всех рассмотренных австронезийских языков.

Есть основания думать, что в языках с долгой литературной традицией, к числу которых относятся малайский и тагальский, лексика обновляется медленнее, чем обычно, или вообще не обновляется. Это значит, что в них всегда реально сохраняется несколько больше слов праязыка, чем в других языках того же возраста, поэтому в них больше слов, общих с другими языками той же группы и семьи, чем ожидалось бы.

Как кажется, обратная ситуация наблюдается с сунданским и особенно яванским языком: число совпадений с родственными языками оказывается меньше, чем ожидалось бы. Мы уже видели, что число общих слов у яванского языка с малайским очень невелико.

Сравним по списку 100 слов четыре интересующих нас малайско-яванских языка с еще двумя языками той же группы — мокен и зарай (джарай). (Они не учтены в классификации И. Дайена.) Зарай — язык чамской (тямской) группы в Южном Вьетнаме. Мокен (селун) — язык «морских цыган»; по моим подсчетам, он стоит немного ближе к чамской группе, чем к малайской (хотя разница незначительна и может быть случайной).

Больше всего общих слов (58) в малайском и мадурском. Сравнивая дальше эти два языка с остальными, мы видим, что лю-

Таблица 4

	сунданский	мадурский	малайский	мокен	зарай
яванский	42,5	41	33,5	22	24,5
сунданский		45	46	28	33,5
мадурский			58	37	35,5
малайский				45	41,5
мокен					43,5

бой из них имеет немного больше совпадений с малайским, чем с мадурским (единственное исключение — яванский). Мы уже знаем об особом положении малайского и можем больше к нему не возвращаться. Мокен и зарай имеют 43,5 общих слова, в среднем столько же с малайским и несколько меньше (37 и 35,5) с мадурским. Но с сунданским и яванским картина сложная — число совпадений их с другими языками колеблется от 22 до 46 (или от 22 до 45, если не считать малайский).

Можно предположить или ускоренное развитие яванского (и, возможно, сунданского), или ошибку в классификации. Для того чтобы выбрать одну из этих двух точек зрения, сравним несколько гесперонезийских языков с двумя полинезийскими — маори и гавайским. Теоретически мы должны были бы во всех случаях получить приблизительно одинаковое число совпадений (отражающее давность разделения гесперонезийских и геонезийских языков).

Таблица 5

		маори	гавайский	среднее
1	чам (тям)	29	29	29
	малайский	29	28	28,5
	тагальский	26	30	28
2	мокен	26	29	27,5
	зарай	26	27	26,5
	дусун	22	24	23
3	сунданский	20	22	21
	яванский	20	19	19,5

По числу совпадений с индонезийскими языки разделились на три группы. Первая нам уже известна: это языки с длительной литературной традицией. Сюда вошел и язык чам — такой же старый литературный язык, как и малайский. Вторая группа не имеет каких-либо специальных отличительных признаков. Третья — это интересующие нас сунданский и яванский. Можно предположить, что есть некоторый фактор, вызывающий уменьшение числа совпадений языков Явы с другими языками.

Таким фактором может быть существование в этих языках степеней вежливости («форм обращения», «этикетных отношений»), то есть необходимость для говорящего использовать разные слова для обозначения одних и тех же понятий в зависимости от социального статуса лица, к которому он обращается. Наличие двойных обозначений для многих понятий может привести к более быстрым, чем обычно, изменениям в лексике.

Несомненно, язык не может «позволить себе» иметь двойные обозначения для *каждого* понятия. По-видимому, вежливые соответствия появляются в первую очередь у слов наиболее обычных. Так, в яванском вежливом языке кроме имеются специальные числительные только до пяти; дальше вежливый язык (кроме) и обычный (нгоко) используют одни и те же слова. Испытательные списки глоттохронологии состоят именно из наиболее обычных слов, поэтому, если ускоренная смена лексики действительно связана с формами вежливости, слова списков должны оказываться в этом отношении более уязвимыми, чем лексика в среднем.

Библиография

Макаренко В.А. О степени родства тагальского и индонезийского языков (с использованием метода лексикостатистической теории М. Свадеша) // Вопросы филологии стран Юго-Восточной Азии / Под ред. Ю.Н. Мазура. М.: Изд-во Московского ун-та, 1965. С. 73–103.

Оглоблин А.К. К вопросу о родственных связях в малайско-яванском ареале (лексикостатистическая оценка по 35-словному списку Яхонтова) // Проблемы этнографии и истории культуры народов Азиатско-Тихоокеанского региона. СПб: Петербургское востоковедение, 2004. С. 258–263.

Dyen I. A lexicostatistical classification of the Austronesian languages. Indiana University publications in anthropology and linguistics. Memoir 19. Supplement to International journal of American linguistics. 1965. Vol. 31. № 1.

Sugiarto Jr., dkk. Kamus Indonesia – Daerah Indonesia, Jawa, Bali, Sunda, Madura / Edisi Kedua, direvisi dan diedit oleh Widya Kirana. Jakarta, 1999.