Российский государственный гуманитарный университет
Russian State University for the Humanities

УДК 81(05)
ББК 81я5

Table of Contents / Содержание

Articles / **Статьи**

Discussion Articles / **Дискуссионные статьи**

Book reviews / **Рецензии**

Reports / Хроника

Note for Contributors

Journal of Language Relationship welcomes submissions from everyone specializing in comparative-historical linguistics and related disciplines, in the form of original articles as well as reviews of recent publications. All such submissions should be sent to the managing editor:

G. Starostin
Institute of Oriental Cultures and Antiquity
Russian State University for the Humanities
125267 Moscow, Russia
Miusskaya Square, 6
E-mail: journal@nostratic.ru

Articles are published preferably in English or Russian, although publication of texts in other major European languages (French, German, etc.) is possible. Each article should be accompanied with an abstract (not exceeding 300 words).

For more detailed guidelines on article submission and editorial policies, please see our Website at: http://journal.nostratic.ru or address the editorial staff directly at journal@nostratic.ru.

Будущим авторам

Журнал *Вопросы языкового родства* принимает заявки на публикацию оригинальных научных статей, а также рецензий, от всех, кто специализируется в области сравнительно-исторического языкознания и смежных дисциплин. Рукописи можно высылать непосредственно заместителю главного редактора по адресу:

125267 Москва
Миусская площадь, д. 6
Российский государственный гуманитарный университет
Институт восточных культур и античности
Г. Старостину
E-mail: journal@nostratic.ru

Предпочтительные языки публикации — английский или русский, хотя возможна также публикация статей на других европейских языках (французский, немецкий и т. п.). К каждой статье обязательно прикладывается резюме (не более 300 слов).

Подробнее о требованиях к оформлению рукописи, редакционной политике журнала и т. п. Вы можете узнать на нашем сайте по адресу: http://journal.nostratic.ru, или же непосредственно обратившись к редакции по электронной почте (journal@nostratic.ru).

В. А. Дыбо, А. В. Дыбо
РГГУ, Москва

Дополнения к 3-му тому «Опыта сравнения ностратических языков» В. М. Иллич-Свитыча. 1

Предлагаемая заметка начинает серию публикаций словарных статей, подготовленных для 3-го тома «Опыта сравнения ностратических языков», но не вошедших в него по техническим причинам. При необходимости даются комментарии об отношении вовлеченных в сравнение сближений к вхождениям в современные этимологические корпуса (прежде всего EDAL по алтайским языкам, также UEW, который к моменту написания выпуска вышел еще не полностью, HSED и 2 тома SED, Takacz EDE, Stolbova 1996, 2007, EWK 1998, Starostin G. 2004). Также дается комментарий об отношении ностратических этимологий, публикуемых нами, к вхождениям Ностратического словаря А. Б. Долгопольского (NED).

Ключевые слова: этимология, компаративистика, ностратика.

Предлагаемая серия публикаций представляет собой словарные статьи, подготовленные для III выпуска «Опыта сравнения ностратических языков» с соблюдением принципов, принятых в этом выпуске, но не вошедшие в него по техническим причинам. Как известно, статьи III выпуска составлялись по картотеке В. М. Иллич-Свитыча. Вначале этой работой занимались А. Б. Долгопольский и В. А. Дыбо, при технической помощи со стороны Р. В. Булатовой, затем, после эмиграции А. Б. Долгопольского в 1978 г., работа по составлению и описанию этимологий практически перешла в руки В. А. Дыбо и А. В. Дыбо; в редактировании этимологий по отдельным ветвям ностратической макросемьи приняли активное участие О. А. Мудрак, О. В. Столбова и Е. А. Хелимский. Таким-то образом при составлении третьего выпуска победила стилистика «длинных этимологий» (ср. Бурлак—Старостин 2005, 110), максимально учитывающих данные по этимологии дочерних семей и не отступающих перед необходимостью переработки «внутрисемейных» этимологий; при этом мы считали, что все такие переработки должны быть эксплицированы в текстах соответствующих статей Ностратического словаря.

Поскольку за прошедшее со времени нашей работы над III выпуском Опыта сравнения в реконструкциях и этимологических корпусах ряда дочерних семей произошел прогресс различных степеней и направлений, в настоящей публикации при необходимости даются комментарии об отношении вовлеченных в сравнение сближений к вхождениям в современные этимологические корпуса (прежде всего EDAL по алтайским языкам, также UEW, который к моменту написания выпуска вышел еще не полностью, HSED и 2 тома SED, Takacz EDE, Stolbova 1996, 2005, 2007, EWK 1998, Starostin G. 2004 [= DRAVET]). Также дается комментарий об отношении ностратических этимологий, публикуемых нами, к вхождениям Ностратического словаря А. Б. Долгопольского (NED).

Мы совершенно уверены в том, что публикуемые нами этимологии не являются истиной в последней инстанции и могут быть впоследствии в значительной мере перера-

ботаны. Но также мы уверены в том, что избранный нами путь максимальной экспликации принимаемых этимологических решений — наиболее продуктивный в этимологической работе и, безусловно, во всех случаях сложной этимологизации, которая неизбежна при исследованиях дальнего родства.

Система сокращений и условных обзначений в публикуемых статьях — та же, что в вышедших выпусках ОСНЯ; некоторые дополнения к литературе даются по мере публикации этимологий, в которых на них есть ссылки.

Ностр. *n-

1. neHrʌ 'тонкий, узкий; (?) проникать сквозь узкое пространство': и.-е. *neHr-* 'проникать, втискиваться, нырять; нора, пещера; тесный, узкий' ~ драв. *nēr̄ʌ* 'тонкий, узкий' ~ алт. *nArA* 'узкий, тонкий'.

И.-е. *neh₁r-* 'проникать, втискиваться, нырять; нора, пещера; тесный, узкий' ‖ лит. *nérti*, praes. 1.sg. *neriù* 1) 'нырять, проникать; проплывать под водой; выскользнуть, убежать; проскользнуть'; 2) 'протягивать, продевать (нить, шнур, веревку) сквозь отверстие; вдевать нитку в иголку; завязать узел, петлю, скрестить (руки); плести, вязать; вывихнуть или вправить сустав'; лит. *nìrti*, praes. 1.sg. *nyrù* (< *nįrù* < *ni-n-rù*) 1) 'погружаться (в воду), нырять'; 2) 'вывихиваться (о суставах)'. Попытка Э. Френкеля разделить основы по первому и второму значениям как принадлежащие к разным корням с разными внешними генетическими связями: 1) и.-е. *ner-* 'нырять...', 2) и.-е. *(s)ner-* 'крутить, плести', — по-видимому, неприемлема. Во-первых, и.-е. корень *(s)ner-* 'крутить, плести', — скорей всего, фикция (см. соответствия в Pok. 975—976). Во-вторых, что важнее, первый и второй комплексы значений представляются вполне сводимыми при условии принятия первичного различия между ними по диатезе: 1) 'проникать в узкое отверстие, нору; нырять в воду'; 2) 'заставлять что-либо проникнуть в узкое отверстие, вдевать'; возможно, в балтийском, это различие по диатезе выражалось первично противопоставлением тематической основы (с нулевой огласовкой корня) и йотированной основы, ср. лтш. *nir̃t*, praes. 1.sg. *niru* (вторично: *nirstu*) 'нырять' и лтш. *ner̃t*, praes. 1.sg. *neṛu* 'зашнуровывать концы лаптя' (об этом способе выражения диатезы в балтийском см. *Chr. Stang* Das slavische und baltische Verbum. Oslo, 1942, S. 104—125). Развитие вторичного различия по диатезе в литовском посредством распространения *n*-инфиксации привело к смешению первичных основ и к современному состоянию. Для семантики ср. также: лит. *išnérti* 'вынырнуть; вывихнуть, развязать' (т.е. *'заставить выйти из суставного гнезда, из узла'); *įnérti* 'нырнуть; вдеть, зацепить' (т.е. *'заставить войти в узкое отверстие'); *į-si-nérti* 'вдеться, одеться (во что-нибудь)' (Юшк. II, 545) (букв. 'втиснуться в одежду'); *nértis iš káilio* 'лезть (вылезать) из кожи вон'; производные имена: *išnara* f. '(змеиный) выползок', лит. *nãras* m. 'нора (диких животных)', *nãras* (4) m. 'водолаз; гагара' (ср. русск. *ныряльщик; нырок*); *narùs* 'юркий, проворный, подвижный' (букв. 'проскальзывающий, вскальзывающий'); *nerìs* 'бобр'; *nar̃vas* m., *narvà* f, 'маточник (сотовая ячейка для личинки матки), птичья клетка'; *užnarvis* m., *užnarvė* 'закоулок, укромное место, убежище, захолустье'; сюда также лит. *nėróvė* (1) и *nėrõvė* (2) NdŽ 'vandens nimfa, undinė; Seejungfrau, Nixe; русалка' LKŽ, *nirà* 'Nixe; русалка', *neretà* 'русалка'. Лтш. *nir̃t*, praes. 1.sg. *nirstu* и *niru* 'нырять', *nir̃tiês* 'нырять, погружаться'; *ner̃t*, praes. 1.sg. *neṛu*, praet. 1.sg. *nēru* 'стягивать (зашнуровывать) концы лаптя (die Spitze des Bastschuhs zusammenziehen)'; *nirêt*, *nirdât* 'нырять'; *nira*, *nire*, *ņira* 'гагара или лысуха; черная приморская ласточка, мартын; гоголь; поганка'; прус. *nurtue* 'hemde (Hemd) — marškiniai' E 479 nom.sg. fem. = pr. *nurtuvē*. (Mažiulis 3: 203) ‖ Слав.

*nerti̯, praes. 1.sg. *nь̏rǫ, 3. sg. *nьrètь, iterat. *nīrắti, *norĭti 'проникать, нырять': ст.-слав. **нрѣти**, praes. 3.sg. **ньретъ**; aor. 3.sg. **понрѣтъ** (Песни VI, 6); отглаг. сущ. **изнертие** (фиксацию форм см. Вайан, 337, § 215) 'погружаться, погрузиться; проникать во что-л.'; южн. ц.-слав. **оуньрѣти** (Исих. CXLI), aor. 3.sg. **понрѣтъ глава моа въ пропасти горьныа** (Іон. II.6, Упырь); ц.-слав. ***изнрѣти**: **правѣднꙑи изъ ловица изньретъ** (Пар. 1271 г. — 1370 г.); др.-русск. **вынерети** 'внезапно появиться из-под земли, вынырнуть', ***понерети** 'погрузиться, нырнуть': **не хоцем из них на старость вынерети** (Сл. О постех XVI в.); **Симон… и вынре и-землѣ** 'внезапно появился из-под земли' (Флавий Полон. Иерус. XVI ~ XI); **исходица водьнаа понереста очи мои** 'в источники водные погрузились очи мои', 'διεξόδους ὑδάτων κατέβησαν οἱ ὀφθαλμοί μου' (Златостр. XII в., Пс. CXVIII, 136 — ошибочен перевод в Срезн. II 1184: 'залить, покрыть', предполагающий согласование, исключенное формой глагола: aor. 3.du.; acc.pl. **исходица водьнаа** выступает в этом предложении в качестве беспредложного вин. пад. направления); **понерло есть море** (Жит. Андр. Юр. XV–XVI в.); **вънерети**: **Опат есмы внерли ако 'в сѣт в первага отрицаниа дьавола** (Сл. Св. отец о постах, XVI в.); **вънърѣша бо нѣции человѣци въ сеи грѣхъ нечестиа** (Окружн. посл. мт. Фотия, XVI в. ~ 1416 г.); укр. зап. _нéрти_, praes. sg. 1. _нру_, 2. _нреш_, 3. _нре_ 'погрузиться', _занéрти_ 'нырнуть, погрузиться', _знéрти, знерáти_ 'вынырнуть, всплыть; сорваться (_кінь з прикóну_); устать'; _понéрти_, praes. sg. 1. _понерý_, 2. _понерéш_ 'нырнуть, погрузиться'; словен. (Plet.) _pondréti_, praes. 1.sg. _pondrèm_ 'untertauchen', _pondrt_ 'untergetaucht'; др.-чешск. _winrzi_ imperat. 3.sg. 'scateat' (Gebauer III, 199); слав. iterat. *norĭti 'нырять, проникать': др.-русск. **пронорити** 'проникать': **тонкоразумноа душе умомъ вса богодуховныа книгы проноради** (Кир. Тур. Пис. Вас.); схрв. диал. _nòriti_ 'нырять' (по Шимкевичу 159: босн. _noriti_, рагуз. _noritti_; нормат. _ro̍niti_ 'нырять' с метатезой); ст.-кайк. (XVII в.) praes. 1.sg. _Norím_ 'Plavam' (Belostenac II, 275); названия животных (и птиц), которые «ныряют»: др.-русск. **поноровъ** и **поноровь** 'дождевой червь': **И есть въ дрѣвци томъ червь великъ, ꙗко поноровъ въболѣ есть, за корою древца того** (Дан. Иг. Пал. 9); ст.-чеш. _pondrav_ и _ponrav_ 'личинка майского жука', польск. _pądroẃ_ 'дождевой толстый червь'; укр. _норúця_ 'полевая мышь', слав. *norьka 'норка', укр. диал. _нур_ 'Seetaucher; гагара', слав. *norъ m., *nora f.: русск. диал. _нор_ 'яма в воде, омут', др.-русск. **нора** 'подкоп': **подкопашаса подъ стѣну двѣстѣ саженъ и подкатиша норами бочки съ зѣльами и зажгоша съ устьа** (Псков. I лет. 7043 г.), русск. _норá_ 'нора'; укр. _норá_; чеш. _nora_, польск. _nora_ 'нора'; др.-серб. **понорь** 'locus ubi fluvius sub terram absconditur' (Даничиħ Pj. II, 362), словен. _pónor_, gen.sg. _ponóra_ 'понижение, пропасть' ‖ др.-инд. _naraka-_ 'ад', пали _naraka-_ 'пропасть, яма', осет. _naræg_ 'теснина; ущелье'; осет. _naræg_ 'тонкий, узкий, тесный', хорезм. _narak-_ 'узкий' (Henning Khwar. 16), афг. _narai_ 'тонкий, узкий' ‖ греч. νέρτερος 'находящийся внизу, нижний; подземный'; греч. νηρίδας· τας κοίλας πέτρας 'пещеры', греч. νέρτος· ίέραξ, οἱ δὲ εἶδος ὀρνέου Hesych. 'ястреб или сокол': собственно 'Taucher, Tauchvogel; ныряющая птица' (Fraenk. 495); греч. Νηρεύς — имя морского бога, Νηρεῖδες 'Meermädchen' — его дочери; ср. лит. _nérovė_ (1) и _nerõvė_ (2) 'Seejungfrau, Nixe'; ? лтш. _nãra_ 'Wassernymphe' (из лив. _nõra_ — то же, ср. эст. фолькл. _narva_ 'ūdensjaunava' Karulis 617) ‖ арм. _нед_ 'узкий'; арм. _nerk'in_ 'unterer' ‖ др.-исл. _nǫrr_ 'enge, schmal', др.-англ. _nearu_ (совр. англ. _narrow_) 'узкий, тесный; плотный'; др.-сакс. _naro_ 'узкий'; др.-исл. _norðr_ n. 'север' (место, где заходит (ныряет) солнце), нов.-исл., фарер. _norður_, нов.-норв. _nord_, нов.-швед. _norr_, нов.-датск. _nord_; др.-англ. _norð_, др.-фриз. _north, nord_, др.-сакс. _north_, др.-в.-нем. _nord_ 'север'; швед. диал. _nor_ 'узкий пролив' ‖ Fraenk. 495; Karulis 617; Mažiulis 3: 203; Frisk II, 315—316, 308.

Драв.: юж.-драв. *nēr- 'thin, lean', тамил. _nēr_ (-v-, -nt-) 'to grow thin, lean, be emaciated, be soft, yield to the touch; n. minuteness, smallness, fineness, slenderness'; _nērmai_ 'fineness,

thinness, minuteness'; мал́то *nēr* 'thin, delicate', *nēruka* 'to be fine'; *nērkka* 'to become thin, fine'; *nērcca* 'fineness'; *nērppu* 'fineness, thinness, liquefaction'; *nērppikka* 'to make thin, fine, attenuate'; *nērmma* 'fineness, softness, delicacy'; *nēriya* 'fine, thin (as cloth, flour)'; тул́у *nērè* 'thin, delicate', *nērmè*, *nervè* 'fine, thin, delicate' ‖ DEDR 3771.

Алт. **nArA* 'узкий, тонкий' ‖ монг. **narijin < *nari-γin* 'тонкий, узкий': сред.-монг. (SH, MA, ИМ, Lig. VMI) *narin*, (LH) *narīn*, халха *нарийн* 'тонкий, узкий', калмыцк. *närn*, бурят. *нарин* 'узкий, тонкий, изящный', ордос. *narīn* (Mostaert 484), хорч., джал., дурб., горлосск., архорч., бар., онн., найм., харч., тум. *нє͟ρīŋ*, шгол., уцаб., чах. *нарīŋ,* дагур. *naῐīn*, могол. *nōrin*, монгор. *narin,* (минхэ) *nariŋ*, дунс. *narun*, баоан. *naroŋ*, сыч. *narən* Тод. Монгор. 348. ПМонг. > маньчж. *narχin* 'тонкий; слабый, нежный; искусный; бережливый, скаредный', сибо *narəhun* 'fine, thin, exact', чжурчен. *nargi* 'narrow' (впрочем, для монг. и тунг. не исключено и алтайское родство; тогда следующее тунг. сопоставление неприемлемо). ‖ <тунг. **nire-* 'худой, тонкий, слабый', если метатеза из ***neri*: эвенк. **nire-kūn* 'худой, тощий'; Мыреева ЭРС *нюркӯн* [*ńurkūn*] ПТ, Н, Е 'сухощавый, жилистый; исхудавший, тощий'; *нэркун-ми* ПТ, Н, Е 'худеть' (пример — о медведях), *нюргэ* И 'тощий, заморенный', *нирку* Учр. 'тонкий (о дереве, человеке)'. Сюда могло бы относиться также маньчжур. *ńere* (если < **nire*) 'тонкий, непрочный; слабый; легкий (об одежде)', но долгота в форме сибо *nīrə* 'weak, weakly, feeble' Yamamoto 1969 №2502 указывает скорее на **ni/ej/γere>* ‖ См. EDAL II, 972. Приведенное в этой же этимологии тюрк. **jAr-* '1 тонкий, тощий 2 бедный 3 плоский' основано отчасти на недоразумении. В сравнение вовлечены, во-первых, тюрк. **jar-liγ* 'несчастный, жалкий; бедный', семантически далекое от монг. сравнения (ранние значения — именно из эмоциональной сферы, см. СИГТЯ 2000, 334), во-вторых, восточно-кыпчакские формы, восходящие к **jara-* 'быть подходящим, подготовленным, тренированным' и его производному **jara-γu* 'подготовленный, тренированный (к походу) > худощавый, жилистый (о лошади)', в-третьих, горно-алт. *jartï-čaq* (источник?) 'плоский' — явное производное от тюркского сибирского слова: телеут. *jartï*, хакас. диал. *čartï* 'доска' (и тув. *čartï* 'щепка', от глагола **jar-* 'рубить', см. Вербицкий 83, ЭСТЯ 1989, 145) с суф. уподобления *-čaq*, т.е. «как доска». Приведенное там же тув. *čarïγ-da-* значит не 'тонкий, тощий', а 'расходовать, тратить', заимств. из монг. (письм.-монг. **ǯaru-*, халха *зарах* 'расходовать, тратить', *зарагдах* 'расходоваться'). По-видимому, не сюда также относятся корейск. **jŕrp-* 'тонкий, негустой, слабый': ср.-корейск. *jŕrp-,* совр. корейск. *jǎlp-, jālp-* KED 1181 (скорее к алт. **ńál'ba* 'молодой, свежий' EDAL 1413) и япон. **nàràs-* 'выравнивать, делать ровным, укатывать, сглаживать; обучать' JLTT 732 (скорее ср. кор. *narä* 'a soil leveller', 'приспособление для выравнивания почвы' KED 288, БКорРС 1, 2316). В общем, без сомнения к рассматриваемому корню может быть отнесено только монг. слово.

◊ Ср. Иллич-Свитыч МС 357 (**nurʌ* 'проникать²': и.-е. **ner-* 'проникать, нырять' ~ драв. **nur̥ʌ-* 'вползать, проникать'). Ср. NED 1560 **ñar[ü]* 'thin, narrow' (и.-е. ~ алт. + груз. *narnar-i* 'zart, fein, sanft', последнее с возможностью отнесения к другому ностратическому корню, № 1624 **ńärE* 'unripe, tender, weak' — но в последнем много ошибочных сопоставлений). Ср. NOSTRET 1115 (предлагаемое там вслед за NED 1624 уральское сравнение следует разбить: прибалто-финское **näre* 'молодая ель' SSA 2, 253 следует относить к **ńȣrȝ* 'прут, молодой побег' UEW, 331; остаются праперм. **nȩr* 'недоспелый, недозрелый, молодой' Лыткин — Гуляев 199 и праобско-угор. **ńār* 'сырой (roh)' Honti 469, которые семантически далеки от рассматриваемого корня).

2. nuHŕʌ 'проникать, втискиваться, нырять': и.-е. *neuHr-* 'проникать, втискиваться, нырять; низ' ~ драв. *nūr̥ʌ* 'проникать, втискиваться, проползать сквозь узкий проход' <~ алт. **nűRu* 'жилище, убежище'>.

И.-е. *nūr- (< *neuHr- или *neHur-): греч. νυρεῖ· νύσσει 'колет, толкает'; греч. гомер. νείαιρα 'inferior, γαστήρ; низ живота' (= тохар. B ñoriya kātso); тохар. B ñor adv. 'below, beneath, under; down; под, ниже; вниз, внизу' < и.-е. *neh₁wṛ- Adams 272 ‖ слав. iterat. *nyrati, praes. 3.sg. *nyrjetъ и *nyrajetъ 'погружаться, нырять': Змиєвы мысли акы непⷮтыревє въ тⷨмоу нырютъ (Иак. посл. Дмитр.); голубь... от мѣсто на мѣсто... в каменную пєщєрицꙋ... внырꙗꙗ (Варлаам и Иосаф, XVI ~ XII–XIII вв.); та жє рѣка понырꙗꙗ идєть подъ зємлю (Пут. Генн. и Позн.); змии вынырнꙋ изъ хврастиа (Ник. лет. 1537); слав. causat. *nȳriti и *nŭriti 'погружать': польск. nurzyć 'погружать', nurzać 'погружать, окунать', nurzać się 'погружаться, окунаться'; н.-луж. nuriś 'погружать'; укр. диал. зап. занýрити 'погрузить в воду', пронýритися 'погрузиться в воду'); русск. диал. занýрити и заны́рити (свайку) — вогнать с головкой в землю (при игре в свайку); характерна нулевая ступень в корне каузатива, что указывает на неподвижную а.п. производящего глагола (а.п. a); вторичен, по-видимому, переход в а.п. b₂: схрв. диал. pf. (Далмация) núriti, praes. 1.sg. nûrīm 'пихнуть', nurati, praes.1.sg. -am 'толкать, пихать; проталкивать; совать, запихивать'. ‖ Frisk I, 514—515; Adams 272.

Драв. *nūṛ‿л, 'проникать, втискиваться, проползать сквозь узкий проход': юж.-драв.: тамил. nuṛai (-v-, -nt-) 'ползти, проползать сквозь узкий проход, проникать; впечатываться в память; напяливать на себя одежду, обувь; занимать должность; быть интерполированным', nuṛai (-pp-, -tt-) 'помещать, вставлять, интерполировать'; n. 'узкий путь; окно, отверстие, щель, расщелина; проём; пещера, полость'; nuṛuntu (nuṛunti-), nuṛutu (nuṛuti-) 'вставлять, втыкать, всовывать; стягивать, связывать (напр. кольцо волос), держать в месте, не свободно устроенном; уносить украдкой; притаиться, красться, выскальзывать крадучись, ускользать из поля зрения; ползти, уползать как рептилии'; nuṛunti 'тот, кто увиливает, ускользает от обязанностей, от ответственности'; nuṛuvu (nuṛuvi-) 'ползти, уползать (как рептилии)'; nuṛuval 'что-либо скользкое; ненадежное, неустойчивое'; nūṛai 'трещина, щель, скважина, дыра, отверстие, высверленное отверстие; вид решетчатого окна, пещера, впадина, полость'; noṛuntu (noṛunti-) 'вставлять, втыкать, вколачивать, вонзать'; малаялам nuṛayuka 'вползать, протискиваться'; nuṛa-vātil 'маленькая, узкая дверь'; nūṛuka, nūḷuka 'вползать, протискиваться, входить, проникать тайком или с трудом'; ? nūṭṭa 'брешь, пролом в заборе, в ограде', ? nokkuka 'проходить сквозь, продевать, пронзать, протыкать, прокалывать; пробуравливать, сверлить'; каннада noṛe 'вползать и т.п.'; nurgu, nuggu 'входить, проходить под нажимом, протискиваться; вламываться, пробиваться; входить без разрешения (в дверь, в дом), вползать внутрь (в отверстие, в нору)'; nuggisu 'заставлять проползать или протискиваться'; nugicu 'вылезать (из руки и т. п.) пробираясь (как птица), ускользать, выскальзывать (из руки, как на пример, веревка, нитка)'; nuguḷ 'входить, проникать в дверь или в маленький узкий проход или отверстие, входить с трудом, протискиваться, вползать'; nuṇ(u)cи, nuṇасu 'ускользать, ускользнуть (например, из рук), уйти (незаметно), проходить (незаметно), скользить, двигаться украдкой или отодвигать'; nusi 'входить в дверь и т.п.'; nusuḷ 'входить в дверь или в узкое отверстие, удаляться, прятаться, скрываться, идти крадучись, красться, уходить'; n. 'входная дверь или маленькое, узкое отверстие; искусная игра в непрямое называние объекта; уклонение, увертка; уловка; ложь'; nusuḷisu 'заставлять войти в дверь'; nūḷ 'ложь, обман'; кодагу nugg- (nuggi-) 'входить силой или без разрешения; проходить; влезть (в пещеру)'; тулу nurguni 'проходить мимо, не замечать', nūruni 'вползать, входить с трудом', nurumpè 'дыра, отверстие, нора', nuripāvuni 'вталкивать в дыру, в отверстие, всовывать' ‖ андхра: телугу (K) nusulu 'отпрянуть, отскочить; сморщиться; избегать, уклоняться; уходить; прятаться, скрываться; говорить ложь'; n. 'сжимание всех частей тела; скрывание, прятание'; (K) nusalu 'избегать, уклоняться, увертываться; лгать'; (K) nusũgu 'идти

крадучись; ускользать; бежать из заключения' ‖ центр.-драв.: колами *no·laŋg-* (*no·laŋkt-* '(змея) ползет'); конда *ṛug-* (*-it-*) 'входить, проникать; незаметно войти, вкрасться; втиснуться (например, в одежду), войти в проход'; пенго *ṛug-* (*rukt-*) 'прятаться, скрываться', *ṛuk-* (*-t-*) tr. 'прятать, скрывать'; манда *ṛug* intr. 'прятаться, скрываться', *ṛuk-* tr. 'прятать, скрывать'; куи *ṛupka* (< *ṛuk-p-; *ṛukt-* 'вталкивать, всовывать между чем-либо; вставлять, приспосабливать, пригонять что-либо между двумя поверхностями; скручивать что-либо в комок; вставлять'; *rūḍa* (*rūḍi-*) 'ползти под, подползать, вползать', *ṛonda* (*ṛondi-*) 'проскользнуть', pl. action *ṛotka* (*ṛotki-*); куви (Sunkarametta) *ṛug-* (*-it-*) intr. 'прятаться, скрываться', *ṛuk-* (*-h-*) tr. 'прятать, скрывать', (F.) *rugali,* (S.) *lugnai* 'прятаться', (F.) *rūga* 'тайно'; *rūkhali* tr. 'прятать, скрывать' ‖ сев.-вост.: курух *nuṛnā* (*nuḍḍas*) 'скрывать, прятать, утаивать, маскировать, скрывать от взгляда, закрывать покрывалом', *nūkhrnā* 'скрываться, прятаться; утаиваться', *nul^ugnā* 'помещать какую-л. вещь в другую скользящим толчком, вталкивать, втыкать или всовывать в или под что-то длинное', *nulg^urnā* 'входить в или под, направляться вперед, вползать'; малто *nuḍe* tr. 'прятать, скрывать, утаивать, укрывать', *nuḍgre* intr. 'прятаться, скрываться; исчезать'; *lulqe* 'вталкивать, всовывать, вонзать'; *lulqre* 'исчезать, пропадать, скрываться (в толпе, в зарослях, в джунглях)', *lulgòre* 'прятаться, скрываться' ‖ См. DED 250 (№ 3076), DEDS 78. [Ср. DRAVET 1048: прадрав. *nūṛ-[ug-]*, праюж.-драв. *nūṛ-*, прагонди-куи *ṛug-*, прасев.-драв. *nuḍ-*, колами *no·laŋg-* возведено к праколами-гадба *nōl-*, которое оставлено без этимологии].

<Алт. *nűRu* (~ -o) 'жилище, убежище': пратюрк. *jū/uR-t* '«большой дом» (жилище и прилегающая территория)', чув. *śort* 'дом и надворные постройки', як. *surt* 'жилище, стойбище', тув., тоф. *čuˊrt* 'стойбище, стоянка, селение, жилье, страна', хак. *čurt* 'жилище, постройка, хозяйство'; др.-тюрк. *jurt* 'опустошенная территория становища', в более поздних текстах — 'место жительства', халадж., туркм. *jūrt* 'жилище, дом, очаг; страна, край; родина'. ‖ EDT 958, VEWT 211, TMN 4, 212—213, ЭСТЯ 1989, 254—255, СИГТЯ 2000, 490—491, 494, Федотов 2, 144—145, Дыбо 2008, 223 (и подробный анализ семантики 246—261). Формант *-t* в тюркской основе, видимо, можно счесть десемантизованным показателем множ.; ср. имеющуюся в древнетюркском форму *tarmut* от *tarym* 'дельта реки' (т.е. не от обозначения лица!) Erdal 1991, 82—83. Что касается фонетики, здесь редкий случай рассогласования по долготе сибирских и огузских форм; позиция перед смычным не дает возможности установить характер пратюрк. сонорного: *-r-* или *-ř-*. ‖ тунг. *nora-* ' находить убежище': маньчж. *noro-* 'находиться на одном месте; находить приют, убежище, укрываться; гнездиться', *noro(-n)* 'местопребывание, убежище, нора, гнездо' ‖ ССТМЯ 1, 606, Захаров 239. Вряд ли сюда сопоставляемое с ностр. корнем в NED эвенк. *niru ~ nirú* 'ямка, впадина' ССТМЯ I, 600, не подходящее ни семантически, ни фонетически. ‖ кор. *nùrí* 'мир, обиталище': ср.-кор. *nùrí*, кор. *nuri* (арх., книжн.) Nam 114, KED 357, БКорРС I 256 ‖ Ср. EDAL 1000. В EDAL в эту этимологию включено еще фантомное монг. *niruγu* '1 общее состояние чего-л. 2 основа, поверхность' — в действительности оно состоит а) из одного из типов стандартной многозначности монг. *niruγu* 'спина, позвоночник': ср. монг. *niriγun:* ср.-монг. SH *niri'un, niru'un* 'позвоночник, спина'; *niri'un* Tk 35а 'позвоночник, спинной мозг'; *nirisun* 'spine' Hua-i 47, *gerün niri'un* 'house top' Hua-i 16, *niru'un* Yy 78b; Лейд. 72 *nirūn* 'спина'; Ст.сл. 57 *niryun* 'dos, reins', ИМ 442 *nirgun* 'спина', письм.-монг. Kow. 672 *niruγun* 'середина спины, крестец', Kow. 673 *niruγu üje* 'спинные позвонки', Kow. 704 *nüru* 'поясница (просторечн.)' (халхас. извод); сев.-монг.: халха *нуруу(н)* 'спина, позвоночник; горный хребет; *ноён нуруу* 'а) конёк, главная балка, верхняя поперечина (крыши); б) становой хребет, 4) хребет'; *уул нуруу* 'горный хребет'; 5) матица, перекладина; *дам нуруу* 'балка'; *майхны нуруу* 'верхняя палка палатки'; *ноён нуруу* 'конёк (крыши)'; *хаалганы нуруу* 'верхний поперечный брус ворот'; *хавирга нуруу* 'балки, идущие

параллельно’; *хавхны нуруу* ‘сжимающиеся части капкана’; *нуруу мод* ‘матица’; 6) ‘скирда’; *гурван нуруу өвс* ‘три скирды сена’; 7) ‘общее состояние чего-л.; общие соображения; очерк, набросок, общий план, эскиз’; *гол нуруу* ‘стержень, ось; основа, центр, самая главная часть’; *дүн нуруу / гол нуруу* ‘общее состояние; набросок, черновик, план’; старо-бур. Castren Bur. 146 *ńuruguŋ* нижнеудинск., *ńurguŋ* тункинск., *ńurgun* хоринск., *ńurū* селенгин. ‘Rücken’, бурят. *нюрган* ‘спина, позвоночник; хребет (горный); балка, матица; рост, возраст; сезон’; (заимств. из письм.-монг.) *нюруу* ‘поверхность чего-л. (земли, воды), гребень стога, волны’, калм. KWb 281b *nurgʋn* ‘Rücken’, (Муниев) ‘спина, позвоночник, хребет, верхняя часть, поверхность, возвышение; рост, возраст’; Тод. Джанг. 347 *нурхн* ‘спина, стан’; ордос. Mostaert DO 502a *nurū* ‘спина, позвоночник; поясничный отдел позвоночника животного; балка, перекладина; хребет горы, вершина холма’; *nurū-la-* ‘поддерживать в к.-л. деле, брать на себя ответственность’; внутр.-монг. (Тод. ЯМВМ 176: хорч., джал., дурб., горл., архорч., бар., онн., найм., хеш., харч., тум., шгол., уцаб., чах., орд.) *нурӯ* ‘спина, позвоночник’; южн.-монг.: монгор. SM 291 *nuru* ‘dos’ ‘спина; верхняя часть (конек крыши, спинка ножа); горный хребет; период’, Тод. Монгор. *нуру* ‘спина’; дунс. Тод. Дунс. 131 *нурун* ‘спина’; дагур. Тод. Даг. 158 *нирō* ‘спина, позвоночник; поясница, стропила’; б) монгор. *nire-* ‘подогнать, закрепить (дверь; крышку; рукоятку к ножу)’ (SM 278), которое Смедт и Мостер сравнивают с могольским *nira-* ‘placer’ (помещать) и письм.-монг. Kow. 644 *nereku* ‘влить в котел воды, наполнить водой сосуд, набить трубку’, халха *нэрэх* ‘набивать трубку табаком’; фонетически и семантически этот глагол не может иметь отношения к вышеприведенным формам со значением ‘поверхность, основа’; к внешним отношениям ср. маньчж. *нэрэ-* ‘накинуть, набросить что-л. на плечи, накрыться’ ССТМЯ 1, 625.

◊ Ср. Иллич-Свитыч МС 357 (*nurʌ ‘проникать[2]’: и.-е. *ner- ‘проникать, нырять’ ~ драв. *нuṛʌ- ‘вползать, проникать’). Ср. NED 1564 *ñur’V ‘to penetrate’ (те же параллели; ср. в таком случае несоответствие и.-е. состава корня драв. огубленному вокализму). Ср. NOSTRET 470 (то же).

Две приведенные этимологии были разделены в процессе работы над картотекой. Объединение их мыслилось В. М. Иллич-Свитычем прежде всего в связи с индоевропейским материалом, который в принципе допускал бы толкование в духе чередования сонантов в корне, как *bheu̯g-/*bheg^u ‘бежать’ ОСНЯ № 15 (ср. еще: 1) и.-е. *kei̯u̯-/keu̯- ОСНЯ № 209; 2) и.-е. *leu̯p-/*lep- ОСНЯ № 268; 3) и.-е. *bher-/*bheu̯r- ‘буря, бушевать’ ОСНЯ № 23). Однако весь набор сопоставлений заставляет формально разделить корни; значительное семантическое сходство по рефлексам можно объяснять древними контаминациями.

Ностр. *p-

1. puḥE ‘дуть’: с.-х. *p₁wḥ* ‘дуть’ ~ картв. *pu-* ‘дуть, (?) кипеть, надуваться’ ~ и.-е. *peu̯-* ‘тяжело дышать; пениться’, *peu̯ə-* ‘дуть, веять’ ~ урал. *риɣʌ* ‘дуть’ ~ алт. *pʻü̆-r-* ‘дуть’.

(?) С.-х. *p₁wḥ* ‘дуть; распространяться (о запахе)’ ‖ семит. *pwḥ* ‘распространяться (о запахе)’: араб. *fwḥ* (perf. *fāḥa*), сирийск. *pwḥ* (perf. *pāḥ*) ‘распространяться (о запахе)’, тигре *fəḥot* ‘запах, вонь’ [LH 654]. Отметим также семит. *pwḥ* ‘дуть’ (> араб. *fwḥ* ‘дуть с шумом (о ветре)’, др.-еврейск. *pwḥ* ‘веять’, арамейск., сирийск. *pwḥ* ‘дуть’), которое скорее всего является собственно семитским дескриптивным образованием либо фонетически преобразовано (*ḥ* на месте *h*) в силу своей дескриптивности. Ср. Ges. 636, Brockelmann Syr. 559, SEMET 1169 (по формальному признаку, невзирая на семантику, все амбивалентно тол-

куемые формы, такие как др.-еврейск. *pwḥ* 'веять', арамейск., сирийск. *pwḥ* 'дуть', отнесены к корню *pwḥ). ‖ бербер.: отметим зенага /ə/ff- 'дуть, надувать', perf. 3 m. sg. *jaf* 'он подул', 1 sg. *åff-əg* 'я подул' (Nicolas 188). ‖ чад. *ph/u/Hλ 'дуть, пахнуть': зап.-чад. *fayVḥ- 'дуть; пахнуть', ангас *fwēr* 'пахнуть'; рон: боккос (Jungraithmayr Ron) *fu'*, perf. *fu'í*, habit. *fwáà'* 'дуть', даффо-бутура (Jungraithmayr) *fu'*, perf. *fù'an*, habit. *fwáà'* 'раздувать (огонь), играть на дудке', (?) ша *fuḍ* (< *fu'* + *-t-* ?) 'раздувать огонь'; *fya'ḥ* 'дуть'; центр.-чад. *ʔi-fiyaH- 'пахнуть', логоне *ʔìfiyàʔà* [By] 'пахнуть'; вост.-чад. *PuH- 'дуть': тумак *pò*. Данные языков рон, по мнению В. М. Иллич-Свитыча, могут указывать на чад. *u. См. Hoffmann Bura 263, 266, Hoffmann Margi 119, 122, 161, Иллич-Свитыч АСб.21, Jungraithmayr Ron 141, 214, 284. Здесь чадские формы выбраны в основном вслед за О. В. Столбовой (CHADET, WCHET 1051, CCET 919, ECHET 630); согласно ее этимологическим решениям ангас (Greenberg) *fȋ* 'дуть'; сура (Jungraithmayr Sura) *fȋ̃*, habit. *fyā* 'дуть, раздувать', анкве *fȋ̃* 'дуть'; рон: ша *fuḍ* (< *fu'* + *-t-* ?) 'раздувать огонь'; центр.-чад. логоне (Nachtigal) *fa* 'дуть', гульфей (Adolf-Friedrich) *fȋ̃* 'ветер'; марги (Hoffmann) *fȋ̃* 'раздуваться', бура (Hoffmann) *fia* 'дуть (о человеке, ветре)', гисига (Lukas Gisiga) *fe, ffə* 'играть на духовом инструменте', которые включает в данную этимологию В. М. Иллич-Свитыч, относятся к другому с.-х. корню со значением «дуть»: *fiʔ-. Формально это решение не может быть обосновано для ангас, сура, анкве, ша, в которых и смычка, и *ḥ* дают нулевой/глайдовый рефлекс (Stolbova 1996, 131); в центрально-чадских для *ḥ* следует ожидать фрикативного *-x-*. На каком основании О. В. Столбова включает в рассматриваемую этимологию, а не в *fiʔ- 'дуть', перо *púyò* 'дуть' (ср., впрочем, огубленный вокализм!), варджи *fày-*, карийа *fày-*, мийа *fay-*, мбурку *fày-*, нгизим *fiyə* 'дуть', неясно. По-видимому, тут возможны альтернативные решения. ‖ кушит.: центр.-кушит. (агав.) *fi/ahw-, хамир *faw* 'отдыхать', *faû-t* 'дышать'; квара *fȋhŭ*, *fȋû* 'дуть; отдыхать'; кемант *fȋw* 'дышать', *fiwā* 'душа, дыхание'; южно-кушит. *faḥ-: алагва *faḥ-* 'дуть'. См. AGW 233, SCU 151. ‖ Корень дескриптивный, и поэтому генетическое толкование корней спорно. Ср. в AFAZET еще корни с близким значением: *fiwaḥ- 'blow, breathe'; *fu(h/wV)k̲- 'breathe (heavily)'. Ср. HSED 813, 814, AFAZET 1650. В NED № 1673 выборка семито-хамитских сравнений — вслед за В. М. Иллич-Свитычем, а не за AFAZET.

Картв. *pu-* 'дуть, (?) кипеть, надуваться' ‖ грузин. *puw-* 'подниматься (о тесте)', *puwn-* 'заквашивать (тесто)' ‖ мегрел. *pu-* 'кипеть', *pu-n-* 'кипятить', чан. *pu-* 'кипеть' ‖ сван. *pw-* 'надуваться, подниматься', масдар *li-pw-e*; *pū-l* 'дуновение' (с суффиксом *-l*), *pwin-*, *pǚjn-* 'надувать, заквашивать' ‖ лаз. *pu-*; *pun-* 'заквашивать хлеб'. См. Кл. 192, EWK 361.

И.-е. 1. *peu̯-, *pou̯- 'дышать тяжело; пениться' ‖ армян. *(h)ogi* (< *pou̯io-) 'дыхание, дуновение, душа', *heval-* (< *peu̯ā-) 'учащенно или тяжело дышать' ‖ сред.-ирланд. *ūan* 'пена', валлийск. *ewyn*, бретон. *eon*, pl. *eien* 'источник' (< *pou̯-ino-), ст.-бретон. *euonoc* 'пенящийся' ‖ ?? др.-инд. *phŭt-karóti* 'дует, тяжело дышит, кричит во все горло' (если это не звукоподражание, возникшее уже на инд. почве) ‖ См. Pok. 847. Корень с точки зрения структуры выглядит странно (нет ларингала при чисто сонантном завершении); дескриптивный? 2. *peu̯ə- 'чистить, очищать от мякины < провеивать' ‖ санскр. *pávate, punā́ti*, inf., *pávitum*, ptc. *pūtá-* 'to make clean, clear, cleanse, purify'; *pavitár-, pavītár-, pótar-, potár-* m. 'purifier', *pavítra-* n. `means of purification, filter, strainer; Soma-Seihe, Sieb'; *pávana-* n. 'Sieb, Worfelkorb', *pāvana-* 'purifying', *pāvaká-* 'pure, clear, bright'; *pū́ti-* f. 'purity, purification'. Из индоарийских рефлексов особенно ср. Turner 8277 *punā́ti* 'purifies' RV > пракрит. *puṇaï* (pass. *puṇijjaï*) 'cleans, winnows'; зап. пахари *pūṇṇõ* 'to winnow'; кховар *phunik* 'to winnow, throw up'; Turner 8320 *pūtá* 'cleaned' RV > синдхи *pū karaṇavā* 'to winnow'; Turner 7843 *paripavana* n. 'cleaning, winnowing' Kull., 'winnowing basket' Nir. > вайгали *parén* 'sieve', пашаи *päerḗn, parín, parūn*, гавар-бати *pairūn*; кховар (Lor.) *pλrūn* 'sieve for flour (made of skin strips)', баш-

карик *páreŋ*, торвали *pōén*, пхалура *parūŋ*; кашмири *paryupu* m. 'sieve, strainer', лахнда *parūŋ* m., *paṛūn*, *prūŋ*, зап. пахари *priūṇi* f., синдхи *perahana*; Turner 7977 *pávana* n. 'sieve, strainer' AV. > пали *pavana* n. 'winnowing'. Ввиду предполагаемого первичного значения более очевидным становится отношение сюда санскр. *pavana* m. 'wind' MBh. > пракрит. *pavaṇa*, *payaṇa* m. 'wind', панджаби *pavaṇ*, *pauṇ* f. (*pauṇ-cakkī* f. 'windmill'); зап. пахари *pauṇ* f. 'wind, storm'; синдхи *pavan* 'wind'; зап. пахари поэт. *pŏ̃ṇ* f. (obl. *-a*) 'wind, storm', гархвали *pɔ̃ṇ* (Turner 7978). Ср. Mayrhofer EWA 2, 105—106 (приводимые там вслед за Й. Койву- лехто пра-индоиранские заимствования в финно-угорский фонетически сомнительны) ‖ Иран.: авест. *pūitika-* 'zur Läuterung dienend'; ср.-перс. *pavāg*, перс. *pāk* 'lauter, rein' ‖ (??) Греч. πτύον, новоаттич. πτεόν `Wurfschaufel', деминутив πτυάριον n. (Hdn., *EM*), ком- позит δίπτυον· Κύπριοι μέτρον, οἱ δὲ τὸ ἡμιμέδιμνον H. Frisk 2, 615—616 (предположен спонтанный переход *p-* > *pt-*), ??? πίτῡρα 'мякина' < *πύτῡρον вследствие диссимиляции Frisk 2, 545 ‖ герм. **fau̯-ja-*: д.-в.-н. *fowen*, ср.-в.-н. *voewen* 'sieben, Getreide reinigen' ‖ лат. прилагательное на *-r-*: *pūrus*, *-a* 'rein; lauter; reinigend; unbefleckt; schlicht; ohne Vorbehalt'; *pūrgāre* (др.-лат. *pūrigāre*) 'reinigen, sühnen, abführen, entschädigen, ebnen, aufräumen'; *pu- tus*, *-a* 'rein', *nepus* 'non pūrus' Paul. Fest. 165; *putāre* 'reinigen' ‖ кельт.: ср.-ирланд. *ūr* 'new, fresh', кимр. *ir* 'fresh, green' ‖ WP II 79, Pok. 847—848.

Урал. *риγл / ришл* 'дуть' ‖ мордов. эрзян. *puva-* 'дуть, подуть; раздуть, задуть (огонь)'; марийск. (лугов., вост.) *pue-* 'дуть (о человеке, ветре), трубить' Bereczki 219 ‖ ст.-венгер. *fu-* (= *fú-* ?) 'дуть (в трубу)', венгер. *fúj-*, *fu-* 'дуть, трубить' EWU 426 (с 1372 г.); обско-угорск. **рйγ-* (Honti), **puwV-* (Живлов 2006), общемансийск. **рйw-* (Honti), **puw-* (Живлов 2006) : мансийск. юж. (по Kannisto) *pŏ̇-*, (по Munkácsi) *puw-* 'дуть', зап. (сев.-вагильск.) *puβ-*, (ниж.-лозьвинск.) *pūβ-*, *pu-*, (сред.-лозьвинск.) *pūγ-*, *powγ-* 'дуть', сев. (сосьвинск.) *puβßi* 'ду- ет'; общехантыйск. (Steinitz) **pŏγ-*, (Honti, Живлов 2006) **puw-* 'дуть': хантыйск. вост. (Вах, Васюган) *pŏγ-*, (Терюган) *pŏγʷ-*, (Юган — по Paasonen'у) *pow-* [= *pŏw-* ?], юж. (В. Демьянка) *pŏu̯-* Honti 495 ‖ самодийск. **pū-* (? **puəj-*): ненецк. (тундров. — Салехард) *pūć* 'дуть (о че- ловеке, ветре)', (лесн. Нялина) *pu'ra·ś* 'быть унесенным ветром', энецк. (хантайск.) *faeŋabo*, (баишенск.) *fuasabo* 'дуть', нганасанск. *fual'i'éma*, *fuarúma* 'дуть'; селькуп. (тымск. — K. Don- ner) *pūab'* 'дуть'; камасинск. *p'и-* 'дуть (о ветре, человеке)', 1 sg. praes. *p'ий'l'em*, *p'uil'àm*; койбальск. *публя*; моторск. **hal-*, *халнамъ* 'дую', *халзы* 'надуваю', *халгамъ* 'пущаю дым', (?) *chásÿmaixa* 'пержу'. См. Janhunen SW 188—189, Hel. 255, 264, Leht. 368, K. Donner MSFOu XLIX. 176, Donner Kam. 55 ‖ См. UEW 411, Coll.12, Wichmann TschT 86, Ӱпымарий 170, Эпин 95, Steinitz OVok. 119, 126, Szin.149, 25. Конечный гласный, скорее, *-а*: на урал. *а*-основу указывают мордовские данные, в марийск. же *pue-* может отражаться и *а-*, и *е*-основа (см. Норманская 2008).

Алт. **p'ǚ-r-* 'дуть' ‖ тюрк. **ür-* 'дуть': чув. *və̂ʷr-*, як., долг. *ür-*, тув., тоф. *ür-* Рас. ФиЛ 239, хак. *ür-*; др.-уйг. *ür-*, крх.-уйг. *ür-* (MK), чаг. *ür-* (Pav. C.), *hür-* (словарь шейха Сулеймана, Кúnos 1902, 79, Радлов II 1810—1811); тур. диал. *ür-* 'надувать (выдохом)'; ст.-кыпч. *ür-* (AH, Houts.), караим. (крым., галицк., трак.) *ür-*, татар., башк. *ör-*, ног., казах., каракалп. *ür-*; телеут. *ür-* 'дуть', туба 'раздувать, дуть'; алт. диал. (какой?) *ǖr-* Радлов I 1825 (долгота, по-видимому, нефонологическая, из записи текста) ‖ EDT 195—196, ЭСТЯ 1, 635—636, Stachowski 253, Баск. Туба 163, Верб. 414 ‖ монг. **(h)uri-* 'дуть (в лицо)', **(h)ur-gi-* 'клубить- ся (о пыли)': письм.-монг. *uri-* (МХТТТ, БАМРС) 'дуть в лицо', халха *uri-* id., *urgi-* 'подни- маться, клубиться (о пыли)', *uŕ(-in)* 'теплый воздух, оттепель'; бурят. *urin* 'теплый воздух', *urjā-* 'клубиться (о пыли)'; ордос. *uri-* 'дуть (о свежем ветре летом)', *urin* 'сильный холод перед восходом солнца (зимой)' Mostaert DO 740 ‖ тунг. **pō-* 'дуть': эвенкийск. *hūv-*, (ер- богочен., илимпийск., вилюйск.) *hub-*, (подкаменно-тунг., сахалин.) *hup-* 'дуть (о ветре), раздувать (огонь), загасить (огонь)', эвен. *hū-* 'дуть', солон. *ūgū-* 'дунуть', негидал. *xūw-*,

ороч. *х̄й-*, ульч. *р̄й-* 'дуть', орок. *р̄й-* 'дуть', нанайск. *р̄й-* 'дунуть, подуть (ртом)'. Сев.-тунг. производное: *рō̆gi-n* 'ветер, метель': эвенкийск. *hū̆wū̆n, hиγип* 'вой ветра; ветер, пурга', эвен. *huḡī* 'буря, метель', солон. *ūḡīī* 'метель, буран' ‖ См. ССТМЯ 2, 336. По-видимому, основа в праалтайском оформлена словообразовательным отглагольным аффиксом интранзитива *-r-*, о котором см. EDAL 191—192 (в тунг. конечный сонант закономерно отпадает).

◊ Ср. Иллич-Свитыч МС 339, Кл.192 (картв. ~ и.-е.), Sauv. 26—27, Räs. 49 (урал. ~ алт.), NED 1673, NOSTRET 135. Придыхательный рефлекс начального согласного в алт. может быть обусловлен воздействием древнего ларингала. NED связывает тюрк. *bū̆g* 'пар' с тунг. *рō̄-* 'дуть'; по-видимому, тюрк. слово относится к отдельному ностратическому корню *bū̆kV* со значением 'дым, пар': и.-е. *bhōg-* 'жарить, печь' WP II 187 ‖ алт. *bū́kà* 'дым, пар' (EDAL 217: тюрк. *bū̆g* 'пар', монг. *baγa-gi-* 'дымить' < *baga-gi-* (по «правилу Владимирцова»), сев.-тунг. *bugar* 'гарь', кор. *p̆ăgil-* 'пузырящийся, кипящий', япон. *bák-* 'кипеть, вариться'[1]) ‖ драв. *poγ-* 'дым' DEDR 4210, DRAVET 1309 ‖ ср. предложенные в заглавную этимологию 'дуть' в NOSTRET эскимос. *puju-* 'дым, сажа', нивх. *boj-* 'дымить, курить' ‖. Предложенное в заглавную этимологию в NOSTRET алт. *p'ubá-ktV* 'мешок, мехи' в силу семантического отличия следует отклонить.

Список дополнительных сокращений

БАМРС — *Большой академический монгольско-русский словарь*. Т. 1—4. М., 2001.

БКорРС — *Большой корейско-русский словарь*. I—II. М., 1976.

Бурлак — Старостин 2005 — БУРЛАК С.А., СТАРОСТИН С.А. *Сравнительно-историческое языкознание*. М.: ИЦ «Академия», 2005.

Вайан — ВАЙАН А. *Руководство по старославянскому языку*. М., 1952.

Варлаам и Иосаф, XVI ~ XII–XIII вв. — Житие и жизнь преподобных отец наших Варлаама пустынника и Иосафа царевича Индийского. Творение преп. отца нашего Иоанна Дамаскина. СПб., 1887 (Изд. ОЛДП, 88)

Даничић Pj. II, 362 — ДАНИЧИЋ Ђ. *Рjечник из књижевних старина српских*, I—III. Београд, 1863—1864.

Дыбо 2008 — ДЫБО А.В. Материальный быт ранних тюрок. Жилище // *Природное окружение и материальная культура пратюркских народов*. М., 2008

Живлов 2006 — ЖИВЛОВ М.А. *Реконструкция праобскоугорского вокализма*. КД. М., 2006

Жит. Андр. Юр. XV–XVI в. — Житіе Андрея Юродиваго по нѣсколькимъ спискамъ XV и XVI вѣка. См. Свбд. и зам., № LXXXVII. (Сокр. см. в Срезн.)

ИМ — Монгольские глоссы в словаре Ибн-Муханны см. в MA.

Іон. II.6, Упырь — Книга прор. Іоны по сп. XV в., скопированному съ рукописи Упыря Лихого 1047 г. (Сокр. см. в Срезн.)

Кир. Тур. Пис. Вас. — Письмо Кирилла Туровского Василию, игумену Печерскому. См. Приб. тв. св. о., т. X, стр. 346 и слѣд. (Сокр. см. в Срезн.)

Мыреева ЭРС — МЫРЕЕВА А.Н. *Эвенкийско-русский словарь*. Новосибирск, 2004.

Муниев — МУНИЕВ Б.Д. *Калмыцко-русский словарь*. М., 1977.

Ник. лет. 1537 — Русская лѣтопись по Никонову списку (Полное собрание русских лѣтописей, т. X). (Сокр. см. в Срезн.)

Норманская 2008 — НОРМАНСКАЯ Ю.В. *Реконструкция прафинно-волжского ударения*. М., 2008.

Окружн. посл. мт. Фотия, XVI в. ~ 1416 г. — Окружное послание митр. Фотія до 1431г. (в спискѣ) // Рус. ист. библ., т. VI, № 59. (Сокр. см. в Срезн.)

Пар. 1271 г.- 1370 г. — Паремейникъ, Имп. публичной библ.; см. Др. пам., изд. I-е. (Сокр. см. в Срезн.)

---

[1] Яп. *bák-* 'кипеть, вариться' лучше, вслед за Мартином, сравнивать с кор. *p̆ăgil-* 'пузырящийся, кипящий', извлекая последний из этимологии EDAL 1823, алт. *p'ó[k]ù* 'опухать (о теле)'.

Псков. I лет. 7043 г. — Псковская первая летопись. // Полн. собр. Рус. лет., т. IV, стр. 173 и след. (Сокр. см. в Срезн.)

Пут. Генн. и Позн. — Путешествие архидьякона Геннадія и гостя Василія Позднякова въ Святую землю, по Копенгагенскому сп. XVIIв. См. Свбд. и зам., т. II, стр. 355. (Сокр. см. в Срезн.)

СИГТЯ 2000 — *Сравнительно-историческая грамматика тюркских языков: Лексика.* М., 2000.

Срезн. — СРЕЗНЕВСКИЙ И.И. *Материалы для словаря древнерусского языка по письменным памятникам.* Т. I—III. СПб., 1893—1912.

Упырь — Списки съ рукописи Упыря Лихого 1047г. Моск. синод. библіотеки и библіотеки Имп. Акад. Наук.

Федотов — ФЕДОТОВ М. Р. *Этимологический словарь чувашского языка.* Т. 1—2. Чебоксары, 1996.

Флавий Полон. Иерус. XVI ~ XI — La prise de Jérusalem de Joséphe le Juif. Par V. Istrin, t. 1. Paris, 1934. XI в., сп. XV–XVI вв., t. 2, Paris, 1938.

Шимкевич 159 — ШИМКЕВИЧ Ф.С. *Корнеслов русского языка, сравненного со всеми главными славянскими наречиями и 24 иностранными языками.* СПб., 1842.

ЭСТЯ 1989 — *Этимологический словарь тюркских языков: Общетюркские и межтюркские основы на буквы «Ж»,* «Ж», «Й» / Авт. сл. статей Э. В. СЕВОРТЯН, Л. С. ЛЕВИТСКАЯ. М., 1989.

Adams — ADAMS Douglas Q. *A Dictionary of Tocharian B.* Amsterdam — Atlanta, GA, 1999.

AGW — Этимологическая база данных центрально-кушитских языков, составлена А. Ю. МИЛИТАРЕВЫМ / www.starling.rinet.ru

AFAZET — Этимологическая база данных афразийских языков, составлена А. Ю. МИЛИТАРЕВЫМ и О. В. СТОЛБОВОЙ / www.starling.rinet.ru

Belostenac II — BELOSTENAC, J., *Gazophylacium.* I–II. Zagrabie, 1740//1972.

Bereczki 1992, 1994 — BERECZKI G. Grundzüge der tscheremissischen Sprachgeschichte I, II. // *Studia Uralo-altaica,* 34, 35, 1992—1994.

Castren Bur. — CASTRÉN A. *Versuch einer Burätischen Sprachlehre.* SPb., 1857

CCET — Этимологическая база данных центрально-чадских языков, составлена О. В. СТОЛБОВОЙ / www.starling.rinet.ru

CHADET — Этимологическая база данных чадских языков, составлена О. В. СТОЛБОВОЙ / www.starling.rinet.ru

DEDR — BURROW Th., EMENEAU M. *A Dravidian etymological dictionary.* 2ⁿᵈ ed. Oxford, 1984.

DRAVET — Этимологическая база данных дравидийских языков, составлена Г. С. СТАРОСТИНЫМ / www.starling.rinet.ru

ECHET — Этимологическая база данных восточно-чадских языков, составлена О. В. СТОЛБОВОЙ / www.starling.rinet.ru

EDAL — STAROSTIN S., DYBO A., MUDRAK O. *Etymological Dictionary of the Altaic Languages.* Vol. I — III. Leiden, Boston, 2003.

EDT — CLAUSON G. *An Etymological Dictionary of Pre-Thirteenth- Century Turkish.* Oxford, 1972.

Erdal (1991) — ERDAL M. *Old Turkic Word Formation.* I–II. Wiesbaden, 1991.

EWK — KLIMOV G. *Kartvelian Etymological Dictionary.* B. / NY, 1998.

Gebauer III — Jan GEBAUER. *Historická mluvnice jazyka českého.* Díl III: Tvaroslový, II. Časováné. Praha, 1958.

Hel. = Helimski 1997 — HELIMSKI E. *Die Matorische Sprache.* Szeged, 1997.

Henning Khwar. 16 — *A Fragment of Khwarezmian Dictionary* by W. B. HENNING, edited by D. N. MACKENZIE. London, 1971.

Hilmarsson 1986 — HILMARSSON, Jörundur. *Studies in Tocharian Phonology, Morphology, and Etymology: with special emphasis on the o-vocalism.* Reykjavik, 1986.

Hoernle 1916 — HOERNLE, A.F.R. *Manuscript Remains of Buddhist Literature Found in Eastern Turkestan,* vol. I. Oxford, Oxford University Press, 1916.

HSED — OREL V., STOLBOVA O. *Hamito-Semitic Etymological Dictionary. Materials for a reconstruction.* NY, 1995.

JLTT = MARTIN S. E. *The Japanese Language Through Time.* New Haven — London, 1987.

Karulis — Konstantīns KARULIS. *Latviešu etimoloģijas vārdnīca.* Rīga, Avots, 1992, 2001.

KED — MARTIN S.E., YANG HA LEE, SUNG-UN CHANG, *A Korean-English Dictionary.* New Haven — London, 1967.

Kúnos 1902 — KÚNOS I. *Šejx Sulejman Efendi's Čagatay-Osmanisches Wörterbuch.* Budapest, 1902.

Lig. VMI — LIGETI L. Un vocabulaire mongol d'Istanboul // *AOH*. T. XIV, fasc. 1—2. Budapest, 1962.

LKŽ — *Lietuvių kalbos žodynas* (2as leidimas) T I–XX. Vilnius, 1968—2002.

Mayrhofer EWA — MAYRHOFER M. *Etymologisches Woerterbuch des Altindoarischen. Heidelberg*, 1986. Bd. I; 1996. Bd. II.

Mažiulis 3: 203 — Vytautas MAŽIULIS. *Prūsų kalbos etimologijos žodynas*. 3 L-P, Vilnius, 1996.

Nam — NAM Kwang U. *Koe sacen*. Seoul, 1960.

NdŽ — *Lietuvių rašomosios kalbos žodynas. Lietoviškaivokiška dalis*. I–IV t. / Sudarė M. Niedermann'as, A. Senn'as, F. Brender'is ir A. Salys. Heidelberg, 1932—1963.

NED — DOLGOPOLSKY A. *Nostratic Etymological dictionary*. Leiden — Brill, 2009.

NOSTRET — Этимологическая база данных ностратических языков, составлена С. А. СТАРОСТИНЫМ / www.starling.rinet.ru

Plet. — PLETERŠNIK M. *Slovensko-nemški slovar*. Ljubljana, 1894 — 1895. D. I–II.

SCU — Этимологическая база данных южнокушитских языков, составлена А. Ю. МИЛИТАРЕВЫМ / www.starling.rinet.ru

SED — MILITAREV A., KOGAN L. *Semitic etymological dictionary*. Münster, I — 2000, II — 2005.

SEMET — Этимологическая база данных семитских языков, составлена А. Ю. МИЛИТАРЕВЫМ и О. И. СТОЛБО-ВОЙ / www.starling.rinet.ru

SSA — *Suomen Sanojen Alkuperä Etymologien Sanakirya*. Helsinki, 1992

Stachowski — STACHOWSKI M. *Dolganischer Wortschatz*. Kraków, 1993.

Stang — STANG Chr. *Das slavische und baltische Verbum*. Oslo, 1942

Starostin G. 2004 — см. DRAVET

Stolbova 1996 — STOLBOVA O. *Studies in Chadic comparative Phonology*. Moscow, 1996

Stolbova 2005 — STOLBOVA O. *Chadic lexical database*. I. Moscow — Kaluga, 2005

Stolbova 2007 — STOLBOVA O. *Chadic lexical database*. II. Moscow — Kaluga, 2007

Takacz EDE — TAKACZ G. *Etymological dictionary of Egyptian*. Leiden — Brill . I. 1999. II. 2001. III. 2008.

TMN — DOERFER G. *Türkische und mongolische Elemente im Neupersischen*. Wiesbaden, 1963. I; Wiesbaden, 1965. II; Wiesbaden, 1967. III.

UEW — REDEI K. *Uralisches etymologisches Wörterbuch*. Budapest, 1986-1989.

VEWT — RÄSÄNEN M. *Versuch eines etymologisches Wörterbuchs der Türksprachen*. Helsinki, 1969.

WCHET — Этимологическая база данных западно-чадских языков, составлена О. И. СТОЛБОВОЙ / www.starling.rinet.ru

Yamamoto 1969 — YAMAMOTO, K. *A Classified Dictionary of Spoken Manchu*. Tokyo 1969.

The paper is the first in an intended series of publications of Nostratic etymologies that had originally been prepared for the 3rd volume of V. M. Illich-Svitych's "Comparative Dictionary of the Nostratic Languages" (1971–1984) but could not be included for technical reasons. Where necessary, we also provide information on how the proposed etymologies relate to more recent corpora for daughter branches of Nostratic (Uralic, Dravidian, Kartvelian, Afro-Asiatic), published after 1984, as well as to the independently prepared Nostratic Etymological Dictionary by Aharon Dolgopolsky.

А. И. Коган
Институт востоковедения РАН, Москва

К вопросу о языке несанскритских фрагментов памятника Mahānayaprakāśa.
Ранний кашмири или поздний апабхранша?

Кашмири — единственный дардский язык, имеющий письменность и литературу. Однако ввиду того, что письменная традиция на этом языке достаточно молода, применение филологического метода при его диахроническом изучении едва ли может дать ощутимые результаты. Со стороны исследователей неоднократно предпринимались попытки обнаружить аутентичные старокашмирские тексты. Вопрос о наличии таких текстов все еще остается открытым, однако разыскания в данной области привели к появлению некоторых интересных проблем. Одна из таких проблем — проблема атрибуции языка несанскритских фрагментов шиваитского философского трактата Mahānayaprakāśa — рассматривается в данной статье. Автор приходит к выводу, что этот язык, несомненно, будучи индоарийским, никоим образом не может являться ранней формой современного кашмири. Более того, он представляет собой разновидность позднесреднеиндийского литературного языка (апабхранша), т. е. является искусственным образованием и, таким образом, не может быть предком ни для одного живого языка.

Ключевые слова: кашмири, дардские языки, индоарийские языки, санскрит, апабхранша, искусственный язык.

Язык кашмири, как известно, является единственным языком дардской группы, обладающим письменной традицией. Значение этого факта, однако, не следует переоценивать. История кашмири достаточно надежно прослеживается по письменным памятникам лишь на протяжении максимум трех последних веков. Не исключено и существование более ранних текстов (XVI и, возможно, даже XIV в.), главным образом поэтических, однако эти последние в течение столетий бытовали исключительно в устной форме и были записаны гораздо позднее на языке, не сильно отличающемся от кашмири XX в. [Захарьин 1978, 88; Коган 2005, 208]. Учитывая, что дардская языковая общность распалась не менее 3 тыс. лет назад, а возможная восточнодардская (охватывающая кашмири, шина, пхалура и кохистанские языки) — по-видимому, более 2 тыс. лет назад[1], следует признать практически полную непригодность филологического метода для исторического и сравнительно-исторического изучения языка кашмири. Данная ситуация, однако, представлялась и представляется аномальной целому ряду исследователей-кашмиристов. По этой причине уже не одно десятилетие активно обсуждается вопрос о возможности наличия старых памятников кашмирской письменности. С сожалением сле-

---

[1] Данные цифры основываются на результатах глоттохронологических подсчетов, приведенных в работе [Коган 2005]. Предварительная датировка распада восточнодардской языковой общности — около 2,5 тыс. лет — вытекает из долей совпадений в стословном списке между кашмири и языками шина (60%) и пхалура (58%). В дальнейшем, как представляется, уточнить данную датировку, возможно, позволит привлечение для лексикостатистического анализа материала кохистанских языков.

дует констатировать, что в настоящее время эта проблема, выдвинувшись на передний план, отвлекла внимание многих ученых от собственно сравнительно-исторических проблем[2]. Именно с этим обстоятельством следует связывать явный застой в диахроническом изучении языка кашмири в последние годы.

Хотя вопрос о наличии древних памятников кашмири до сих пор остается открытым, разыскания в этой области привели к появлению некоторых интересных проблем. Пожалуй, самая значительная из них — проблема атрибуции языка несанскритских фрагментов памятника Mahānayaprakāśa. Mahānayaprakāśa — религиозно-философский трактат, созданный в традиции кашмирского шиваизма. Языком основной части текста является санскрит, однако памятник включает также 94 стихотворные вставки, написанные на языке, отличном от санскрита. Каждая такая вставка снабжена обширным санскритским комментарием, который обычно оказывается весьма полезным для понимания несанскритских стихов и, в частности, для установления значения многих неясных слов, встречающихся в последних. Надежная датировка текста отсутствует, хотя на этот счет высказывались разные точки зрения. Наиболее обоснована из них гипотеза, предложенная Дж. Грирсоном [Grierson 1929, 74], согласно которой памятник следует датировать концом XV в. Основным аргументом Дж. Грирсона является наличие другого текста Bālabodhinīnyāsa, полностью санскритоязычного, имя автора которого (Шитикантха) совпадает с именем автора Mahānayaprakāśa. В тексте Bālabodhinīnyāsa имеется указание на приблизительную дату написания: сообщается, что памятник был написан в годы правления кашмирского султана Хусейн-шаха, взошедшего на трон в 1482 г. Разумеется, данная точка зрения является сугубо гипотетической и в принципе может быть пересмотрена в будущем, поскольку нельзя исключить наличие двух авторов-тезок, живших в разное время. Однако нельзя не признать, что на сегодняшний день гипотеза Дж. Грирсона является единственной, имеющей под собой какие-либо веские основания. Альтернативная точка зрения, предложенная индийским лингвистом С. К. Чаттерджи [Chatterji 1963, 258] и относящая рассматриваемый текст к XIII в., представляется нам безосновательной. Автор не приводит ни бесспорных контраргументов, исключающих грирсоновскую датировку, ни весомых доводов, подтверждающих его собственную (XIII в.). Между тем точка зрения С. К. Чаттерджи широко распространилась в современной литературе и, к сожалению, проникла даже в энциклопедические издания.

Главным предметом полемики послужила, однако, генетическая принадлежность языка несанскритских фрагментов. Дж. Грирсон, первым начавший глубокое изучение этого языка и посвятивший ему специальную работу [Grierson 1929], полагал, что в Mahānayaprakāśa отразилось наиболее архаичное (из известных нам) состояние языка кашмири. Ему вторили многие индийские исследователи, в частности, уже упоминавшийся выше С. К. Чаттерджи. Однако данная точка зрения разделялась не всеми. Так, известный французский индолог Ж. Блок утверждал, что язык Mahānayaprakāśa не есть кашмири в собственном смысле слова [Bloch 1934, 15]. Сам Дж. Грирсон указывает, что многие слова в тексте Mahānayaprakāśa выступают в стадии апабхранша [Grierson 1929, 77]. Язык памятника характеризуется при этом как переходная форма от апабхранша к современному кашмири [Grierson 1929, 73].

Взгляд на язык Mahānayaprakāśa как на одну из форм позднейшего апабхранша пользуется в настоящее время определенной популярностью. Следует, однако, иметь в

---

[2] Подобное положение вещей возникло, на наш взгляд, во многом из-за того, что большинство исследователей кашмири, являясь индологами, едва ли могли не перенести в кашмиристику традиции исторической индологии, где филологический метод является основным.

виду, что термин «апабхранша» понимается сегодня во многом иначе, чем в начале XX в. Дж. Грирсон обозначал им форму позднего среднеиндийского, по его мнению, вполне отражавшую реальный народно-разговорный язык соответствующей эпохи. В соответствии с этим для каждого новоиндийского языка предполагалось наличие своего собственного апабхранша[3]. К настоящему времени этот взгляд признан необоснованным [Елизаренкова 2004, 125; Tagare 1948, 4]. В современной индологии признается многозначность термина «апабхранша» [Елизаренкова 2004, 123; Bubenik 1996, 16; Tagare 1948, 1–5]. При этом чаще всего данным термином обозначают три конкретных литературных языка Северной Индии, во многом утративших связь с живыми диалектами и, бесспорно, не тождественных языку Mahānayaprakāśa[4]. В случае же с последним понятие «апабхранша», по-видимому, употребляется Дж. Грирсоном и другими исследователями в более общем значении — позднесреднеиндийский язык вообще. Принимая во внимание данный факт, не следует забывать, что значительная часть текстов среднеиндийского периода, включая и поздний, написана на стилизованных, по существу искусственных языках, произведенных из санскрита согласно определенным правилам пересчета, описанным грамматистами[5]. Поэтому одним из потенциально возможных решений проблемы языка Mahānayaprakāśa (и это обстоятельство ни в коем случае нельзя упускать из виду) является его трактовка именно как искусственного языка, который, разумеется, не только не является предком для современного кашмири, но и вообще не может иметь живых языков-потомков.

Как уже говорилось, взгляды Дж. Грирсона на генетическую принадлежность языка Mahānayaprakāśa отличаются противоречивостью. Неоднократно называя этот язык «старым кашмири», он вместе с тем отмечает, что почти вся лексика несанскритской части памятника является индоарийской, то есть восходит к определенным древнеиндийским прототипам [Grierson 1929, 77]. Данное утверждение, хотя и находится в неразрешимом противоречии с остальными идеями Дж. Грирсона, представляется нам весьма близким к истине. Действительно, в указателе, приложенном к работе [Grierson 1929], огромное большинство слов (об исключениях будет сказано ниже) могут быть выведены из прототипов, засвидетельствованных в древнеиндийских текстах, причем, что представляется нам особенно показательным, в текстах на позднем санскрите. Наиболее примечательны в этом смысле примеры отражений заведомо поздних (во всяком случае, никоим образом не претендующих на общеарийскую или общеиндоиранскую древность) санскритских композитов[6]: kamathu при др.-инд. kramārtham; mahakama- при др.-инд. mahākrama-, cucakkēśara- при др.-инд. catuścakrēśvarī, ṛppada- при др.-инд. hṛtpada-, cijjalanē при др.-инд. cijjvalanena (cit-jvalanena); jagaghasmaru при др.-инд. jagadghas-

---

[3] Следует отметить, что выводя язык Mahānayaprakāśa из апабхранша и, одновременно с этим, считая его языком-предком кашмири, Дж. Грирсон входит в непреодолимое противоречие со своей собственной идеей генетической самостоятельности дардской группы внутри арийской языковой общности.

[4] Эти языки условно называют западным, восточным и южным апабхранша [Tagare 1948, 15–16]. Имеются также разновидности литературного апабхранша (например, врачада-апабхранша, упанагара-апабхранша), упомянутые в средневековых индийских грамматиках, но не представленные дошедшими до нас текстами.

[5] В этой связи уместно также привести слова крупнейшего российского специалиста по среднеиндийским языкам В. В. Вертоградовой: «От среднеиндийского периода до нас не дошло никаких текстов, воспроизводящих разговорные языки того времени» [Вертоградова 2002, 6].

[6] В качестве отражений мы рассматриваем здесь исключительно словоформы, обнаруживающие следы фонетических изменений. Наряду с ними в тексте Mahānayaprakāśa чрезвычайно много примеров лексических заимствований из санскрита (так называемых слов tatsama), сохраняющихся в неизменном виде.

maraṃ; tulaggu при др.-инд. tallagnaḥ; divyōgu при др.-инд. divyaughaḥ; padāthu при др.-инд. padārthaḥ; savvaga при др.-инд. sarvagā; haṃkārē при др.-инд. ahaṃkārena.

Примеры, подобные приведенным здесь, довольно многочисленны в тексте Mahānayaprakāśa. Значительная их часть не имеет не только соответствий за пределами индоарийского, но и продолжений в позднем индоарийском[7]. Это позволяет считать подобные композиты новообразованиями, характерными для позднего санскрита и чуждыми живому разговорному языку. Таким образом, гипотеза о языке Mahānayaprakāśa как об искусственном языке получает дополнительное подтверждение, хотя полное прояснение картины возможно лишь при наличии системы установленных регулярных звукосоответствий.

Здесь, однако, исследователя ожидает разочарование: все попытки установить такую систему кончаются неудачей. Примеры ничем не мотивированной неоднозначности в фонетических соответствиях встречаются в столь огромном количестве, что имеет смысл говорить не об исключениях, а скорее об отсутствии правил. Наблюдаются, например, такие интересные типы нерегулярности, как наличие разных соответствий одной и той же др.-инд. фонеме в одной и той же позиции внутри одного слова (ср. padipātō при др.-инд. pratipātam) и разных соответствий одному и тому же др.-инд. слову (ср. uditu и udiyo при др.-инд. uditaḥ; gaü и gatō при др.-инд. gataḥ; tōdaśa- и trōvaha '13' при др.-инд. trayodaśa; tiyu и triya при др.-инд. trikaṃ; śatta и śatti при др.-инд. śaktiḥ). Случаи наличия нескольких соответствий одной и той же др.-инд. фонеме в одинаковой позиции в разных словах весьма многочисленны. Среди них можно выделить следующие:

1) соответствия t в интервокальном положении: t (ugghātu при др.-инд. udghātaḥ; vyugata- при др.-инд. vyudgata-; pativiccī при др.-инд. prativr̥ttyā; uditu при др.-инд. uditaḥ); d (padi при др.-инд. prati); 0/y (gaü при др.-инд. gataḥ; thiya- при др.-инд. sthita-; biyu при др.-инд. dvitīyaḥ; udiyo при др.-инд. uditaḥ);

2) соответствия d в интервокальном положении: d (padāthu при др.-инд. padārthaḥ; udiyō при др.-инд. uditaḥ; tōdaśa- при др.-инд. trayodaśa; sattadaśa при др.-инд. saptadaśa; pasādē при др.-инд. prasādena; sēdu при др.-инд. svedaḥ); 0/y/v (bāha при др.-инд. dvadaśa; uya- при др.-инд. udaya-; vāyēna при др.-инд. vādayanti, vādena; cupāvō при др.-инд. catuṣpāde);

3) соответствия k в интервокальном положении: k (pakāru при др.-инд. prakāram; pakāśu при др.-инд. prakāśaḥ; vikasiya при др.-инд. vikasitā); y (tiyu, triya при др.-инд. trikaṃ);

4) соответствия ś в интервокальном положении: ś (pakāśu при др.-инд. prakāśaḥ; tōdaśa- при др.-инд. trayodaśa; sattadaśa при др.-инд. saptadaśa); h (bāha при др.-инд. dvadaśa; aṣṭadaha при др.-инд. aṣṭadaśa; cuvīha при др.-инд. caturviṃśati); s (isaü при др.-инд. īdr̥śaḥ).

Подобный разнобой в рефлексах способен поставить в тупик исследователя-компаративиста. Следует, однако, иметь в виду, что основные постулаты сравнительно-исторического языкознания и, в частности, такой важный из них, как принцип регулярности фонетических соответствий, показали свою справедливость для естественных языков. От языков же искусственных, сознательно созданных людьми, по-видимому, исключительно для письменного употребления, едва ли следует требовать их неукоснительного выполнения. В этой связи уместно опять же провести параллель с литературными

---

[7] Из приведенных выше примеров средне- и новоиндийские продолжения имеются лишь у последнего — ahaṃkāra 'гордыня, самовлюбленность' (ср. пали ahaṃkāra то же, кумауни hangār 'дух').

апабхранша. Здесь также наблюдается нерегулярность в отражении древнеиндийских фонем, причем степень этой нерегулярности (количество разных рефлексов одной прафонемы или прафонемного сочетания в одинаковой позиции в слове) тем больше, чем позднее датировка памятника или, иными словами, чем сильнее разрыв между языком памятника и народно-разговорным языком[8]. Данную ситуацию обычно объясняют тем, что авторы текстов не всегда четко выполняли рекомендации грамматистов, иногда следуя при пересчете с санскрита более ранним пракритским моделям или же моделям, характерных для апабхранша других регионов. Влияние пракритов на апабхранша было весьма значительным на всех языковых уровнях (пракритизмы отмечаются даже среди словоизменительных морфем), что дало некоторым исследователям, в частности Ж. Блоку, основание утверждать, что апабхранша не является самостоятельным языком [Елизаренкова 2004, 123; Bloch 1934, 12].

Ситуация в языке Mahānayaprakāśa представляется нам во многом аналогичной таковой в литературных апабхармша. По всей видимости, автор «производил» этот язык из санскрита по нескольким разным моделям, возможно отражавшим фонетическое развитие различных эпох. Из приведенных выше примеров можно видеть, что разные рефлексы одной и той же древнеиндийской фонемы нередко могут представлять разные ступени одного историко-фонетического процесса (например, сохранение интервокального t и его озвончение, сохранение интервокального d и его выпадение). Отличие от основных среднеиндийских литературных языков, как кажется, состоит в гораздо меньшей степени нормированности: как уже говорилось, для языка Mahānayaprakāśa трудно предполагать исключения из правил ввиду невозможности отследить сами правила. Таким образом, гипотеза об искусственном характере рассматриваемого языка представляется нам наиболее вероятной. Скорее всего, мы имеем здесь дело с не вполне удачной попыткой создать еще один литературный апабхранша. Данный язык, будучи произведенным из санскрита искусственным конструктом, как уже говорилось, по определению не может являться предком для какого-либо живого языка, в том числе и для кашмири, принадлежащего к тому же к иной (дардской) ветви индоиранской языковой общности.

Тот факт, что язык Mahānayaprakāśa, скорее всего, является искусственным, сам по себе отнюдь не лишает его интереса для исследователя дардской и, в частности, кашмирской диахронии. Как известно, в текстах на литературных среднеиндийских языках встречаются отдельные элементы, заимствованные из местных разговорных языков[9]. Нельзя исключить, что в нашем случае такими элементами могут являться заимствования из языка — предка современного кашмири. В случае если такие элементы действительно имеются в наличии, их можно было бы считать наиболее ранними кашмирскими словами и/или морфемами, зафиксированными письменно. Выделить их можно было бы по ряду критериев, главным из которых является историко-фонетический: следует попытаться найти в языке Mahānayaprakāśa примеры специфически кашмирского (и, шире, дардского) фонетического развития.

Подобные примеры действительно обнаруживаются. Так, в числительном akka '1' прослеживается специфичная для ряда дардских языков, включая кашмири, рефлексация общеарийского краткого дифтонга *ai в преконсонантной позиции (*a* при древне-

---

[8] См., например, детально рассмотренное Г. В. Тагаре развитие в различных апабхранша интервокальных смычных и консонантных групп [Tagare 1948, 78–82, 87–99]. В таблицах, приводимых в этой работе, достаточно надежно прослеживается корреляция между датировкой текстов и степенью регулярности звукосоответствий.

[9] Авторы пракритских грамматик обычно классифицируют их как элементы особого уровня (krama) языка — деши. Об этом см. [Вертоградова 2002, 9–10].

индийском отражении в виде *e*)[10]. В консонантизме отмечаются отдельные случаи десас-
пирации древних звонких придыхательных[11] (ср. gōnāna при др.-инд. ghonānām; dari
при др.-инд. dharati; aggu при др.-инд. argham; divyōgu при др.-инд. divyaughaḥ; majjā
при др.-инд. madhyāt; ūda- при др.-инд. ūrdhva-). Однако, как справедливо отмечает
Дж. Грирсон, потеря придыхания в подобных случаях может объясняться ошибками
переписчиков — носителей кашмири, языка, лишенного звонкой придыхательной се-
рии[12]. В любом случае, приведенным здесь примерам противостоит значительно (в не-
сколько раз) большее число примеров индоарийского историко-фонетического разви-
тия: отражения общеарийского краткого дифтонга **ai* в виде *ē* (ср. ēśu при др.-инд. eṣaḥ;
ēhu при др.-инд. etad; jēṭhī при др.-инд. jyeṣṭhā; pēkṣēta при др.-инд. prekṣya; pavēśē при
др.-инд. praveśena; sēdu при др.-инд. svedaḥ; pattēkasa при др.-инд. pratyekasya) и сохра-
нения древних звонких придыхательных в неизменном виде (ср. bhāji при др.-инд. bhrā-
jate; bhāya при др.-инд. bhāti; bhuttē при др.-инд. bhukte; bhūma при др.-инд. bhūmiḥ;
ubbhāvō при др.-инд. udbhāvitaḥ; khambhēta при др.-инд. skambhitvā; pabhāvē при др.-
инд. prabhāveṇa; patibhōgē при др.-инд. pratibhogena; sabbhāvē при др.-инд. sadbhāvena;
dhāmi при др.-инд. dhāmni; apabōdhu при др.-инд. aprabodhaḥ; udhiadha при др.-инд.
ūrdhvādhaḥ; kōdha при др.-инд. krodhaḥ; dōdhā при др.-инд. dvidhā; niddhāmi при др.-
инд. nidhāmni; ugghātu при др.-инд. udghātaḥ; -ghanō при др.-инд. ghanaḥ; dīgha- при
др.-инд. dīrgha-; nigghātu при др.-инд. nirghātaḥ; nighāriśā при др.-инд. nigharṣāt).

Среди историко-фонетических расхождений между языком Mahānayaprakāśa и
кашмири следует отметить также примеры различного развития сибилянтов и некото-
рых консонантных групп, содержащих *r*. Такие примеры хотя и не свидетельствуют од-
нозначно против принадлежности языка Mahānayaprakāśa к дардской группе, но все же
бесспорно указывают на то, что этот язык не может являться предком современного
кашмири.

Рефлексы древних сибилянтов **ś* и **ṣ* в Mahānayaprakāśa совпадают в виде *ś* при со-
хранении оппозиции в кашмири (**ś* > h, **ṣ* > š[13]). Ср.:

śatta при др.-инд. śaktiḥ, кашм. hĕkun 'мочь' (от того же корня); śamēta при др.-инд.
śamitva, кашм. hamun 'успокаиваться, остывать, быть потушенным' (от того же корня);
śiru при др.-инд. śiraḥ, кашм. hīr 'голова', но

śa '6' при др.-инд. ṣaṭ, кашм. še то же; ēśu при др.-инд. eṣaḥ).

В языке Mahānayaprakāśa обычны переходы **pr* > p и **bhr* > bh, в то время как в каш-
мири конечный *r* сохраняется (ср., например, развитие указанных кластеров в ряде эти-
монов, общих для обоих языков: pāvēya при др.-инд. prāpayati, pāvēna при др.-инд. prā-
payanti, но кашм. prāwun 'получать'; padi при др.-инд. prati, но кашм. prath 'каждый';
bhāji при др.-инд. bhrājate, но кашм. brazun 'сверкать, блестеть, быть ярким').

Характерное для кашмири развитие церебральных из консонантных групп типа
«r + зубной» не обнаруживается в Mahānayaprakāśa, где нормальными являются перехо-
ды **rt, *rth, *rdh* > tt, th, dh:

---

[10] Об этом см. [Коган 2005, 19–22].

[11] Данное явление, вероятнее всего, представляет собой общедардскую классифицирующую историко-
фонетическую инновацию.

[12] Примечательно, что среди «кашмирских пандитов» (представителей небольшой общины кашмир-
цев, исповедующих индуизм) распространена особая традиция произнесения санскритских текстов, пред-
полагающая дезаспирацию изначальных звонких придыхательных согласных [Grierson 1929, 83].

[13] О развитии сибилянтов в кашмири см. также [Коган 2009].

kattiku при др.-инд. kartr̥kaḥ, но кашм. waṭun 'сворачивать, складывать' < *wartaya-, ср. др.-инд. vartayati 'поворачивает', кашм. kaṭun 'резать' < *karta-, ср. др.-инд. kartati (наряду с kr̥ntati) 'режет';

kamathu при др.-инд. kramārtham, padāthu при др.-инд. padārthaḥ, но кашм. coṭh 'четырехдневная малярия', ср. др.-инд. caturtha- 'четвертый';

udhiadha при др.-инд. ūrdhvādhaḥ, ūda- при др.-инд. ūrdhva-, но кашм. wŏḍ 'макушка' (из общеарийск. *r̥̄dhwa- > др.-инд. ūrdhva-).

Возвращаясь к вопросу об общих изоглоссах, объединяющих кашмири и язык Mahānayaprakāśa, необходимо коснуться морфологии. В Mahānayaprakāśa засвидетельствованы два личных глагольных показателя, характерных для кашмири и не имеющих несомненных параллелей в индоарийских языках: -ā 1 Sg Prs при кашм. -i̇ (< *ā[14]) то же; -īva 2 Sg Imp при кашм. -iw то же. Отмечено также соответствие кашмирскому субъектному суффигированному местоимению 3 л. ед.ч. -n в виде -na. В индоарийских языках данный элемент не встречается[15].

На лексическом уровне язык Mahānayaprakāśa также имеет ряд специфических параллелей с кашмири. Речь идет о лексемах, представленных в кашмири (а иногда также и в других дардских языках), но отсутствующих (вовсе или же в данном фонетическом облике) в индоарийском: cāvu 'вошедший, вошел' при кашм. cāw то же[16]; patta 'после' при кашм. pati̇ 'после, позади', шина pati, майян patō, торв. pat, г.-б. pata 'позади', башк., паш. pat 'после', катарк. padē 'назад, снова'; ju '2' при кашм. zi̇ то же; ci 'тебе' при кашм. sĕ то же. Как можно видеть, приведенные здесь слова из Mahānayaprakāśa фонетически довольно слабо отличаются от их кашмирских соответствий. Явными инновациями современного кашмири являются дентализация палатальных аффрикат[17] и падение конечного краткого гласного *u*. Однако, как было установлено нами в специальном исследовании, оба эти явления носят относительно поздний характер, о чем свидетельствует, например, тот факт, что они обнаруживаются в заимствованиях из персидского и арабского языков [Коган 2009]. Кроме того, нужно принять во внимание и особенности использованного для фиксации ранних редакций текста Mahānayaprakāśa письма шарада, не располагающего специальными символами для дентальной аффрикаты *c* и для *z*.

На основании вышесказанного следует признать необоснованным (или, во всяком случае, крайне сомнительным) утверждение С. К. Чаттерджи о какой-то чрезвычайной архаичности языка Mahānayaprakāśa и о необходимости на этом основании ранней датировки памятника [Chatterji 1963, 258][18]. Встречающиеся в тексте кашмирские элементы отражают языковое состояние, которое вполне могло быть хронологически близким к современному.

---

[14] О происхождении данного окончания в кашмири и его связях за пределами дардской группы см. [Коган 2005, 151–153].

[15] Кашмирское суффигированное местоимение -n продолжает общеарийскую местоименную основу *ana-. Рефлексы этой основы отмечены в ведийском, а также (в форме инструментального падежа) в пали и пракритах [Turner 1966, 14], однако там они не являются местоименными энклитиками или суффиксами.

[16] Возможно < *atyāgata- [Turner 1966, 12].

[17] При транскрибировании примеров мы, в соответствии с установившейся традицией, обозначаем латинской буквой *c* глухую дентальную аффрикату в кашмири и глухую палатальную аффрикату в языке Mahānayaprakāśa.

[18] Впрочем, основания для предлагаемой С. К. Чаттерджи датировки (XIII в.) в любом случае остаются неясными.

Говоря о дардских элементах в языке Mahānayaprakāśa, нельзя не отметить их относительную малочисленность. Приведенный в настоящей работе перечень таких примеров является практически полным. При этом принадлежность некоторых упоминавшихся выше слов к дардскому пласту представляется сомнительной. В первую очередь это относится к примерам дезаспирации звонких придыхательных, появление которых, как уже отмечалось, может быть связано с ошибками писцов[19]. Таким образом, наиболее точно язык Mahānayaprakāśa может быть, на наш взгляд, охарактеризован как среднеиндийский литературный язык, содержащий определенное (притом довольно малое в процентном отношении) количество дардских заимствований[20].

Хотя данное заключение является наиболее вероятным при современном уровне знаний, оно все же носит максимально общий характер и не дает ответов на целый ряд вопросов, связанных с языком Mahānayaprakāśa. Так, несмотря на установленный нами факт наличия в этом языке кашмирских элементов, в вопрос о субстрате, т.е. о родном языке автора текста, оказавшем влияние на язык несанскритских фрагментов, все еще не внесена окончательная ясность. Дело в том, что наряду с дардским этимологическим пластом в Mahānayaprakāśa налицо и другие пласты, которые также могут являться субстратными. Отмечается, например, наличие ряда индоарийских элементов, имеющих соответствия в поздних индийских языках, но не в древнеиндийском: haḍyu 'кости'[21], ср. лахнда haḍ, зап. пах. (бхадравахи) haḍḍ 'кость'[22]; utthē 'именно там'[23], ср. пандж. utthe, лахнда utth то же.

Определенный интерес представляют также некоторые фонетические архаизмы языка Mahānayaprakāśa, в частности, повсеместное сохранение древнеиндийского сонанта y[24] в начале слова и группы kṣ[25] во всех положениях, в то время как для всех известных нам литературных апабхранша характерны переходы y > j, kṣ > (k)kh, (c)ch. Уникальными для индоарийских языков являются некоторые типологические характеристики рассматриваемого языка. К таковым относится, например, почти полное отсутствие послелогов и выражение синтаксических отношений исключительно при помощи падежных флексий[26].

Подытоживая все вышесказанное можно с уверенностью констатировать, что язык несанскритских фрагментов Mahānayaprakāśa еще может стать объектом для интереснейших исследований в будущем, хотя вопрос о его отношениях с языком кашмири представляется нам вполне ясным: язык Mahānayaprakāśa никоим образом не может считаться дардским языком и не является предком ни для современного кашмири, ни (ввиду своего искусственного характера) для какого-либо другого живого языка.

---

[19] Один из примеров перехода gh > g (divyōgu при др.-инд. divyaughah) представляет собой явное отражение позднего санскритского композита (см. выше) и поэтому не может считаться дардским по происхождению.

[20] К числу последних можно отнести и перечисленные выше глагольные показатели. Как уже говорилось, заимствование словоизменительных морфем известно в литературных среднеиндийских языках.

[21] Соответствует asthīni в санскритском комментарии.

[22] У приведенных примеров, возможно, имеется соответствие и в кашмири — aḍi 'кости предплечья или голени'. Это соответствие, однако, семантически отстоит дальше указанных нами индоарийских примеров, а, кроме того, само может являться индоарийским заимствованием.

[23] Соответствует tatraiva в санскритском комментарии.

[24] Список примеров с указанием древнеиндийских параллелей см. в [Grierson 1929, 126].

[25] Примеры см. в [Grierson 1929, 21].

[26] Данная черта сближает язык Mahānayaprakāśa с литературными пракритами и отличает от литературных апабхранша, а также от новоиндийских и других арийских языков региона: дардских, нуристанских, многих иранских.

Литература

Вертоградова 2002 — В. В. ВЕРТОГРАДОВА. *Пракриты*. Издание второе. М., Восточная литература, 2002.

Елизаренкова 2004 — Т. Я. ЕЛИЗАРЕНКОВА. *Языки мира. Индоарийские языки древнего и среднего периода.* М., Academia, 2004.

Захарьин 1978 — Б. А. ЗАХАРЬИН. Генезис видо-временных отношений в языке кашмири // *Индийское языкознание*. М., Наука, 1978.

Коган 2005 — А. И. КОГАН. *Дардские языки. Генетическая характеристика*. М., Восточная литература, 2005.

Коган 2009 — А. И. КОГАН. К вопросу о ряде фонетических изменений в языке кашмири и их относительной датировке // Аспекты компаративистики 4 [Aspects of Comparative Linguistics 4] / Под ред. Г. С. Старостина. М.: Изд-во РГГУ, 2009. (Orientalia et Classica: Труды Института восточных культур и античности. Выпуск XXVIII.) С. 25–54.

Bloch 1934 — J. BLOCH. *L'indo-aryen. Du Veda aux temps modernes*. Paris, 1934.

Bubenik 1996 — V. BUBENIK. *The Structure and Development of Middle Indo-Aryan Dialects*. Delhi, Motilal Banarasidass, 1996.

Chatterji 1963 — S. K. CHATTERJI. *Languages and Literatures of Modern India*. Calcutta, 1963.

Grierson 1929 — G. A. GRIERSON. The Language of the Mahānayaprakāśa, an Examination of the Kashmiri as Written in the 15th Century // *Memoirs of the Asiatic Society of Bengal*, vol. XI, №2. Calcutta, 1929.

Tagare 1948 — G. V. TAGARE. *Historical Grammar of Apabhraṃśa*. Poona, 1948.

Turner 1966 — TURNER R. L. *A Comparative Dictionary of the Indo-Aryan Languages*. L.: Oxford University Press, 1966.

Kashmiri is the only Dardic language that has its own script and literary tradition. However, since the latter is not very old, the regular philological method can hardly be helpful in its diachronic study. Researchers have made a lot of attempts to find authentic old Kashmiri texts. Although the existence of such texts remains unproved, some interesting problems have arisen in the course of research; one of them is the issue of the attribution of the language of non-Sanskrit fragments of Mahānayaprakāśa — a philosophical treatise belonging to the tradition of Kashmiri Shaivism. This problem is dealt with in the present article. The author comes to the conclusion that the language in question is undoubtedly Indo-Aryan and can thus by no means be considered an early form of Kashmiri. Moreover, being actually a variety of the late literary Middle Indo-Aryan (Apabhramsha), this language is, in fact, artificial, and therefore, could not have been an ancestor for any modern language.

Alexey Lyavdansky
Moscow, Russian State University for the Humanities

Temporal deictic adverbs as discourse markers in Hebrew, Aramaic and Akkadian[1]

The categorial shift from temporal deictic adverb to discourse marker is observed in many languages of the world. There are three Semitic languages — Hebrew, Aramaic, and Akkadian — where similar developments were attested for a temporal adverb with present time reference. This article is dedicated to the comparison of non-adverbial usages for Hebrew (*wə)ʕattā*, Aramaic *kʕt/kʕnt/kʕn* and Akkadian *inanna* and *anumma.* The preliminary results of this investigation, based on the findings of Rhetorical Structure Theory and discourse markers research, show that in most of the uses these adverbs function as discourse markers. As is the case with Hebrew (*wə)ʕattā*, the specific discourse function is attested also for Aramaic *kʕt/kʕnt/kʕn* and Akkadian *inanna*: an adverb with the meaning 'now' marks a transition from assertive discourse unit to directive discourse unit within directive utterances. The range of usage for Aramaic *kʕt/kʕnt/kʕn* and Akkadian *inanna* is broader than for Hebrew (*wə)ʕattā*. Akkadian *anumma* is another type of lexeme: not being used as a temporal deictic adverb, it also appears in directive utterances, but, unlike Hebrew (*wə)ʕattā* and Aramaic *kʕt/kʕnt/kʕn*, it usually has an assertive discourse unit in its right co-text, the transition from assertive to directive usually left unmarked.

Keywords: Semitic languages, historical syntax, discourse relations, discourse markers.

It is probably a universal phenomenon that a temporal deictic adverb with the present time reference like English *now* appears in specific contexts, where it assumes non-temporal meaning. Yet it is only recently that such non-adverbial uses of temporal adverbs have drawn special attention. The interest to these uses is related to the growth of discourse studies over the last three decades. It has been noticed that *now* and its sister-words in other languages (further designated as "now-words") quite often appear in contexts that are very typical for connectors/connecting particles, or discourse markers. Accordingly, there are two main types of meaning which are usually distinguished for these words: the temporal (adverbial) meaning and the discursive (textual) meaning. This distinction may be demonstrated by the following examples:

(1a) *sicut fortis equos, spatio qui saepe supremo vicit Olympia,* **nunc** *senio confectus quiescit* (A fragment from Ennius, quoted by Cicero, Cato Maior 5)

(1b) *redeo* **nunc** *ad epistulam tuam* (Cicero. Ad Atticum 14.13.5)

(2a) *Either do it* **now** *or not at all* (RHWUD).

(2b) RICHARD PLANTAGENETH: *Lord Buckingham, methinks, you watch'd her well:*
A pretty plot, well chosen to build upon!
Now*, pray, my lord, let's see the devil's writ.*
What have we here? (Shakespeare. Henry VI, 693–696)

Examples (1a, 2a) demonstrate "now-words" (Latin *nunc*, English *now*) in a temporal adverbial meaning; in the examples (1b, 2b) "now-words" appear as discourse particles/markers[2]. Consequently, these two types of meaning can constitute the basis for postulating two distinct lexemes, for which I propose the symbols *now* I (adverbial) and *now* II (non-adverbial, discursive).

The non-adverbial usage of temporal deictic adberbs being almost universal, it is, however, difficult to predict the range of the discursive meaning that these now-words assume in a given language. The present article is dedicated to the comparative study of now-words in three Semitic languages: Hebrew, Aramaic and Akkadian. The analysis is concentrated on Biblical Hebrew (including Epigraphic Hebrew of the same epoch) and Egyptian Aramaic. In addition, some parallel examples from Old Babylonian and Standard Babylonian dialects of Akkadian are discussed.

The aim of this paper is to show the preliminary results achieved in the analysis of the development of meaning for now-words in the said three Semitic languages. Within the cadre of this analysis I am trying to understand to what extent it is possible to distinguish between adverbial and non-adverbial uses for dead languages like the chosen ones. Another problem which can be treated only in part at the present state of the research: What more can be said about the comparison between Hebrew (*wə*)*ʕattā*, Aramaic *kʕt/kʕnt/kʕn* and Akkadian *inanna* and *anumma*, beyond the fact that they are sometimes similar in usage[3]?

Although this study is comparative in the sense of "Languages in contrast and comparison", it does not pursue any etymological goals. Nevertheless, its results, if proven viable, may shed a new light on the problems of the origin of Semitic temporal deictic adverbs and the words related to them in meaning.

1. Method and theoretical background

The main theoretical frameworks upon which my investigations are based are Discourse Markers (DM) research[4] and Rhetorical Structure Theory (RST) [Mann & Thompson 1988]. I will not go into a detailed presentation of these theories, fairly well-known from many publications. Suffice it just to point out the most important issues for the present discussion. The authors of RST have suggested a catalog of rhetorical (i. e. discourse) relations and provided an analytical instrument that is useful to determine the functions of discourse markers. At a certain point it was understood that some of the discourse particles and functionally similar expressions signal or mark discourse relations[5]. In fact, connectivity took very firm ground and usually the main position in the definitions of that specific class of words and expressions called "discourse markers" [Fraser 2006; Schourup 1999].

Since one of the main tasks of this paper is to distinguish between adverbial and non-adverbial (discourse) uses of now-words in Hebrew, Aramaic and Akkadian, I need to pay

---

[2] On the criteria of this distinction see par. 1 of the present article.

[3] Hebrew *wʕth* and Aramaic *kʕt* were compared by many authors, e. g. [Lande 1949; Schwiderski 1997]. As to the comparison between *wʕth* and Akkadian *inanna*, there is a short remark by Moran [2003: 16].

[4] There is no universal theory of Discourse Markers, but there is a certain line of thought represented in the works of Deborah Schoffrin [1997], Bruce Fraser [1996, 1999], Lawrence Schourup [1999] and ADP.

[5] As it was put, e. g., by Fraser in one of his earlier works: "…discourse marker, an expression which signals the relationship of the basic message to the foregoing discourse" [Fraser 1996: 186]. "The characteristic figuring most prominently in definitions of DMs is their use to relate utterances or other discourse units" [Schourup 1999: 230].

some attention to the problem of what is to be understood under the term "discourse marker". There are many approaches to DM now, but I will only mention a summary list of DM characteristics suggested by Lawrence Schourup [1999], who tried to survey the most important findings in the field of DM research. According to him, the main characteristics of DM include: 1) connectivity, 2) optionality[6], 3) non-truth conditionality[7], 4) weak clause association[8], 5) initiality[9], 6) orality, 7) multi-categoriality[10] [Schourup 1999]. The following observations on *now* according to the above list of DM features may show the applicability of this approach; it was shown that *now* when used as a discourse particle:

1) marks a transition "from a disputable issue to evaluation of it" [Aijmer 1988: 19]; cf. "l'enonciateur ce sert de *now* pour marquer une transition avec son discours antérieur" [Brunaud 1991: 85];

2) is much less frequent in non-surreptitious conversation as compared to surreptitious (recorded using hided devices) conversation [Aijmer 2002: 68],

3) loses its ordinary temporal meaning [Aijmer 1988: 15];

4) is phonologically independent, "constituted most often a (prosodic) phrase on its own" [Aijmer 2002: 59];

5) "has a fixed position in the utterance as the leftmost element in the larger utterance" [Aijmer 1988: 18];

6) is "characteristic of speech rather than of writing" [Aijmer 2002: 70];

7) have as its synonyms expressions from other syntactic classes.

One of the main achievements of DM research is the understanding that the discourse model should account for several layers of discourse coherence, or "planes of talk" [Schiffrin 1987]. It is significant for the discussion of 'now-words' that discourse markers are sensitive not only to rhetorical relations in the sense of RST, they sometimes point to the new speech act in the flow of discourse. In other words, DM may contribute to what has been called "action structure[11]" [Schiffrin 1987].

Both RST and DM analysis draw attention to the immediate context of DM, not only to the textual unit which immediately follows a DM[12], but also to the textual unit which immediately precedes a DM. These units, relevant as the intra-textual context of a DM, may be called "text spans" [Mann & Thompson 1988: 245], "discourse segments" [Fraser 2006: 191] or "discourse units" [Schourup 1999; Redeker 2006]. Consequently, this should be the rule for the presentation of the linguistic material in a discussion of concrete DMs — the uses of a DM should be presented with both adjacent discourse units[13], which is rarely done in works on Semitic discourse particles of a connective nature.

---

[6] In the sense that "if a DM is omitted, the relationship it signals is still available to the hearer, though no longer explicitly cued" [Schourup 1999: 231]. The statistical data on the ratio of cued/non-cued discourse relations were presented in [Taboada 2006].

[7] DMs usually "contribute nothing to the truth-conditions of the propositions expressed by an utterance" [Schourup 1999].

[8] It is often indicated by the phonological independence of a DM.

[9] DMs tend to appear at the beginning of a sentence or a discourse unit. Though there are DMs sometimes or even exclusively placed within clauses, e. g. English *after all*, *now*, Biblical Hebrew *ʔēpō*.

[10] The class of DM may include adverbs, conjunctions, interjections, verbs, clauses [Schourup 1999].

[11] Analogous to G. Redeker's "rhetorical structure", as suggested in [Müller 2002: 30].

[12] Sometimes considered as "host utterance" for a DM, which is problematic because DMs are very often seen as syntactically unintegrated entities [ADP: 8].

[13] The problem of limits for these units is not discussed here; see on this [ADP] and [Mann & Thompson 1988].

Finally, one more application of RST to DM research deserves our attention. As has already been shown in a number of RST studies, the text in the final analysis may be looked upon as a complex of discourse units, organized hierarchically and sequentially and related to each other by one (or maybe more than one) of the rhetorical relations. Potentially, most of these relations may be lexically marked in a given language, in a given text type. Therefore, one of the ways to determine the functional distribution of DMs in a language is to parse texts rhetorically, in order to show which relations are marked and to what degree, as has been done by M. Taboada for two English corpora [Taboada 2006]. This is one of the paths —already somewhat trodden —which must lead to tangible results, important for the typology of discourse marking.

As to the languages investigated here — Hebrew, Aramaic and Akkadian — there is obviously quite a lot of work to be done as long as rhetorical relations are concerned. In Biblical Hebrew, where discourse particles have been explored but partially, there is a very promising field of study constituted by the constructed literary dialogue with its plethora of discourse particles (*wə-, kī, hēn/hinnē, (wə)ʕattā, lākēn, ʕal kēn, ʔēpō, ʔăbāl*). Imperial Aramaic provides a relatively small, but sufficient (for that sort of study) corpus of letters from Egypt and probably other regions; the field of Imperial Aramaic discourse particles is almost untouched. Finally, besides many interesting corpora in Akkadian, there is the corpus of Amarna correspondence, where numerous discourse particles present serious problems, often avoided by translating these particles automatically by "now" or "moreover".

2.1. Classical Hebrew (wə)ʕattā.

The Hebrew expression *ʕattā* is the main word for "now" in Biblical Hebrew. It occurs 433 times [Jenni 1972: 6], which is quite a figure for such restricted corpus as Hebrew Bible[14]. In approximately 60% of its usages it occurs with the preceding conjunction *wə-*, namely *wəʕattā* (272 times [Jenni 1972: 6])[15]. There are also 20 occurrences of *wʕt* in epigraphic material, all of them in letters. These statistical data already suggest the general tendency in the usage of *ʕattā*: if *wə-* is attached to *ʕattā*, it obviously represents the clause-combining *wə-*, which means that in almost all of these 272 cases *ʕattā* is placed at the beginning of a clause. According to the 'initiality' characteristic of a DM, the clause-initial position of *ʕattā* points to its discursive character at least in these cases.

For quite a long time, the string *wəʕattā* had been considered a compound with its own specific range of meaning [Laurentin 1964; Brongers 1965]. It was Ernst Jenni who rightly noted that there is no principal difference between the expressions *wəʕattā* and *ʕattā*. In fact, *wəʕattā* is frequent and most conspicuous in its non-adverbial discursive usage, but both *wəʕattā* and *ʕattā* may have temporal adverbial and discursive functions [Jenni 1972]. Moreover, if we look at the expression *wə-ʕattā* from the vantage point of DM analysis, it should be treated as a collocation, a juxtaposition of two discourse markers with similar meanings. Inclusion of the coordinating conjunction *wə-* into the class of Hebrew DMs is corroborated by my observation that its usage in dialogue differs very much from its usage in narrative. In narrative, *wə-* is a default coordinating and subordinating conjunction: it opens every clause if it is not (rarely) substituted by other coordinators (*ʔap, raq*) or subordinators (*ʔăšär, ləmaʕan*), or omitted before clauses with specialized function in narrative (author's remarks). In dialogue

---

[14] Approximately 300000 words.

[15] On the problems of these statistical data, mostly related to text corruption, see [Jenni 1972].

every appearance of *wə-* should be accounted for in terms of its specific discourse functions, one of which has been analyzed in [Miller 1999].

2.2. (wə)ʕattā as a temporal adverb.

First of all, it is important to demonstrate the usage of *(wə)ʕattā* in different types of temporal adverbial meaning. Typologically, as shown in [Pérennec 2002], the range of temporal reference attested for an adverb may suggest the type(s) of its discourse function.

We already said that *ʕattā* is the main word for 'now' in Biblical Hebrew; consequently, it appears in all the most typical contexts for a temporal adverb with present time reference. The adverbial meaning of *ʕattā* is highlighted in contexts where the situation of speaking is contrasted with the (more often) past or future situation:

(3) *zākarnū ʔät haddāgā ʔăšär nōkal bəmiṣrayim hinnām ʔēt haqqiššuʔīm wʔēt häʔăbaṭṭīḥīm wəʔät häḥāṣīr wəʔät habbəṣālīm wəʔät haššūmīm* **wəʕattā** *napšēnū yəbēšā ʔēn kōl biltī ʔäl hammān ʕēnēnū*

We remember the fish we used to eat in Egypt for nothing, the cucumbers, the melons, the leeks, the onions, and the garlic; **but now** our strength is dried up, and there is nothing at all but this manna to look at.[16] (Num 11:5f)

The beginning of the new state of affairs, starting from the present moment, is usually expressed by a prepositional phrase, constituted by the preposition *min* (*mē-*) and adverb *ʕattā* (4); the same phrase is used also in the context of the comparison of the present situation with the past (5).

(4) *ləmarbē hammiŝrā ūləšālōm ʔēn qēṣ ʕal kissē dāwīd wəʕal mamlaktō ləhākīn ʔōtāh ūləsaʕădāh bəmišpāṭ ūbiṣdāqā* **mēʕattā** *wəʕad ʕōlām*

His authority shall grow continually, and there shall be endless peace for the throne of David and his kingdom. He will establish and uphold it with justice and with righteousness **from this time** onward and forevermore (Isa 9:7).

(5) *wəʔāmərā ʔēləkā wəʔāšūbā ʔäl ʔīšī hārīšōn kī ṭōb lī ʔāz* **mēʕattā**

… she shall say, "I will go and return to my first husband, for it was better with me then **than now**." (Hos 2:9)

ʕattā may also denote recent past[17]:

(6) *ʔānōkī ʕāŝītī ʔät hāʔāräṣ ʔät hāʔādām wəʔät habbəhēmā ʔăšär ʕal pənē hāʔāräṣ bəkōḥī haggādōl ūbizrōʕī hannəṭūyā ūnətattīhā laʔăšär yāšar bəʕēnāy* **wəʕattā** *ʔānōki nātattī ʔät kōl hāʔărāṣōt hāʔēllä bəyad nəbukadnäṣar mäläk bābäl abdī wəgam ʔät ḥayyat haŝŝādä nātattī lō ləʕōbdō*

It is I who by my great power and my outstretched arm have made the earth, with the people and animals that are on the earth, and I give it to whomever I please. **But now** I have given all these lands into the hand of King Nebuchadnezzar of Babylon, my servant, and I have given him even the wild animals of the field to serve him (Jer 27:5f).

---

[16] The translation of biblical texts is according to New Revised Standard Version if not specified otherwise.

[17] The specific nature of the prophetic texts here allows for a double meaning according to time reference: the giving of lands has happened just before the moment of speaking, but it is understood as a decision, or promise of Yahweh, the event itself will happen in (near?) future.

Summing up, ʕattā as a temporal adverb denotes the moment of speaking, the recent past, the imminent future. The syntactic position of ʕattā in its temporal adverbial uses is not always non-initial: it may occur at the beginning of a sentence, in which cases it is often preceded by the conjunct wə-[18]. The adverbial meaning applies here because it is highlighted by the contrast with mostly past situation. The contrast is expressed lexically: zākarnū 'we remember'(4), (mē)ʔāz 'then' (5; Josh 14:11), mīmē qädäm "from early days" (2Kgs 19:25). The sense of contrast is sometimes suggested by knowledge shared by the communicators: in (6) it is known that the "making of the earth etc." happened in the remote past. The morphological criterion in this last case and in many other cases does not work: Classical Hebrew verb forms are often ambiguous in their time reference, as is the case with the form nōkal 'we used to eat' (3), which, in other contexts, could be translated as 'we eat' or 'we shall eat'. The occurrences with prepositions m- 'from' (13 cases) and ʕd 'until' (Gen 32:4 and 8 more cases) are clearly adverbial; here, this simple syntactic criterion of ʕattā's adverbial usage works successfully.

2.3. (wə)ʕattā as a discourse marker.

The most typical context of wəʕattā (and sometimes ʕattā without wə-) is when it has injunctive verb forms in its immediate right co-text:

(7) *wayyišlaḥ malʔākīm ʔäl ʔăbīmäläk bətormā lēmōr hinnē gaʕal bän ʕäbäd wəʔäḥāw bāʔīm šəkēmā wəhinnām ṣārīm ʔät hāʕīr ʕāläkā* **wəʕattā** *qūm laylā ʔattā wəhāʕām ʔăšär ʔittāk wäʔărōb baśśādä*
He sent messengers to Abimelech at Arumah, saying, [DU1] "Look, Gaal son of Ebed and his kinsfolk have come to Shechem, and they are stirring up the city against you. **Now therefore**, [DU2] go by night, you and the troops that are with you, and lie in wait in the fields (Jud 9:31f).

There are two most important issues to be noted here: first, wəʕattā appears on the border between two discourse units within direct speech (DU1, DU2); DU1 and DU2 are related, the text before wəʕattā being a background or motivation for the text after wəʕattā; second, the texts (discourse units) before wəʕattā and after it are different in their mood — indicative changes to imperative.

The relatedness of both discourse units is not expressed by any cohesive devices such as pronominal anaphora, it is suggested by the narrative context and may be revealed with the help of semantic analysis: the words 'city' and 'field' (i. e. open space around the city) are related in the mental world, shared by speakers of Classical Hebrew. But there is another type of cohesion between DU1 and DU2: the text immediately preceding (wə)ʕattā serves as a justification or motivation for the imperative utterance in DU2. In this and many other cases the text before (wə)ʕattā describes a state of affairs which is supposed to urge the addressee to undertake an action. Or it may be said that the speaker presents his request or order as justified or motivated by the preceding discourse unit[19]. In terms of RST, the rhetorical relation JUSTIFY is obtained between the two discourse units. It is for that reason that in many cases the adverb (wə)ʕattā is translated as 'therefore' or 'now therefore, now then'.

As has already been said above, DU1 and DU2 differ in their mood (indicative vs. imperative). For the analysis of (wə)ʕattā it is more important to note that they differ in their illo-

---

[18] The discursive non-adverbial meaning may be present in these "temporally contrasted" contexts; on this problem see par. 2.3.

[19] This type of a relation between two unites is described also as "resultative" [Müller 2005: 82].

cutionary force: DU1 assertive, DU2 directive. In fact, in the overwhelming majority of cases the text before (*wə*)*ʕattā* is illocutionary assertive, while the text after (*wə*)*ʕattā* is in most cases directive[20]. Therefore, (*wə*)*ʕattā* may be tentatively called a "speech act marker"[21]. One of the earlier analyzes of Hebrew *wəʕattā* comes to a similar conclusion: *wəʕattā* is "Illokutionsindikator" [Wagner 1997: 236].

Let us now look at the applicability of the main DM characteristics for this use of *ʕattā*. The connectivity is suggested by a rhetorical relation between two discourse units discussed above. Non-truth-conditionality is seen in that the removal of (*wə*)*ʕattā* does not affect the meaning of the imperative sentence: the slot of temporal adverbial is filled by *laylā* 'by night'. As to the characteristics of initiality, *ʕattā* is posited at the beginning of the sentence, as indicated by the preposed sentence coordinator *wə*- 'and'. As far as the 'optionality' characteristics is concerned, we need to look at all the potential slots for (*wə*)*ʕattā*. Let us suppose that (*wə*)*ʕattā* appears mainly at the border between indicative DU (assertive illocution) and imperative DU (directive illocution), if they are related by the rhetorical relation JUSTIFY. This is a simplified procedure, not taking into account the more complicated contexts (e. g., with indirect illocutionary force), but it gives the idea of what I mean by the "potential slot" for a DM. All such potential slots were checked in the Hebrew text of the Old Testament Book of Judges and it was found that it includes 16 dialogic utterances (turns), built as a succession of discourse units assertive-directive, where the discourse unit before (*wə*)*ʕattā* is interpreted as a justification or motivation of the following directive[22] (Jud 7:2; 9:38; 10:15; 11:36; 13:3f, 7; 14:2; 15:18; 16:10, 13; 18:14; 19:9(2), 30; 20:4–7, 12–13). Among these 16 potential slots for a discourse marker eight are filled with (*wə*)*ʕattā* (7:2; 13:3–4, 7; 14:2; 15:18; 16:10; 18:14; 20:12–13), the rest are unmarked. Here is one of the examples where the would-be slot for (*wə*)*ʕattā* is not filled:

(8) *wattōmär ʔēlāw ʔābī pāṣītā ʔät pīkā ʔäl yhwh ʕăŝē lī kaʔăšär yāṣā mippīkā ʔaḥărē ʔăšär ʕāŝā ləkā yhwh nəqāmōt mēʔōyəbäkā mibbənē ʕammōn*

She replied, [DU1] "Father, you have made a promise to Yahweh; [DU2] treat me as the promise that you have made requires, since Yahweh has granted you vengeance on your enemies the Ammonites." (Jud 11:36; NJB)

Thus, from this rather short but representative number of examples it is seen that (*wə*)*ʕattā* may be omitted in that type of contexts where its appearance is expected, so its usage is optional. It should also be noted that there is no competing DM to fill the potential slots for (*wə*)*ʕattā*. These generalizations surely need to be checked on bigger amounts of text, but I believe that the rate of filled/unfilled slots will not change drastically after the investigation of the whole corpus of Classical Hebrew.

In this analysis of the most typical usage for (*wə*)*ʕattā* I attempted to show that the main DM characteristics are applicable here; (*wə*)*ʕattā* usually marks or cues a certain rhetorical relation (JUSTIFY) and is tightly related to directive utterances, expanded by the preceding as-

---

[20] The directive illocution is expressed, besides imperative including prohibitive (134 times), also by jussive (16 times) and cohortative (7 times). According to A. Wagner [1997: 238], (*wə*)*ʕattā* is never used "vor einfachen Mitteilungen (REPRÄSENTATIVEN)".

[21] It is not a very widespread term, used, e. g., to describe the functions of English *so* [Müller 2005]. For French *car* and *puisque* as speech act markers see [Delort & Danlot 2005], following [Groupe λ-l 1975]. It was noticed also by Helbig [1988] that "discourse particles function as illocutionary indicators" (quoted in [Fischer 2006: 437]).

[22] In one of the cases (Jud 20:12) the rogative (question) has assertive force; in Jud 15:18 rogative has directive force. Both cases are interpreted in terms of indirect illocutionary force.

sertives, being the only Classical Hebrew DM with this specific function[23]. I admit that these characteristics of the discourse function of (wə)ʕattā are not exhaustive, e. g. I did not discuss some other functions and interpretations assigned to (wə)ʕattā and other now-words[24], but the higlighted features are, in my opinion, most relevant to the present comparative investigation.

Concluding this short representation of (wə)ʕattā in its discursive usage, I shall discuss some of its controversial uses. There are certain contexts where wəʕattā appears as a temporal adverbial but its discursive interpretation is not to be excluded. Typologically, it is predictable and fairly well-known from works on German *nun* and English *now* [Pérennec 2002: 342; Aijmer 2002: 59]. The problem is sometimes solved by saying that both meanings apply in such a case, but one of them prevails over another.

(9) *ūdəbar ʔabnēr hāyā ʕim ziqnē yiŝrāʔēl lēmōr gam təmōl gam šilšōm hăyyītäm məbaqšīm ʔät dāwīd ləmäläk ʕälēkäm* **wəʕattā** *ʕăŝū kī yhwh ʔāmar ʔäl dāwīd lēmōr bəyad dāwīd ʕabdī hōšiaʕ ʔät ʕammī yiŝrāʔēl miyyad pəlištīm ūmiyyad kōl ʔōyəbēhäm*

Abner sent word to the elders of Israel, saying, "For some time past you have been seeking David as king over you. **Now then** bring it about; for the LORD has promised David: Through my servant David I will save my people Israel from the hand of the Philistines, and from all their enemies." (2 Sam 3:17f)

The contrast between the moment of speaking and the past appears highlighted here. Since ʕattā is here in the same extreme left position as in cases without pronounced temporal contrast (exemplified by (7)) where discourse function of ʕattā applies, it is not focused and it must not be integrated into the clause. If its temporal adverbial meaning were focused, it would be placed at the other end of the clause. On the contrary, the focused item is imperative. The question at stake is not *when* "to bring it about", but the necessity of the action itself is highlighted here.

2.4. wəʕattā in Ancient Hebrew letters.

This section may appear to be a curious addendum, since the material observed here is very scarse (21 occurrences of *wʕt*)[25], but its significance is increased by the fact that the bulk of Aramaic and Akkadian material discussed in this paper comes from letters. The corpus of Ancient Hebrew letters is very small: it includes 50 letters, only 20 among them in relatively good condition [Schwiderski 1997: 128]. The particle *wʕt* is used very consistently in the letters, always marking the transition from the introductory part, usually containing the name of the addressee and greetings, to the body of the letter[26]:

(10) *ʔl ʔlyšb* **wʕt** *ntn lktym b 1 2 yyn lʔrbʕt hymm w 300 lḥm wmlʔ hḥmr yyn whsbt mḥr ʔl tʔḥr wʔm ʕwd ḥmṣ wntt lhm* [Ahituv & Mazar 1992: 56]

---

[23] As shown in [Lyavdansky 2007], there is the following functional distribution for different inferential discourse markers in Biblical Hebrew: (wə)ʕattā marks directive utterances; lākēn marks commissives; ʔēpō marks rogatives (interrogatives).

[24] The Hebrew (wə)ʕattā was also interpreted as "attention arouser" (Aufmerksamkeitserreger) by D. Schwiderski [1997] and in very similar terms by E. Jenni [1972].

[25] Two texts found in the territory of Edom (Horvat Uzza ostracon) and Ammon (Tell Mazar ostracon) are included.

[26] This usage is typologically significant and is paralleled by certain usages of English *now*: according to Halliday & Hasan [1976: 268]; quoted in [Aijmer 2002: 69], in 'a transaction situation such as a shop encounter, the transition from phatic communion to transactional relations is often made by *now*'.

To Eliyashib: **And now** — give to Kittiim 1 bat and 2 hins of wine for four days, and 300 [loaves of] bread, and a homer full of wine. You should send [it] out tomorrow, do not tarry. Also, if there is any vinegar left, give [it] to them.

The introductory part of a letter may also include a greeting/salutation formula:

(11) *ʔl ʔdny ʔlyšb yhwh yšʔl lšlmk* **wʕt** *tn lšmryhw…* [Ahituv & Mazar 1992: 74]
To my lord Eliyashib. May Yahweh ask for your peace! **And now** — give to Shemaryahu…

Lexical marking of the border between the introductory part of a letter and the body of the letter is a widespread phenomenon in Ancient Northwest Semitic epistolography. Thus, the use of *wʕt* in letters from Iron Age Judaea may be compared to similar use of Egyptian Aramaic *kʕn/kʕnt/kʕt*, Mishnaic Hebrew *š-* and Hellenistic Aramaic *d-*. It is not to be excluded that the ultimate origin of this phenomenon is the Akkadian epistolary style, but a cursory look through different Akkadian letter corpora does not reveal any consistent usage of a lexical marker on the border between the introductory part of a letter and its body, as it happens in Ancient Hebrew and Egyptian Aramaic letters.

The function of *wʕt* in letters is similar to its function in dialog in at least two respects:

1. *wʕt* is inserted at the transition from the subsidiary part of the text of a letter (address, salutation) to its main part, traditionally called the "body" of the letter. In terms of RST analysis the introductory part of a letter is a *satellite*, whereas the body of a letter as a whole is a *nucleus*. The same terminology is applicable to the typical dialogic utterances hosting (*wə*)*ʕattā*: assertive DU1 is a satellite, directive DU2 is a nucleus.

2. In most of its attestations in letters *wʕt* marks a transition to an utterance with directive illocutionary force: in 9 cases, before imperative, in 6 cases — before absolute infinitive with the imperative function; together with one prohibitive there are 16 volitive utterances. The rest are 4 assertives with the perfect in the main clause and one commissive (promise). It may be argued that the function of *wʕt* in letters is to mark the transition from the introductory part to the body of the letter, irrespective of the illocution of the first utterance in a letter or of the body of a letter as a whole. It is hard to come to any definitive conclusion with this scarse material, but the statistical data given above support the comparison of *wʕt* in letters with (*wə*)*ʕattā* in literary dialog.

Consequently, the epistolary usage of (*wə*)*ʕattā* is in line with its usage in dialogue. In my opinion, the genealogy of the usage of (*wə*)*ʕattā* in Classical Hebrew may be presented as follows it is born in spoken interaction and all the other uses (literary dialogue, epistolary usage, liturgical and prophetic poetry) derive from it.

It is beyond the scope of this paper to discuss the uses of (*wə*)*ʕattā* in Psalms and Prophets; let me just refer to the observations of Jenni [1972] who, not contradicting my own findings, did not notice any major deviations from the picture drawn above.

3.1. Egyptian Aramaic kʕn / kʕnt / kʕt.

There are three particles with the meaning "now" in Egyptian Aramaic[27] which are almost identical in meaning and function and often interchange in the same contexts: *kʕn, kʕnt* and *kʕt*.

---

[27] The term "Egyptian Aramaic" refers to the language of a relatively large corpus of Aramaic texts from Egypt as represented in TAD, dating from seventh to third century BCE, but the bulk of this material is dated

The nature of this variation is not altogether clear; the attempt to understand it, undertaken in [Folmer 1995: 661–71], is full of interesting observations as to the distribution of *kʕn*, *kʕnt* and *kʕt* in different epistolary archives, groups of texts and in the notation of different scribes, but it did not reach any definitive overall conclusions. One thing that these three expressions all have in common is that they are construed with the preposition *k-* attached to three different (but probably related) words: *k-ʕn*, *k-ʕnt* and *k-ʕt*. It is not a common opinion that these three words are etymologically related[28], but their usage, which will be discussed below, points to that possibility. According to Ribera i Florit [1983], the common etymology of these three expressions may be shown as follows:

*ʕadtu/ʕidtu > *ʕattu > ʕt*
*ʕadtu/ʕidtu > *ʕattu > *ʕantu > ʕnt > ʕn*

In light of their usage and probable common etymology, the expressions *kʕn*, *kʕnt* and *kʕt* are treated as allomorphs[29]. There are also variants *kʕn / w-kn*, *kʕnt / w-kʕnt* and *kʕt / w-kʕt*. The uses of *kʕt* and its allomorphs with preceding *w-* are considered as collocations (see 2.1).

3.2. Adverbial kʕn.

My observations on the 164 attestations of *kʕt* and its allomorphs in the corpus of Egyptian Aramaic [Porten & Lund 2002] show that there are only 9 clear instances where the Aramaic word for "now" is used adverbially[30], and in all of these instances only *kʕn* is attested. They are interpreted as adverbial based on criteria applied to the analysis of *(wə)ʕattā* above: in 6 cases *kʕn* appears within a prepositional phrase (*ʕd kʕn* A4.3:7; D2.29:1; D7.19:5, 7; *mn zy kʕn* A4.7:3; A4.8:2); in 2 cases there is a contrast with the past state of affairs (B3.8:41; B6.4:7). In one more case (A6.4:3) *kʕn* is found in a relative clause:

(12) *kʕn ps]mšḫ brh zy ʕḥḥpy zy **kʕn** pqyd ʕbd ḥlpwhy byn bgyʔ zyly zy bʕlytʔ wtḥtytʔ šʔl lmnšʔ dšnʔ zky z[y] mn mlkʔ wmny [y]hb lʕḥḥpy*
Now, Psamshek the son of Aḥḥapi who **now** has been made an official in his stead in my domains which are in Upper [and Lower (Egypt) asked to carry on] that grant which was given by the king and by me to Aḥḥapi (TAD A6.4:3–4).

Beyond intuitive contextual considerations, which suggest an adverbial function for *kʕn* in this context, it should be noted that, according to the principles of RST, restricted relative clauses are not considered separate discourse units. In the above example we, however, have an unrestricted relative clause. The problem is solved by the observation that DMs like *(wə)ʕattā* and *now*, with which *kʕn* is comparable, always mark a transition to nucleus. It is natural to assume that the relative clause never constitutes a nucleus. Therefore, I suppose that the appearance of *kʕn* in any relative clause forbids its interpretation as a discourse particle.

---

around the fifth century BCE. The corpus of Egyptian Aramaic is fairly representative for the larger linguistic entity usually designated as Imperial Aramaic, or Achaemenid Aramaic.

[28] On this problem see the discussion and references in [Folmer 1995].

[29] See, e. g., [Schwiderski 1997: 132].

[30] TAD A4.3:7; A4.7:3; A4.8:2; A6.4:4; B3.8:41; B6.4:7; D2.29:1; D7.19:5, 7. Here and below the indices for Aramaic texts are given according to the TAD edition, divided into four thematic volumes: A — Letters; B — Contracts; C — Literature, Accounts, Lists; D — Ostraca & Assorted Inscriptions.

The list of temporal adverbial uses of *kʕn* may be expanded by one more example (TAD C1.1:51), discussed below, where both basic functions — adverbial and discursive — are probably concomitant.

3.3. Egyptian Aramaic kʕt/kʕnt/kʕn in letters.

There are approximately 150 attestations of *kʕt* and its allomorphs in Aramaic letters included into the corpus of Egyptian Aramaic as represented in TAD. In 60 cases it is placed at the beginning of the body of the letter. Most of the other attestations (excluding 7 clearly adverbial uses) are clause-initial uses of *kʕt/ kʕnt/kʕn*, where it marks a transition from one paragraph of a letter to another, thus having a connective function in the text. For the present comparative investigation it is relevant to note that in 41 cases it marks the transition to a directive utterance; only 5 among them open the body of a letter. These usages are exemplified by the following two texts from different letter corpora:

(13) *šlm ʔwryh* **kʕn** *hlw tʔtʔ zylk rbtʔ mṭʔt lmgz ʕmrʔ zylh qdmʔ mtmrṭ bkbʔ* **kʕn** *ʔtʔ wgzh bywm zy tr ḥmnh tgznh…*
Greetings, Uriyah! **Now** your big ewe is ready for shearing. The one you sent over before is being combed. **So** you can come shear her whenever you please (TAD D7.8:1).

(14) *mn wrwhy ʕl nḥtḥwr wkndsyrm wknwth* **wkʕt** *tnh ʔnh qblt lʔršm ʕl ʔḥtbsty pqydʔ zyly zy m[nd]t[ʔ] mndʕm lʔ mhyth ly ʔḥ[r] t mhytyn bb[ʔl]* **kʕt** *ʔntm ʔtnṣḥ[w] whndrzw ʕbdw lpqydʔ [zy] ly ʕd mndt [bgyʔ ʔlk yhy]th ʕly bbʔl*
From Varuvahya to Nakhthor and Kendasirama and his colleagues. **And now**, I complained here to Arsames about Aḥatubasti my official who is not bringing me anything of [the] r[en]t. The[n…] … they are bringing to Baby[lon]. **N[o]w**, you, be diligen[t] and issue instruction to [m]y official that he [bri]ng to me to Babylon the rent of [those domains] (TAD A6.14:1–3).

Naturally, there are some local and register-related peculiarities in the usage of *kʕt*. The letters from Hermopolis apparently demonstrate a deviation from the more widespread usage. I adduce a rather lengthy example, omitting the original Aramaic text; not a single case of *wkʕt* has any equivalent in the translation:

(15) Greetings to the temple of Bethel and the temple of Queen of Heaven. To my sister Nanaiham from your brother Nabusha. I bless you by Ptah — may he let me see you again in good health! Greetings to Bethelnetan. Greetings to Nikkai, Asah, Tashai, Anati, Ati, and Reia.
wkʕt The tunic you sent me has arrived. I found it all streaked; I just don't like it at all! Do you have plenty of other kinds? If I knew, I would exchange it for a dress for Ati.
wkʕt As to the tunic which you brought for me to Syene, I wear it.
wkʕt Please have some castor oil sent to us, so we can exchange it for olive oil.
wkʕt Don't worry about me and Makkebanit; let us worry about you instead! Take care of Bethelnetan; keep Habib away from him!
wkʕt If I can find anyone dependable, I will send you something. . . . . . (TAD A2.3)

The paragraphs/units that are introduced by *wkʕt* are pragmatically of different nature — constative (the paragraph about the tunic), directive (request to send castor oil), commissive (a promise to send something). But what is more important here is that the paragraphs marked by *wkʕt* are not understood as being explicitly related to each other. Only the second *wkʕt* marks a switch to sub-topic within the paragraph; it is, apparently, a case of the rhetorical re-

lation ELABORATION in terms of RST. Nevertheless, there is still cohesion in this letter, which we may call "global cohesion": explicitly unrelated discourse units are related by having the same speaker and the same speech situation.

Thus, in the letters from Hermopolis the particle *wkʕt* becomes the default transition marker. It does not matter whether there is any specific discourse relation between the paragraphs of a letter or not; what matters is that every other unit introduced by *wkʕt* represents a new topic in the broadest sense of this word.

3.4. kʕn in dialogue.

Naturally, it is problematic to discuss the dialogic usage in a dead language, but it is possible with certain restrictions. Within strictly Egyptian Aramaic material we have only rare instances of reported speech in letters and in the Story of Ahiqar. To these scarse data we may add literary texts, written in the idiom traditionally called Biblical Aramaic (BA). Since BA in many respects follows Egyptian Aramaic[31], the Aramaic passages from Daniel and Ezra can be also included into the present discussion.

In the Aramaic Story of Ahiqar[32] *kʕn* appears in a context that is typical for a discourse marker, before directive utterance (the Aramaic text is omitted for the sake of brevity):

(16) I am Ahiqar who formerly rescued you from an innocent killing… I brought you to the house of mine. There, I was supporting you as a man with his brother, and I hid you from him. I said, "I killed him," until at [an]other time and many days later I presented you before Sennacherib the King and I removed your sins before him and evil he did not do to you. Moreover, abundantly Sennacherib the King loved me because I let you live and did not kill you.

Now (**kʕn**), you, just as I did for you, so, **then** (**ʔpw**), do for me. Do not kill me. Bring me to your ho[u]se un[til] later days (TAD C 1.1:46–52).

The temporal adverbial function of *kʕn* is probably concomitant here with its textual function. There is one more discourse marker (*ʔpw*) here; it occurs only twice in Egyptian Aramaic, but in its meaning it follows Biblical Hebrew *ʔēpō*. Its appearance in the same sentence as *kʕn* may affect the interpretation of the meaning of the latter, because if *kʕn* and *ʔpw* mark the same rhetorical relation, then the discourse function of *kʕn* will appear redundant.

The text of the following short letter, written on an ostracon, is not without problems, but it is unique, because it includes reported speech that demonstrates one of the rare examples of *kʕn* in dialogue:

(17) *ʕl ḥgy ʔmrt lʔšn ʕl ksp mrzḥʔ kn ʔmr ly lm lʔyty*[33] **kʕn** *ʔntnnh lḥgy ʔw ygdl dbr ʕlwhy wyntnhy lkm*

To Haggai: I talked to Ashina about the money for the marzeah society. He told me " [If?] there is not, **so I** will give it to Haggai or to Yigdal." So go see him and get him to give it to you![34] (TAD D7.29)

The supposed scenario behind this implies that Ashina is a sponsoring agent for the ritual communal banquet (*mrzḥʔ*); seeing that there is no (*lʔyty*) money for it, he promises the author

---

[31] Sometimes Biblical Aramaic is included into Imperial Aramaic [Beyer 1986].

[32] On the peculiar dialect of this text see [Kottsiepper 1990].

[33] RÉS: *lʔytw,* this was understood as personal name Ito (see [Lindenbrger 1993: 39]), which is hardly plausible here.

[34] Translation follows [Lindenberger 1993: 39], modified according to the emended text in [TAD 4: 177].

of the letter to give the needed money to Haggai or to Yigdal. The older interpretation 'to Ito' (*lʔyty*) would, naturally, not destroy this scenario but would not support the inferential meaning of *kʕn* in this context. This reading, which runs into certain problems, was rejected in the newer edition of the text (TAD 4: 177).

The following pragmatically complicated example from Biblical Aramaic cannot be discussed at length here, but it demonstrates some typical problems and ways to overcome them:

(18) Nebuchadnezzar said to them, "Is it true, O Shadrach, Meshach, and Abednego, that you do not serve my gods and you do not worship the golden statue that I have set up? **Now** (*kʕn*) if you are ready when you hear the sound of the horn, pipe, lyre, trigon, harp, drum, and entire musical ensemble to fall down and worship the statue that I have made, well and good. But if you do not worship, you shall immediately be thrown into a furnace of blazing fire, and who is the god that will deliver you out of my hands?" (Dan 3:13–15)

The textual units before *kʕn* and after *kʕn* are difficult to interpret pragmatically. I believe that, in this case, the interrogative in the left co-text of *kʕn* may be interpreted as an indirect assertive, because Nebuchadnezzar does not seem to expect an answer to his rhetorical question. The immediate right co-text of *kʕn* may be interpreted as a directive, because the purpose of the whole utterance is to urge three young men to worship Nebuchadnezzar's idol.

3.5. Results for Aramaic.

It should be taken into account that, generically, the discussed Aramaic textes are different in comparison with Classical Hebrew: they are mostly letters, and examples from reported speech are quite few. If we consider only letters, there are many examples of directives in the right co-text of *kʕt*: among 83 inner-body uses of this particle 36 have a directive in said position. The relatively low percentage of directive uses is explained by the observation that *kʕt*, at least in some subcorpora (Hermopolis), assumed a new generalized function: it could mark every turn (paragraph) in the letter, irrespective of the type of rhetorical relation. This new function of *kʕt* implies that it appears not only at satellite-nucleus junctures, but also at nucleus-nucleus junctures.

The function of *kʕt* in the text may be seen from a new perspective if it is compared to and contrasted with other discourse particles in Egyptian Aramaic, e. g. *ʔp, hlw, hʔ*. This work is yet to be done, but it may be preliminarily noted that *hlw* and *hʔ*, contrary to *kʕt/kʕnt/kʕn*, are used mainly with the indicative (assertive) in the right co-text, which will be relevant for the discussion of Akkadian *anumma* and *inanna*.

4.1. Akkadian words with the meaning "now".

Let me start with a quotation from [Moran 2003: 16], discussing particles in the Amarna letters from Byblos: "Note also the phrase *u inanna* in 102.24–28 and compare Heb. *wěʕattāh*, where "and now" is not temporal but interjectional."

The term 'interjectional' may look obsolete, but the remark as a whole is right to the point, because the "temporal" function of *inanna* is here taken for granted, and attention is drawn primarily to its non-adverbial usage.

If we look at what standard descriptions of Akkadian tell us about temporal adverbs, we find that there are two adverbs with the meaning "now": *inanna* and *anumma*. CAD, together with many other descriptive sources, suggests that both are used as connectors or discourse particles: the "introduce topic of a letter" [CAD 7: 144]. Thus, the situation in Akkadian looks perplexing: we have two words for "now" which are probably competing not only for temporal adverbial slots, but also for discursive slots. But things like that never happen: there must be some rule(s) of distribution for these two words. In fact, one of the recent treatments of *inanna* together with *anumma* by Loesov [2004] forcefully draws these two words apart: as to the core meaning, for *inanna* it is "now", for *anumma* it is "here". It is true that the typical adverbial usages for *inanna* may be easily shown: *inanna* is used in such prepositional phrases as *adi inanna* "until now", *ištu inanna* "from now on", but *anumma*, apparently, never appears in such contexts (e. g., all the examples in CAD are sentence-initial). Thus, it appears that — if we keep the translation "now" for lexicological purposes — *inanna* is used in both senses, 'now I' (adverbial) and 'now II' (non-adverbial, discursive), while *anumma* is used only as 'now II'.

Let us now look a little closer at some of the uses of *inanna* and *anumma*.

4.2. Akkadian *anumma* in letters.

The description of *anumma* in CAD A₂ is interpreted as follows: there is one specific usage of *anumma* in letters ("used to introduce the message, its bearer and what he brings") in different corpora and periods of Akkadian, and there are many other usages that are not classified. All the examples in CAD show that *anumma* is inserted at the beginning of a sentence; it allows to suppose that it may function (or, probably, always functions) as a discourse particle. The remark in [Huehnergard 1989: 195] that *anumma* is "a sentence-modifying adverb that introduces a new thought" and the findings of Rainey [1988] for Amarna and Loesov [2004] for Old Babylonian letters from Harmal support this supposition.

Taking into account one of the functions of Aramaic *kʕt* and Hebrew *wʕt* — to introduce the body of the letter — it is tempting to also find such a device in Akkadian, and *anumma* is one of the probable fillers of this slot. First of all, I must say that the situation with this slot in Akkadian is different when compared with Classical Hebrew and Biblical Aramaic: no one of the Akkadian letter corpora that I have looked through (core Old Babylonian, OB Mari, Amarna, Neo-Assyrian) have a regular lexical device to mark the transition from the introductory part to the body of the letter. The assertion found in CAD that "the Mari letters use *anumma* to introduce the first topic of a letter" is not altogether wrong (see ARM 5 5:4, 9:4, 13:5, 41:4, 78:5; ARM 4 3:5), but it must not be understood in the sense of the above discussion; in Mari letters the first topic is introduced sometimes by *šanītam*, sometimes by *aššum* and very often does not have any specific lexeme at the transition to the body of the letter. A similar picture is found in the rest of the letter corpora mentioned above.

If we turn to instances of *anumma* inside the body of the letter, we find that, most often, it introduces an assertive utterance with indicative verb forms in the main sentence. Looking a little beyond the immediate right context, we find that *anumma* quite often appears before injunctive utterances, but in all of these cases *anumma* is separated from the injunctive by an assertive sentence.

In the letters of Hammurapi, analyzed by Sallaberger, among 170 letters with a directive utterance (ANORDNUNG) in their nucleus there are 30 where *anumma* is detached from an injunctive by one sentence or clause, which may be considered a discourse unit. Here is one of these letters:

(19) [*a-na*] *ᵈEN.ZU—i-din-nam* [*qí*]-*bí-ma um-ma ḫa-am-mu—ra-bi-ma aš-šum* ᵖ *ᵈEN.ZU—ra-bi ša it-ti* ᵖ*nu-úr—
    išt[*ar*] *ta-at-ru-da-aš-šu* ᵖ *ᵈEN.ZU—ra-bi šu-a-ti a-na ma-aḫ-ri-ia ú-še-ri-bu-nim-ma aš-šum i-din-ᵈEN.ZU ú-
    lam-mi-da-an-ni **a-nu-um-ma** *ᵈEN.ZU-ra-bi šu-a-ti a-na ṣe-rí-ka aṭ-ṭar-dam* ᵖ *ᵈE*[*N.Z*]*U ù* ˡᵘ*ši-i-bi ša i-qá-ab-bu-
    kum a-na ma-aḫ-ri-ia ṭu-ur-dam

    Zu Sîn-iddinam sprich: also (sagt) Hammu-rabi: Was den Sîn-rabi betrifft, den du mit Nūr-Ištar zu mir
    geschickt hast, (so) hat man diesen Sîn-rabi vor mich geführt und er hat mir betreffs Iddin-Sîn berichtet.
    Hiermit schicke ich **jetzt** den betreffenden Sîn-rabbi zu dir. Schicke den Iddin-Sîn und die Zeugen, die
    er dir nennen wird, zu mir. (AbB II 2)

It is impossible to see from just one example, yet it appears that this usage is close to for-
mulaic, because the sentence immediately following *anumma* in these contexts is restricted as
to its content and lexicon [Sallaberger 1999: 146]. From the discussion in [Sallaberger 1999:
146f] it is seen that the DU introduced with *anumma* (Sallaberger calls it "Initiative") is de-
pendent on the following directive DU[35]; thus, they are related as satellite and nucleus. It is
also clear that *anumma* marks the transition from the informative part of the letter ("In-
formieren") to "Initiative". This metatextual (= discourse) function of *anumma* is discussed in
[Loesov 2004] and I do not go into it here, although it is also relevant for the discussion.

The corpus of Akkadian letters from Mari from the time just before Hammurapi (1792–
1750 BCE), or coterminous with his rule, demonstrates another type of formulaic or quasi-
formulaic usage of *anumma* in accounts of prophecies that are found in letters. Usually at the
very end of the letter there appears a phrase, introduced with *anumma* and often followed by
the injunctive:

(20) *sinništum šī annêtim idbubamma* [*aw*]*āt pīša ana bēlīya ašpuram* **anumma** *šārassa u sissiktaša ana bēlīya ušābi-
    lam bēlī têrētim lišēpišma ana kī ilum bēlī ippalu lī*[*p*]*uš*

    This is what this woman said, and I have written her [wor]ds to my lord. I have **herewith** sent her hair
    and a fringe of her garment to my lord. My lord should let oracles be taken. Let my lord act according
    to what the god answers. (Nissinen 2003 27:17–31 = ARM 27 217:27–31)

The scenario behind these phrases implies that the author of the letter attaches prophet's
hair and garment "to be used as representing the prophet during the process of authenticating
the prophecy by extispicy ("oracles" in our letter — A.L.)" [Nissinen 2003: 16]. Thus, the as-
sertive after *anumma* is related to the following injunctive, and the relation between these two
DUs may be interpreted as JUSTIFY. Among 50 prophetic letters from Mari adduced by Nissi-
nen there are 11 letters where *anumma* appears in the described type of context.

Both groups of contexts — the letters of Hammurapi and prophetic letters from Mari —
are quasi-formulaic, but they are different in their Sitz im Leben and are not dependent on
each other; therefore, I conclude that they are based on free or 'natural', non-formulaic usage
of *anumma*.

There are also some *anumma* + Perfekt contexts, where its usage goes beyond formulaic,
because the phrase after *anumma* speaks about the actual affair (Sallaberger's "Thema") of the
letter and is therefore intimately related to the following imperative (e. g. AbB XIII, 10).

Note that in the above examples the transition from assertive to injunctive (directive) was
never marked. This does not always happen in the Akkadian directive utterances, as will be
seen in the following paragraph. It was important to show that *anumma* in that sort of contexts
does not appear in the slot filled by Hebrew *wʕt* and in many cases by Aramaic *kʕt*; it appears in

---

[35] "Dieses Teil spielt als … Übergang vom Informations- zum Aufforderungsteil" [Sallaberger 1999: 146].

the slot filled in Hebrew directive utterances by *hinne* (Jud 16:10; 19:9; 20:7), in Aramaic sometimes by *hlw*.

4.3. Akkadian *inanna* in reported speech and in letters.

A very detailed discussion of *inanna* in Old Babylonian (OB) by Loesov [2004] seems to me sufficient to demonstrate the various discourse functions of this temporal adverb, though the approach to the description of *inanna* chosen by Loesov is different from mine. For Loesov's analysis, it was the consideration of concomitant verb forms that served as the decisive factor. I admit that the shift from one type of verb form to another is significant for the discourse structure. At the present stage of my research I do not take into account the shifts within the indicative sphere; I am interested in usages which mark (or are somehow related to) the shift in the verbal mood or in the illocutionary force of the utterance.

If we look at the uses of *inanna* in reported speech, we find that it mostly appears at the transition from one type of speech act to another. I have checked all the attestations of *inanna* (*eninna*) in the Standard Babylonian (SB) version of The Epic of Gilgamesh. The "corpus" of reported speech in The Epic of Gilgamesh is not big, but in relation to all reported speech that may be found in OB or SB literature it is hardly small. According to the edition of A. George [2003], there are 16 unrestored attestations of *eninna* in the SB version of The Epic of Gilgamesh, plus one in the Ischchaly tablet. In the majority of cases *eninna* marks a shift in illocution: 5 times from assertive or interrogative[36] to directive (I 96; V 102; V 180= V 238; XI 198), 3 times from assertive to interrogative (VII 59; VIII 55; X 73, 150; XI 207), one time from assertive to declarative (III 122), once from interrogative to assertive (III 47), once from interrogative to commissive[37] (VII 139). 3 times it is turn-initial (Ischchaly 11 — directive; X 73, 150 — both interrogatives). Note the slight prevalence of directives in the right co-text; however, the data are too limited to reach any definitive conclusions. There are more attestations of *inanna* before directives among the examples from OB letters, discussed in [Loesov 2004].

The usage of *eninna* before directives may be demonstrated by the following text with the injunctive form in the immediate right co-text:

(21) *at-ti ᵈa-ru-ru tab-ni-[i amēla (lú)]* **[e]-nin-na** *bi-ni-i zi-kir-šú*
      'You, O Aruru, cteated [man:] **now** create what he suggests! (I 95f)

A more complicated case is represented by the following passage, where the injunctive verb forms are not in the immediate right co-text of *eninna*:

(22) 100 *am-mi-ni ib-ri pi-is-nu-qiš* [*ta-qa*]*b-bi*
      101 *ù pi-i-ka ir-ma-am-ma tu-lam-*[*man l*]*ib-bi*
      102 **e-nin-na**-*ma ib-ri iš-ta-at* [(x)]-*pi?-*[x]
      103 *ina ra-aṭu ᴸᵘSIMUG e-ra-⟨a⟩ šá-ba-šá-a*
      104 *tu-ú-ru ana 1 DANNAᵃᵐ na-pa-ḫu na-pi-iḫ-tu ana 1 DANNAᵃᵐ x-ˊlu-úˋ*
      105 *šá-par a-bu-bu iš-tuḫ-ḫu la-pa-tu*
      106 [*e*] ˊtaˋ-*as-suḫ GIRᵐⁱⁿ-ka e ta-tu-ur ana ár-ki-ka*

---

[36] Interrogatives are usually not included within the classification of Speech Acts [Allan 1998]. They need further analysis, e. g. rhetorical questions are often assertives or directives.

[37] The phrases in the following passage (140–147) constitute a promise to Enkidu.

107 [. . . . . . .]x x x *mi-ḫi-iṣ-ka du-un-nin*

100 'Why, my friend, [do you] speak like a weakling?

101 With your spineless words you [make] me despondent.

102 '**Now**, my friend, but one is [our task,]

103 the copper is already pouring into the mould!

104 To stoke the furnace for an hour? To … the coals for an hour?

105 To send the Deluge is to crack the whip!

106 '[Don't] draw back, don't make a retreat!

107 …… make your blow mighty!' (V 100–107)

If it is possible to say that a DM is hosted by an utterance or a discourse unit, it should be noted that a discourse unit may include from one up to several clauses or discourse units of lower rank. In the above text, *eninna* marks a transition to that part of the text where directive force is dominating; in terms of RST, the passage (106–107) with injunctives is a nucleus with two satellites — (102) and (104–105)[38]. Thus, *eninna* focuses the hearer's attention upon the whole passage (102–107), organized as a complex directive utterance.

It is also to be noted that assertives are rare in the right co-text of *eninna* in reported speech, as represented by the SB version of Gilgamesh. The significant amount of interrogatives draws our attention in this case, because they are rare in the right co-text of Hebrew (*wə*)*ʕattā* and Aramaic *kʕt*. Apparently, the functions of *inanna* are broader compared to the *now*-words in Hebrew[39] and Aramaic.

To conclude with *inanna*, let me briefly discuss the analysis of its meaning suggested in [Loesov 2004]. I quote the results of this analysis in a slightly shortened manner, omitting the references to examples within Loesov's article:

> These and related data (presented below) permit one to posit two inanna-lexemes:
>
> *inanna*$_A$: temporal deictic adverb, further subdivided into
>
> *inanna*$_{A_1}$ pointing to the moment of speaking, used with the Present, employed in non-future sense and with the semantically "present-tense" Stative: … (speaker-orientation);
>
> *inanna*$_{A_2}$ pointing to the moment of speaking; it is used with the Perfect and locates its resultative component: … (speaker-orientation);
>
> *inanna*$_{A_3}$ pointing to the future, used with injunctive forms and with E[pistolary] P[er]f[ect]: … (addressee-orientation).
>
> *inanna*$_B$: metatextual "particle" marking a turn in discourse, i. e. a means of discourse deixis. It is formally set apart from *inanna*$_A$ through the combination of two features: *inanna*$_B$ is used only with the Preterite and, unlike *inanna*$_A$, is incompatible with injunctive utterances in its immediate right context/co-text. [Loesov 2004: 96]

As stated by the author, one of the principles of this classification is "to illuminate the verb usage". I am not against the postulation of two distinct *inanna*-lexemes, but the distinction between *inanna*$_{A_3}$ and *inanna*$_B$ seems to be exaggerated, because *inanna*$_{A_3}$ also marks "a turn in discourse". Moreover, *inanna*$_B$ is also orientated to the addressee: if *now* is metatextual (*now* II), it is that type of *now*, which is shared by the speaker and the addressee, denoting textual time which is common to both participants of the communicative act.

---

[38] These are proverbial phrases which serve to enhance the rhetorical effect of Enkidu's speech as the latter tries to urge a frightened Gilgamesh to "swift action" [George 2003: 467].

[39] Interrogatives related to the foregoing assertives by the relation JUSTIFY are sometimes marked in Hebrew by *ʔēpō.*

4.4. Results for Akkadian.

According to my observations, there are components of meaning that are common to both *inanna* and *anumma*, and their frequent appearance together (the collocations *inanna anumma* or *anumma inanna*) is one of the facts that prove this. Both *anumma* and *inanna* draw the attention of the addressee to what happens in the text. Or, in the words of Deborah Shiffrin (said about English *now*), they "focus attention on what the speaker is about to do" [Schiffrin 1987: 241]. Another side of this function is that they mark a turn in the discourse, a turn which is marked also by a shift from one verb form to another, with or without a shift in verbal mood. When *inanna* or *anumma* introduce an assertive utterance, the difference between them is yet to be clarified[40]. There are certain contexts where *anumma* would not usually appear (purely temporal adverbial usages; immediately before directives, interrogatives and commissives). The comparison of *anumma* to Hebrew *hinne* is probably restricted to just one type of contexts, where *anumma* is in the vicinity of a directive DU but is separated from it by an assertive DU. The percentage of such uses for *anumma* should be checked on a wider textual basis, but it is unlikely to be small. The comparison with Hebrew *hinne* may also be taken into consideration when solving the problem of the origins of *anumma*. As to the similarity in usage between Hebrew (*wə*)*ʕattā* and Akkadian *inanna* (examples from reported speech), I think it is possible to interpret it as a parallel development from a temporal deictic adverb to a discourse particle, which marks a transition from one illocution (usually assertive) to another (usually directive) for discourse units related to each other by the rhetorical relation JUSTIFY. This comparison is made with one reservation: not all instances of *inanna* are comparable to Hebrew (*wə*)*ʕattā*.

5. Conclusion.

As the above investigation has shown, there are some common developments in the usage of Hebrew (*wə*)*ʕattā*, Aramaic *kʕt*/*kʕnt*/ *kʕn* and Akkadian *inanna*. Used as temporal deictic adverbs with an extralinguistic reference to the present time, they are more often attested in contexts where their adverbial meaning is bleached, combined with discursive (metatextual) meaning or does not apply at all, ceding its place to their discursive meaning. The aforementioned 'now-words' in three Semitic languages are thus included into the class of words with a salient discourse function, usually called 'discourse markers'.

Triggered by the specific usage of Hebrew (*wə*)*ʕattā*, the focus of this study primarily concerned directive utterances. It was noticed quite a long time ago by Teun van Dijk [1979], one of the "fathers" of discourse studies in Europe, that directive utterances are often preceded by assertives, helping the speaker to fulfill her/his communicative goals, e. g. to make a request/demand/ more acceptable to the addressee. Most often, the assertive discourse unit is placed before the directive discourse unit; transition from assertive to directive constitutes the slot for a discourse marker. Languages vary not only as to the type of lexemes that can fill this slot, but also as to the percentage of slots filled with a DM. Thus it appears that among the three discussed languages this slot is filled most frequently in Classical Hebrew with (*wə*)*ʕattā*, less often in Egyptian Aramaic with *kʕt* and (sometimes) in some varieties of Akkadian with *inanna*. In all three languages there are no lexemes that could be considered as really competing for this slot: the lexemes at the focus of this study are the main fillers of the slot for the di-

---

[40] See the discussion and references in [Loesov 2004], an important step in this direction.

rective speech act marker. This type of contexts was taken as the basis and the starting point for the comparison of the discussed 'now-words'.

An attempt was made to demonstrate in what respect the usage of these words is different. Thus, in Egyptian Aramaic the temporal adverb *kʕt/kʕnt/ kʕn* assumes specific functions in letters: it is used as a transition marker to the body of the letter[41], in some of the corpora it is used as the marker of a new topic (paragraph), irrespective of the type of rhetorical relation between paragraphs. In Standard Babylonian *inanna* often marks transition from different illocutions to interrogatives, and there are also uses of *inanna* before commissive utterances and declarations[42]. In general, the usage of *inanna* is significantly broader than that of *(wə)ʕattā* and *kʕt/kʕnt/ kʕn* as far as the existing evidence shows us: *inanna* very often appears before assertives, e. g. in the narrative part of letters. The distribution of *inanna* and *anumma*, both appearing before assertives, is not altogether clear; at present it appears that *anumma*, sharing with *inanna* only the discursive meaning "now" (i.e. *now* II), is a different type of lexeme. It is never used as a temporal deictic adverb and may be compared in its usage to Biblical Hebrew *hinnē* and Egyptian Aramaic *hlw* and *hʔ*.

Finally, I would like to define the field for further research. The discussed discourse markers in Hebrew, Aramaic and Akkadian should be considered as members of their class; their function may be further clarified by juxtaposing them with other DMs such as *hinnē*, *lākēn*, *yaʕan* and *kī* in Classical Hebrew. The use of discourse particles in Aramaic letters is a promising field of study, especially because it may be compared to their use in Akkadian letters, with which there are clear parallels [Fales 1987].

Literature

AbB — *Altbabylonische Briefe* (series).

ADP — FISCHER, K. (ed.) (2006) *Approaches to Discourse Particles*. Amsterdam: Elsevier.

ARM — *Archives royales de Mari* (series).

Ahituv & Mazar 1992 — AHITUV Sh. & MAZAR, A. (eds.) (1992) *Handbook of Ancient Hebrew Inscriptions*. Jeruasalem: The Bialik Institute (Mosad Bialik).

Aijmer 1988 — AIJMER, K. (1988). Now may we have a word on this: the use of *now* as a discourse particle // KYTO M. et al. (ed.) *Corpus linguistics: hard and soft*. Amsterdam and Atlanta: Rodopi.

Aijmer 2002 — AIJMER, Karin. 2002. *English Discourse Particles: Evidence from a Corpus*. Amsterdam / Philadelphia: John Benjamins.

Allan 1998 — ALLAN, K. Speech Act Classification and Definition // J.L. MEY (Ed), *Concise. Encyclopedia of Pragmatics*. Amsterdam: Elsevier, pp. 922–927.

Bayerl 2004 — BAYERL P.S. Disambiguierung deutschsprachiger Diskursmarker: Eine Pilot-Studie // *Linguistik online* 18.1 (2004).

Beyer 1986 — BEYER, K. (1986) *The Aramaic Language. Its Distribution and Subdivisions*. Göttingen: Vandenhoek & Ruprecht.

Brongers 1965 — BRONGERS, H.A. (1965). Bemerkungen zum Gebrauch des adverbialen *we'atta* im Alten Testament // *Vetus Testamentum* 15(3).

Brunaud 1991 — BRUNAUD, N. (1991) "Now" et "then" dans la coherence argumentative // *Langages* 104, p. 79–91.

Delort & Danlot 2005 — DELORT, L. & DANLOT, L. (2005). Coordination of Causal Relations in Discourse // AURNAGUE, M.; BRAS, M.; LE DRAOULEC, A. & VIEU, L. (eds.), *Proceedings of Simposium on the Exploration and Modelling of Meaning* (SEM'05). Biarritz, p. 75–84.

Fales 1987— FALES F.M. Aramaic Letters and Neo-Assyrian Letters: Philological and Methodological Notes // *JAOS* 107.3 (1987).

---

[41] Also attested for Hebrew *(wə)ʕattā*, but the evidence is too sparse.

[42] Sometimes attested for Classical Hebrew *(wə)ʕattā*.

Fischer 2006 — FISCHER K. (2006). Frames, constructions, and invariant meanings: the functional polysemy of discourse particles // FISCHER, K. (ed.) *Approaches to Discourse Particles*. Amsterdam: Elsevier.

Folmer 1995 — FOLMER, M.N. (1995). *The Aramaic Language in the Achaemenid Period: A Study in Linguistic Variation* (Oreintalia Lovaniensia Analecta 68). Leuven.

Fraser 2006 — FRASER, B. (2006) Towards a Theory of Discourse Markers // K. FISCHER (ed.), *Approaches to Discourse Particles*. Amsterdam: Elsevier.

George 2003 — GEORGE, A.R. (2003). *The Babylonian Gilgamesh epic: introduction, critical edition and cuneiform texts*. Oxford University Press.

Groupe λ-l 1975 — Groupe λ-l. Car, parce que, puisque // *Revue Romane*, 10.2, p. 248–280.

Huehnergard 1989 — HUEHNERGARD, J., (1989). *The Akkadian of Ugarit* (Harvard Semitic Studies 34). Atlanta: Scholars Press.

Jenni 1972 — JENNI, E. (1972). Zur Verwendung von ʿattā "jetzt" im Alten Testament // *Theologische Zeitschrift* 28, pp. 5–12.

Kottsieper 1990 — KOTTSIEPER, I. (1990). *Die Sprache der Ahiqarsprüche* (BZAW 194), Berlin — New York.

Lande 1949 — LANDE, I. (1949). *Formelhafte Wendungen der Umgangssprache im Alten Testament*. Leiden: Brill.

Laurentin 1964 — LAURENTIN, A. (1964). *Weʾattah — kai nun*. Formule charactéristique des textes juridiques et liturgiques à propos de Jean 17,5 // *Biblica* 45, pp. 168–97.

Lindenberger 1993 — LINDENBERGER, J. (1993). *Ancient Aramaic and Hebrew Letters* (SBL Writings from the Ancient World Series 4), Atlanta: Society of Biblical Literature.

Loesov 2004 — LOESOV, S. (2004). T-Perfect in Old Babylonian: The Debate and a Thesis // *Babel & Bibel* 1.

Lyavdansky 2007 — LYAVDANSKY, A. Inferential Markers in Biblical Hebrew, presented in the *Linguistics and Biblical Hebrew Section at the Society of Biblical Literature Internetional Meeting*, July 2007, Vienna.

Mann & Thompson 1988 — MANN, B. & THOMPSON, S. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization // *Text*, 8(3), pp. 243–281.

Miller 1999 — MILLER, C. (1999) The Pragmatics of waw as a Discourse Marker in Biblical Hebrew Dialogue. *Zeitschrift für Althebraistik* 12,2, pp. 165–191.

Moran 2003 — MORAN, W. (2003) Amarna Studies // *Collected Writings* (Harvard Semitic Studies 54). Winona Lake, Ind. : Eisenbrauns.

Müller 2005 — MÜLLER, S.(2005). *Discourse Markers in Native and Non-native English Discourse* (Pragmatics & Beyond New Series 138). Amsterdam-Philadelpia: John Benjamins Publishing House.

Muraoka 1985— MURAOKA, T. (1985). *Emphatic Words and Structures in Biblical Hebrew*. Jerusalem-Leiden.

NJB — *New Jerusalem Bible.*

Nissinen 2003 — NISSINEN, M. (2003). *Prophets and Prophecy in the Ancient Near East* (Writings from the Ancient World Society of Biblical Literature — WAWSBL 12). Atlanta: Society of Biblical Literature.

Pérennec 2002 — PÉRENNEC, M. (2002). Von Zeitdeiktika zu Text- und Diskurskonnektoren. Überlegungen zur sprachlichen Tämporalität // PÉRRENC, M., *Sur le text: énonciation et mots du discours en allemand*. Presses Universitaires de Lyon, pp. 339–356.

Porten & Lund 2002 — PORTEN, B. & LUND, J.A. (2002) *Aramaic Documents from Egypt: A Key-Word-In-Context Concordance*. Winona Lake: Eisenbrauns.

Rainey 1988 — RAINEY, A.F. (1988). Some presentation particles in the Amarna Letters from Canaan // *Ugarit-Forschungen* 20, p. 209–220.

Redeker 2006 — REDEKER, G. (2006). Discourse markers as attentional cues at discourse transitions // *Approaches to Discourse Particles*. Amsterdam: Elsevier.

RÉS — *Répertoire d'épigraphie sémitique.* 1900–68, 8 vols. Paris: Imprimerie Nationale.

RHWUD — *Random House Webster Unabridged Dictionary on CD-ROM*. 1999.

Ribera i Florit 1983 — RIBERA I FLORIT, J. (1983) Evolución morfológica y semántica de las partículas kʿn y ʔry en los diversos estadios del arameo // *Aula Orientalis* 1, pp. 227–33.

Sallaberger 1999 — SALLABERGER, W. (1999) *"Wenn Du mein Bruder bist, …". Interaktion und Textgestaltung in altbabylonischen Alltagsbriefen*. Groningen: Styx.

Schiffrin 1987 — SCHIFFRIN, D. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.

Schourup 1999 — SCHOURUP, L. (1999) Discourse markers // *Lingua* 107, p. 227–265.

Schwiderski 1997 — SCHWIDERSKI, D.(1997) 'Wer ist dein Knecht? Ein Hund!'. Zu Aufmerksamkeitserregern und Überleitungsformeln in hebräischen Briefen // WAGNER A. (ed.) *Studien zur hebräischen Grammatik*

(Orbis Biblicus et Orientalis 156). Freiburg: Universitätsverlag; Göttingen: Vandenhoeck & Ruprecht, pp. 127–41.

Taboada 2006 — TABOADA, M. (2006). Discourse markers as signals (or not) of rhetorical relations // *Journal of Pragmatics* 38(4), pp. 567–592.

TAD — PORTEN, B., & YARDENI, A., (1986–1999) *Textbook of Aramaic Documents from Ancient Egypt*. Vol. 1: Letters. Vol. 2: Contracts. Vol. 3: Literature, Accounts, Lists. Vol. 4: Ostraca & Assorted lnscriptions. Jerusalem: The Hebrew University.

van Dijk 1979 — VAN DIJK, T.A. (1979). Pragmatic connectives // *Journal of Pragmatics* 3, pp. 447-456.

Wagner 1997 — WAGNER, A. (1997) *Sprechakte und Sprechaktanalyse im Alten Testament* (Beihefte zur Zeitschrift fur die alttestamentliche Wissenschaft 253). Berlin: Walter de Gruyter.

Во многих языках мира дейктические наречия времени претерпевают категориальный сдвиг: во многих случаях они употребляются как дискурсивные маркеры. В трех семитских языках — древнееврейском, имперском арамейском и аккадском — отмечены общие черты в употреблении наречий времени, указывающих на момент речи. Данная статья посвящена сравнению ненаречных употреблений (*wə*)*ʕattā* в древнееврейском, *kʕt/kʕnt/kʕn* в имперском (египетском) арамейском, а также *inanna* и *anumma* в аккадском. Теоретико-методологическую базу данного исследования составляет Теория риторической структуры и традиция изучения дискурсивных маркеров. Предварительные результаты показывают, что указанные наречия в большинстве своих употреблений функционируют как дискурсивные маркеры. На основе анализа употреблений древнееврейского (*wə*)*ʕattā* в качестве главного объекта сравнения выдвинута специфическая дискурсивная функция: маркирование границы между ассертивным дискурсивным отрезком и директивным дискурсивным отрезком в директивных высказываниях. Отмечено, что как *kʕt/kʕnt/kʕn*, так и *inanna* употребляются в данной функции, однако сфера их употредления шире, чем у древнееврейского (*wə*)*ʕattā*, что требует дальнейшего изучения. *Anumma* представляет собой лексему иного типа: *anumma* не используется как дейктическое наречие времени; употребление этой частицы также в ряде случаев связано с директивными высказываниями, но в отличие от (*wə*)*ʕattā*, *kʕt/kʕnt/kʕn,* правый по отношению к *anumma* дискурсивный отрезок, как правило, является ассертивным, а переход от ассертивного отрезка к директивному в таких случаях остается немаркированным.

Alexander Militarev
Russian State University for the Humanities / Santa Fe Institute

A complete etymology-based hundred wordlist of Semitic updated: Items 1–34

The paper presents a detailed etymological analysis of the first 34 lexical items on the Swadesh 100-wordlist as attested in most of the living and extinct Semitic languages, aiming at a maximally precise lexical reconstruction of these items for Proto-Semitic as well as intermediate stages (West Semitic, South Semitic, etc.). All the etymologies are meticulously accompanied with evaluations of alternative possibilities of reconstruction, potential external parallels in other Afroasiatic languages, and — occasionally — discussions of a more generally methodological character.

Keywords: Semitic languages, lexicostatistics, Swadesh list, etymology, lexical reconstruction.

This study is the author's second attempt at compiling a complete one hundred wordlist ("Swadesh's List") for most Semitic languages, fully representing all the branches, groups and subgroups of this linguistic family and including the etymological background of every item whenever possible. It is another step toward figuring out the taxonomy and building a detailed and comprehensive genetic tree of said family and, further, of the Afrasian (Afroasiatic) macro-family with all its branches on a lexicostatistical basis.

Several similar attempts, including those by the author (Mil. 2000, Mil. 2004, Mil. 2007 and Mil. 2008), have been made since Morris Swadesh introduced his method of glottochronology (Sw. 1952 and Sw. 1955). In this paper, as well as my previous studies in genetic classification, I rely on Sergei Starostin's method of glottochronology and lexicostatistics (Star.) which is a radically improved and further elaborated version of Swadesh's method. One of the senior American linguists told me he had heard from Swadesh that his goal was "to get the ball rolling". I am absolutely sure that in a historical perspective this goal should be regarded as brilliantly achieved in spite of all criticism, partly justified, of Swadesh's method from various points of view.

That said, it is no secret that Swadesh did not care much about regular sound correspondences, the quality of etymologies or the problem of borrowing (being, in these aspects, very close to the mass comparison method authored by J. Greenberg[1]) in his diagnostic lists. This negligence toward the fundamental principles of the comparative method was unfortunately

---

[1] Joseph Greenberg, an outstanding American linguist who recently passed away at a respectable age (one of the creators of linguistic typology, a pioneer in the area of root-internal phonotactics as well as plenty of others) introduced this method as a way to envisage the preliminary and approximate genetic classification of linguistic families that comprise a huge number of languages, poorly studied in the comparative aspect, with "relatively little carnage" — without establishing sound correspondences and reconstructing protolanguage states. Endowed with a remarkable intuition, Greenberg has advanced far ahead that path, which cannot be said for most of his followers, few as they are, whose handling of the mass comparison method is as distinct from the much more labor-intensive comparative-historical method (which the Moscow school steadfastly holds on to) as the job of a lumberjack is distinct from that of a jeweler — and thus, somewhat discredits the very idea of distant language affinity in the eyes of the skeptics.

inherited by most of the students who have so far applied lexicostatistics to Afrasian (V. Blažek being a conspicuous exception). Even those who have claimed to follow these principles practically never adduce consistent etymological arguments in favor of their cognate scoring decisions[2]. (I regret to say that my own earlier studies, with their scarce and brief etymological remarks and only occasionally reconstructed protoforms, are no exception from this lamentable rule.)

Starostin's method, in my opinion, yields far more coherent results; however, it requires a thorough etymological analysis to distinguish between inherited and borrowed lexemes. His rule concerning the latter is that a loanword, if, of course, reliably qualified as such, (1) when matching the inherited lexeme in a related language, should not be scored as its cognate (or counted as a +), and (2) when not matching the corresponding inherited lexeme in a related language, should not be scored as its non-cognate (or counted as a -), (3) when matching another loanword in a related language, should not be scored as its cognate, and (4) in all the above cases it should be *eliminated* from the scores (counted as *0*), therefore equaling the not infrequent case of a lexeme missing in a given language in a given position on the 100-wordlist.[3]

This paper is an attempt to meet these requirements to the extent that the present state of comparative Semitic linguistics allows, and supply the scoring choices, wherever possible,

---

[2] In view of these considerations, I was surprised at the publication in the Proceedings of the Royal Society (B — Biological Sciences) of a study, obviously arranged as a novel discovery and a serious breakthrough in scholarship, by Andrew Kitchen, Christopher Ehret, Shiferaw Assefa and Connie J. Mulligan, entitled "Bayesian phylogenetic analysis of Semitic languages. Supplementary data identifies an Early Bronze Age origin of Semitic in the Near East" (*Proc. R. Soc. B* published online 29 April 2009). The study refers to a supplement containing a modified version of the Swadesh list that includes 96 words for 25 extant and extinct Semitic languages, compiled by Chr. Ehret and subjected to "Bayesian phylogenetic analysis". While the choice of the most representative lexemes for each language is also fraught with multiple problems, it is the etymological aspect, the basis of the scoring, that serves as argumentation for this or that etymological/scoring decision and is responsible for the resulting genealogical tree and the chronology of branching for a given linguistic phylum. Without this argumentation, the application of any methods, be it Bayesian-based phylogenetics, or the old Swadesh or Starostin methods or any others, no matter how advanced and sophisticated, remains fruitless: it is calculating nothingness. Being well acquainted with Prof. Ehret's work, I am more than assured that, when (and if) his etymological/scoring argumentation comes to light, there will be an enormous number of debatable — and objectable — issues; I am fully prepared to participate in these debates. Until this has happened, I can regard the sensational study in question only in a Shakespearean light, as "much ado about nothing". Another detail that struck me was the absence of several of my studies on the subject (SED I, XV–XVI, etc.) from the list of sources referred to. This is more than strange, not only because of the incompleteness of references, but also in view of the fact that some of the non-trivial results, presented in the quoted paper and obtained in my studies, surprisingly coincide in regard to both classification and chronology.

[3] A conclusion to which both of us, Starostin and myself, came independently and, surprisingly, simultaneously (somewhere around 1984) after much hesitation and checking. I was finally convinced by the following: Tigre and Amharic, although undoubtedly belonging to the same (Ethiopian) group of Semitic, yielded incoherent results when compared lexicostatistically with Jibbali or Mehri: Tigre showed a much closer cognation with the latter languages than Amharic. That was simply impossible: a well-known Russian-Jewish joke tells us that the distance from Zhmerinka to Odessa cannot be longer than the distance from Odessa to Zhmerinka. The absurd situation that first seemed a deadblock for the whole method, cleared up only after I had eliminated the loanwords from the Ethiopian lists: 13 or 14 Cushitisms from Amharic (*wušša* 'dog', *ṭäṭṭa* 'drink', *žoro* 'ear', *laba, läboba* 'feather', *asä* 'fish', *ṭägur* 'hair', *gulbät* 'knee', *awwäḳä* 'to know', *səga* 'meat', *ṭənnəš* 'small', *dəngay* 'stone', *čəra* 'tail', *zaf* 'tree', probably *wəha* 'water') and only four Cushitisms (*čəgär* 'feather', *ʕasa* 'fish', *čəgär* 'hair', *səga* 'meat') and one Arabism (*näfär* 'person') from Tigre. The lists, now reduced to 86–87 (Amharic) and 95 (Tigre) items, showed quite an even result for Amharic and Tigre, on one hand, and Jibbali and Mehri, on the other. The distance between Odessa and Zhmerinka turned out to be the same from both ends, and the method was — luckily, not posthumously — rehabilitated.

The following dates (some of them fairly conventional, some chosen after much hesitation and discussions with specialists in individual languages) have been attributed to individual languages: Akkadian, *1450* B.C.E.; Ugaritic, *1350* B.C.E.; Hebrew, *650* B.C.E.; Phoenician *850* B.C.E.; Biblical Aramaic, *200* B.C.E.; Palestinian Judaic, *200* C.E.; Syrian Aramaic, *200* C.E.; Mandaic, *750* C.E.; Urmian Aramaic *1900*; Qur'anic Arabic, *600* C.E.; Lebanese Arabic, *2000*; Meccan Arabic, *2000*; Maltese Arabic, *2000*; Sabaic, *200* B.C.E.; Geʕez, *500* C.E.; Tigrai, *2000*; Tigre, *2000*; Amharic, *2000*; Argobba, *2000*; Gafat, *1900*; Soddo, *2000*; Harari, *2000*; Wolane, *2000*; Chaha, *2000*; Harsusi, *2000*; Mehri, *2000*; Jibbali, *2000*; Soqotri, *1950*.

Abbreviations of languages, language periods and sources:

Afras. — Afrasian (Afroasiatic, Semito-Hamitic); Akk. — Akkadian; Amh. — Amharic; Arb. — Arabic; Arg. — Argobba; Arm. — Aramaic; BD — Book of the Dead; Brb. — Berber; Bib. — Biblical Aramaic; C. — Central; Chad. — Chadic; Clas. — Classical; Cush. — Cushitic; Daṯ — Daṯna Arabic; Dem. — Demotic; Ḍof. — Ḍofar; Dyn. — Dynasty; E. — East; Egyp. — Egyptian; ESA — Epigraphic Sout Arabian; Eth. — Ethiopian; Gaf. — Gafat; Gez. — Geʕez; Gur. — Gurage; Har. — Harari; Ḥḍr — Ḥaḍramaut; HEC — Highland East Cushitic; Hbr. — Hebrew; Hrs. — Harsusi; Jib. — Jibbali (= Shaḥri); Jud. — Judaic Aramaic; LL = lexical lists; Leb. — Lebanese Arabic; LEC — Lowland East Cushitic; Mlt. — Maltese Arabic; Mec. — Meccan Arabic; Med. — Medical Texts; Mhr. — Mehri; MK — Middle Kingdom; Mnd. — Mandaic Aramaic; Mod. — Modern; MSA — Modern South Arabian; N. — North; NK — New Kingdom; OK — Old Kingdom; Omot. — Omotic; P. — Proto; Pal. — Palestinian Aramaic; pB. — postbiblical; Pho. — Phoenician; Pyr. — Pyramid Texts; Qur. — Qur'anic Arabic; S. — South; Sab. — Sabaic; Sel. — Selti; Sem. — Semitic; Sod. — Soddo; Soq. — Soqotri; Syr. — Syrian Aramaic; Tna. — Tigriñña (= Tigray); Tgr. — Tigre; Ugr. — Ugaritic; Urm. — Urmian Neo-Aramaic; W. — West; Wol. — Wolane.

Transcription and transliteration:

c — alveolar voiceless affricate [ts], *ȝ* — alveolar voiced affricate [dz], *č* — palato-alveolar voiceless affricate [tš], *ǯ* — palato-alveolar voiced affricate [dž], *ṣ* — hissing emphatic voiceless fricative, * c̣* — emphatic voiceless affricate, *ẓ* — conventionally stands for what was likely *ḍ*, emphatic voiced interdental, or *ṯ̣*, emphatic voiceless interdental, *č̣* — palato-alveolar emphatic affricate, *ŝ* — lateral voiceless fricative, *ĉ* — lateral voiceless affricate, *ĉ̣* — lateral voiceless emphatic affricate, *ẓ̂* — lateral voiced emphatic fricative or affricate, *ẑ* — lateral voiced fricative, *ḳ* or *q* — emphatic velar stop, *γ* — uvular voiced fricative (Arabic "ghain"), *ḫ* — uvular voiceless fricative, *ẖ* — uvular voiceless fricative (only in Egyptian), *ḥ* — pharyngeal voiceless fricative, *h* — laryngeal voiceless fricative, *ʕ* — pharyngeal voiced fricative, *ʔ* — glottal stop, *H* — unspecified laryngeal or pharyngeal, *y* — palatal resonant.

1 ALL:

(1) Akk. *kalû*; Ugr. *kl*; Hbr. *kōl*; Pho. *kl*; Bib. *kōl*; Pal. *kwl, kol*; Syr. *kul*; Mnd. *kul*; Urm. *kəl*; Qur. *kull-*; Leb. *kəll*; Mec. *kull*; Mlt. *kolla*; Sab. *kll*; Gez. *kʷǝllu*; Tna. *kʷǝllu*; Tgr. *kǝllu*; Amh. *hullu*; Gaf. *yǝlhʷä* (<**yǝlkʷ-*, met.); Sod. *kullǝm*; Cha. *ǝnnǝm*; Har. *kullu*; Wol. *hullǝm*; Hrs. *kal(l)*; Mhr. *kal*; Jib. *ka(h)l* // < Sem. *kʷall-u (cf. in LGz 281).

(2) Arg. *muli* // < Sem. *mlʔ 'to be full' (v. FULL No. 1).

(3) Soq. *faḥere* // < Sem. **paḫr-* 'totality, gathering': Mhr. Jib. *fáḫrǝh* 'together' (JM 110, JJ 67), Akk. *paḫāru* 'sich versammeln' (AHw 810), 'to assemble, congregate, gather, collect' (CAD *p* 23), Ugr. *pḫr* 'assembly, cluster; group, faction, family' (DUL 669), *pḫyr* 'whole, totality'

(DUL 670),[5] Pho. *m-pḫr-t* 'assembly', ESA: Qatabanian *ftḫr* (*-t*-stem) 'to enter into partnership, associate with' (Ricks Qat. 129).

→ **Proto-Semitic**[6] *k^wall-u* (#1) < Afras. *k^wal-* 'all, each, much': (?) Brb.: Ahaggar *tu-klə-t* 'ê. réuni en masse'; Egyp. (Pyr.) *ṯnw* 'each' (<*kVlw*?[7]); W. Chad.: Pero *kálù* 'to gather', C. Chad.: Gude *kálà* 'every'; C. Cush.: Waag *täkäl* 'all' (likely <*ta-kal*)[8]; S.: Iraqw *kila* 'very much, completely', Dahalo *ʔákkale* 'all'; S. Omot.: Dime *kull* id. (cf. EDE I 136).

2 ASHES:

(1) Akk. *tumru*; Ugr. *ʕmr* (also 'dust' // <*(tV-)ʕVmr-*.[9]

(2) Hbr. *ʔēpär*, Amh. *afär* // < Sem. *ʔapar-* 'dust, soil; ashes' (v. HALOT 80; LGz 10).

(3) Pho. *ʔry* (Tom. 29) // < Sem.: Hbr. *ʔūr* 'firelight, fire' (HALOT 25), Arb. *ʔry* II 'allumer le feu' (BK 1 27), *ʔirrat-* 'feu' (ibid. 22), *ʔawwara* 'enflammer' (ibid. 68), Tgr. *ʔarwa* 'to flame, to blaze' (LH 359).

(4) Pal. *ḳṭm*; Syr. *ḳeṭm-*; Mnd. *giṭm-*; Urm. *ḳiṭm-* // < Sem. (Arm-Arb.) *ḳiṭam-*: Arb. *ḳatām-* 'poussière' (BK 2 675).

(5) Qur. *ramād-*; Leb. *rməd*; Mec. *rumād*; Mlt. *ərmit* // In the absence of direct cognates,[10] one wonders whether it may be a metathesis < Sem. *midr-* 'dust, dirt': Hbr. pB *mädär* 'ordure (material used for vessels)' (Ja. 735); Syr. *medr-* 'gleba (terrae), terra, lutum, pulvis' (Brock. 373); Arb. *madar-* 'boue sèche et tenace, sans sable' (BK 2 1078), Gez. *mədr-* 'earth, ground, soil, etc.' (LGz 330), Mhr. *mdêr* 'Lehmziegel' (Jahn 210), v. LGz 330.

---

[5] Surprisingly overlooked in Kog. Ug. 466, wherein Akk. *paḫāru* is referred to as "the only reliable Sem. cognate to the MSA forms." In this respect, I would like to polemicize with my friend and co-author Leonid Kogan whom I consider one of the (if not the) best today's Semitists. This is an example of our long-term controversy about what he regards as "unreliable cognates" — in this case, implicitly, the Phoenician and Qatabanian forms that he does not even quote in the main text, but rather in a footnote. My position is that in such cases, one should strictly observe the "presumption of innocence". What is wrong with the two examples? Or with the fact that both of them represent *hapax legomena*? They do match the Akkadian form (let alone the Ugaritic and Soqotri ones) perfectly, both phonetically and semantically. Does this assertion cause doubts? Or are there doubts in the philological aspect —about their correct reading or interpretation? If there are, they should be explicitly exposed, otherwise they are invalid. Are there doubts as to their authenticity? If so, any suspicions about their having been borrowed and any suggestions about the source of borrowing should be openly discussed. Are there doubts about the qualifications of the author(s) of the corresponding source if he/she quotes the form in question without a question mark or any other sign of *his/her* doubts? Could he/she, for some whimsical reason, have forged the form in question, deliberately adjusting it to match the "reliable cognates"? Or can this affinity be the result of a chance coincidence? Perhaps there is some other *rational* justification that I have overlooked — besides the simple intuitive mistrust of the philologist, which is eventually of the same nature as the mutual mistrust that is often felt between students of "classical" literary ancient languages and those of non-literary modern living languages towards the data of each other. If not — what is the point of this self-restrictive overcautiousness?

[6] By "Proto-Semitic", which I use in a somewhat conventional opposition to "Common Semitic" (see n. 10) I refer to a term represented in all the main branches of Semitic, according to my genealogical classification based on lexicostatistics: South Sem. (MSA), North Sem. (Akkadian), and West Sem. (all the rest).

[7] Tentatively compared in Vyc. 216 with *ṯr* 'number' (since the 20th Dyn.) with the following comment: "The writing *ṯnw: ṯrw* speak in favor of reading as *ṯlw*".

[8] Cf. PNAgaw *t-aḫər/-əḫər* 'aunt', a fem. derivative with prefixed -*t* (App CDA 26).

[9] These two forms, undoubtedly related, are not compared either in AHw 1370 or in DUL 165 where the Ugr. term is viewed as having no definite etymology; direct, if tentative, comparisons (ibid.) with *ʕpr* 'polvo, tierra' and other Hbr. and Arb. terms are unacceptable unless viewed as instances of *m : p* root variation, which in this case, however, is hardly possible to prove or disprove (on this phenomenon v. Maizel and SED I pp. LX–LXIII).

[10] Note what can be viewed as a variant root: Arb. *rubd-at-* 'colour of ash, ashen'; cf. also Hausa *rúbúḍí* 'hot fine ash' (an Arabism?).

(6) Gez. *ḥamad*; Tna. *ḥamäk̲ʷəsti* or *ḥamäd k̲ʷəsti*;[11] Tgr. *ḥamäd*; Amh. *amäd* (syn.); Arg. *hamäd*; Sod. Cha. *amäd*; Har. *ḥamäd*; Wol. *amäd* // From the semantic point of view, more likely < Sem. **ḥmd* (in which case *ḥ-* in Gez. must be treated as a purely graphic variant of **ḫ-*): Arb. *ḫmd* 'cesser de flamber (se dit du feu, quand la flamme s'éteint, quoqu'il y ait encore des tisons qui brûlent)' (BK 1 630), Mhr. *ḫəmūd* 'to be extinguished, burnt out' (JM 443), Jib. *ḫõd* 'to extinguish; to be extinguished' (JJ 301); very probably also Hbr. pB. *ḥmd* 'to produce shrivelling by heat' (Ja. 475). Somewhat less likely < Sem. **ḥmd* 'to be hot' represented by Arb. *ḥmd* 'être intense (se dit de la chaleur)', *ḥamadat-* 'pétillement du feu qui brûle' (BK 1 488) and Daṯ. *ḥamad* 'to be hot, burn' (quoted in LGz 232 together with Arb. *ḫmd*).[12]

(7) Tgr. *rämäč̣* ('hot ashes', syn.);[13] Mhr. *ramẑ̇*; Jib. *rɛmč̣* // < Sem. **ramẑ̇-* 'hot ashes', **rmẑ̇* 'to burn' (v. LGz 470).

◊ Hrs. *remēd*, Mhr. *rmid*; Jib. *rīd* and Soq. *rimid* are rather Arabisms than original retentions; no terms for 'ashes' found in the available sources for Bib., Sab. and Gaf.

→ **Common Semitic 1**:[14] **(tV-)ʕVmr-* (#1). The only plausible, if isolated, parallel is W. Chad.: S. Bauchi **m/ŋuruŋ-* 'ashes' (cf. EDE III 244), probably <**muruH-* and consequently < **murʕ-*: Jum *mùrún*, Mangas *mwúrùn* and *ŋúruŋ*, Kir *mwùrŋ* and *ŋúrəŋ*.

Common Semitic 2: **ramẑ̇-* (#7).

Common Semitic debatable[15] (# 2) **ʔapar-* < Afras. **far-* 'dry soil': Egyp. (Pyr.) *fˀ.t* 'dust (?); W. Chad.: Hausa *fàrà̀*, C. Chad.: Gaanda *fìr-tà*, Boka *fur-tà* 'ground', E. Chad.: Mokilko *pùùré* 'dust'; E. Cush.: LEC: Dasenech *faara* 'clay' (ADB)[16].

3 BARK:

(1) Akk. *ḳuliptu, ḳulpu*; Hbr. *ḳəlippā*; Syr. *ḳəlāpət-*; Urm. *ḳalpa*; Hrs. *ḳelfēt*; Mhr. *ḳalifūt*; Soq. *ḳalifoh* // < Sem. **ḳal(i)p-* (v. in LGz 427).

(2) Mnd. *masik-* // < Sem. **ma/išk-* 'skin' (v. in SED I No. 190).

(3) Leb. *ʔišri*; Mec. *gišra*; Jib. *ḳaŝrot* // < Sem. (Arb.-Eth.) **ḳVŝr-*: Gez. *ḳaŝŝara* 'to peel, scrape, take off scales', *ḳəŝŝār* 'fish scales, shell' (LGz 448).

(4) Gez. *ləḥṣ*; Tna. *ləḥṣi*; Tgr. *ləḥaṣ*; Amh. *ləṭ*; Arg. *lihinṭo*; Har. *inči ləḥiṭ* ('thin bark of tree'); Wol. *ləč̣č̣ač̣e* // A deverbal noun < Common Eth. **lḥṣ* 'to peel, bark' (LGz 312), likely < Sem. **lḥṣ/ḫlṣ* 'to draw off, peel': Arb. *laḥḥaṣa* 'épurer en séparant les parties moins propres; enlever, tirer, extraire la partie la plus pure et la meilleure' (BK 2 980), Akk. *ḫalāṣu* 'to press, squeeze out; clean by combing' (CAD ḫ 40), Hbr. *ḥālaṣ* 'to draw off', pB. 'to take off (shoe)' (HAL 321), (nif.) 'to be peeled off (skin)' (Ja. 472), Jud. *ḥalaṣ* 'to take off, undress' (ibid.

---

[11] *k̲ʷəsti* is a variant stem of *k̲ʷäsk̲ʷäsä* 'to stir, poke, revive, relight a fire, to shake cinders, ash from a firebrand' (Kane T 966–7; cf. Bulakh Dis. 119–120).

[12] Cf. Kog. Eth. 379 ("None of the two alternative etymological approaches to this Proto-Ethiopian root outlined by Leslau is fully convincing"), where Hbr. pB. *ḥmd* 'to produce shrivelling by heat' and Akk. *ḫamadīru* 'shrivelled or withered' (CAD ḫ 57; the form has an affixed fossilized -r, v. Mil RE) are compared not to Arb. *ḫmd* 'to subside (of fire)', which is more attractive in view of Akk. *ḫ-*, but to Arb. *ḥmd* and *ḥamadat*, which implies an irregular, though not unattested correspondence: Akk. *ḫ* vs. Arb. *ḥ.*

[13] Borrowed into C. Cush. Aungi *areméc* 'embers' (App. CDA 61).

[14] What I conventionally call "Common Semitic" are cognate terms — provided they are definitely "above suspicion" of having been borrowed — represented at least in two of the three branches of Semitic (at least in one language of each branch).

[15] On such cases as Hbr. *ʔēpär*, Amh. *afär*, when a similar meaning evolution from a different meaning of the common proto-form seems quite transparent, see note 18.

[16] The Egyp. word (meaning debatable) is tentatively compared in EDE II 553 with Sem. **ʔapar-* ~ **ʕapar-* 'sand' (two different roots, comparable as variants) and Mokilko.

473). It is hard to say whether -ḥ- in the Gez. root is a graphic variant of *ẖ or reflects Sem. *ḥ; cf. what looks like two variant roots with ẖ vs. ḥ in Arb., both probably with the underlying meaning 'bark': lẖṣ 'av. la paupière de l'oeil supérieur très charnue' (BK 2 980) and laḥaṣ- 'contraction de la paupière supérieure, au point qu'il s'y forme des plis' (ibid. 974). Cf. Kog. Eth. 377.

(5) Tna. ḳʷərbät (syn.; also 'skin, rind, peel') // < Sem. (Arb.-Eth.) *ḳʷirb-at-: Arb. ḳirbat- 'grande outre à lait ou à eau faite d'une seule peau cousue au milieu' (BK 2 704), Gez. ḳʷərbābit 'leather bag' (LGz 440), Amh. ḳorbät 'skin'.

(6) Tgr. ḳärəf (syn.); Amh. ḳärfit (syn.); Sod. ḳərfit (syn.); Har. ḳärfit ('hard bark of tree') // < Sem. (Arb.-Eth.) *ḳVrp-: Arb. ḳirf- 'bark (n.)', ḳrf 'to peel' (v. LGz 441).

(7) Sod. kana, Cha. ḵara // only Gur.[17]

◊ Mlt. barka is a lw., likely < English; no terms in Ugr., Pho., Bib., Pal., Qur., Sab. and Gaf. Note: *ḳal(i)p-, *ḳVrp- and *ḳʷirb-at- are scored differently as they go back to three different variant roots as early as in Afras. For *ḳVrp- ~ *pVrḳ- cf. Brb.: E. Tawllemmet e-fărăɣ 'coquille'; W. Chad. *ḳʷarip-: Tsagu ḳorōpe, Barawa kworəp, Wangday kwòrip 'bark'; Egyp. (Med.) pꜣḳ-t 'shell (of turtle, skull)' (v. EDE II 403–4); and, perhaps, E. Cush.: Somali fuuruq 'smallpox' (met. and a meaning shift 'bark' > 'scab'); for *ḳʷirb-at-, C. Chad.: Mandara kwàlàbàa 'bark' (possibly < *ḳʷarab-), N. Cush.: Beja kurbe 'skin' (<*ḳurb-), Omot.: Male ḳurubi 'bark', etc. (ADB).

→ **Proto-Semitic:** ḳal(i)p- (#1) < Afras.*ḳalp-: E. Cush. *ḳolf-: Somali qolof, Konso qolfa 'bark', Oromo qolofa 'foreskin', Gawwada qoffol 'bark' (met.).

4 BELLY:

(1) Akk. karšu; Syr. kars-; Mnd. kars-; Urm. ki(r)s-; Tgr. käršät; Arg. kärs, hars; Gaf. ərsä, Sod. kärs; Har. kärsi; Hrs. kērəŝ; Mhr. kīrəŝ; Jib. s̃irŝ // < Sem. *kar(i)ŝ- (SED I No. 151).

(2) Ugr. kbd (?); Gez. kabd; Tna. käbdi; Tgr. käbəd (syn.), Amh. hod // < Sem. *kabid(-at)-,[18] v. in LIVER No. 2.

(3) Hbr. bäṭän; Qur. baṭn-; Leb. baṭan; Mec. baṭin // < Sem. *baṭn- (SED I No. 42).

(4) Bib. *məʕē (pl. suff. məʕốhī) // < Sem.: Hbr. mēʕayim (pl.) 'entrails', Arb. maʕy- 'intestins', etc. (SED I No. 185).[19]

(5) Wol. däl; Cha. dän // according to LGur 210, "represents däl" with the l ~ n variation; if, indeed, < *dal 'abdomen, belly, stomach, interior' (including Selti dälmūt 'intestine') ibid., these forms are related to Amh. (Gondar) dulät 'mets de tripes de chèvre ou de mouton' and Arb. dawlat- 'jabot, gésier' (DRS 235) going back to Sem. (Arb.-Eth.) *dawl- 'stomach, interior'. If, otherwise, the Gur. forms represent dän, they should be compared to redupli-

---

[17] Tentatively compared in LGur 344 to E. Cush. Burji kán-oo 'bark', which can hardly be a source of borrowing into Gur. One wonders whether Sidamo ḳonnonna id. could be such a source, with ḳ- rendered as k- in Gur.

[18] The treatment of such cases is a serious problem for lexicostatistics: on one hand, it seems obvious that the shift from 'liver' to 'belly' in Ugr. and Proto-Eth. should be estimated as two independent processes, not reflecting a common *inherited* feature; following this logics, the Ugr. and Eth. forms should be scored as unrelated which, however, would have looked strange. On the other hand, 'belly' could have been a secondary meaning of *kabid- as early as in Proto-West Sem., accounting for the later semantic shift in both Ugr. and Proto-Eth. caused by this inherited common feature and allowing to score them as related.

[19] There are isolated parallels worth mentioning: E. Cush.: LEC: Bussa mắyē 'liver', which, according to EDE III 160, may be a borrowing from N. Omot. *mayy-, regularly from *mayz- (corresponding to Egyp. myz.t 'liver'); cf., however, E. Chad.: Gadang mùyò 'liver' (derived by Jungraithmayr and Ibriszimow from Chad. *m-l-d, at first glance, rather suspicious). Could this stunning resemblance in root structure point to the vestiges of Afras. *maʕVy- 'entrails, liver'?

cated Gur. *dənäddänä* (LGur 212), Gez. *dandana* 'to be fat, stout', Amh. *dänäddänä* id. (LGz 136), probably further related to Akk. *dandann-* 'tout puissant' (compared in DRS 280; 'al-mighty' in CAD *d* 87) connected with *danānu* (*daʔānu*) 'strength, might, force' (CAD *d* 81) and/or Sem. **duhn-*, **daha/in-* 'fat' (SED I No. 48).

(6) Hrs. *hōfel*; Mhr. *hōfel*; Jib. *šofəl* (all syn.) // < Sem. **ŝₓV(n)pVl-*:[20] Arb. *mišfalat* 'gésier; esomac', Tgr. *šənfəlla* 'one of ruminant's four stomachs', etc. (SED I No. 271).

(7) Soq. *mer* (*mher*) // likely < Sem. **marʔ-* 'fat' (cf. LGur 418 and FAT No. 9); less likely < Sem. **mi/ar(V)r-(at-)* 'gall, gall-bladder' (SED I No. 188).

(8) Soq. *ḥant* (syn.) // with the assimilation of **-m-* > *-n-* to the dental *-t* (< **ṯ*) in a contact posi-tion < Sem. **ḥVmṯ-* '(lower) belly, uterus, womb' (SED I No. 122).

◊ Mlt. *stonku* is a lw. from a European language (Italian or English?); no term in Ugr., Pho., Pal., Sab.

→ **Proto-Semitic:** **kar(i)ŝ-* (#1).

Common West Semitic **baṭn-* (#3) < Afras. **ba/uṭ(Vn)-*: Brb.: Semlal *a-buḍ* 'navel', Ntifa *a-buḍ* 'belly' (and *i-biniḍ* 'navel', met. <**biḍin-*?), etc.; W. Chad.: Mupun *a-buḍ*, Angas *ḅwut*, Fyer *ḅúto'*, etc. 'belly, stomach' (see more details in EDE II 241–2).

Common West Semitic (debatable) (#2): **kabid-*

5 BIG:

(1) Akk. *rabû*; Ugr. *rb*, *rabbu*; Pho. *rb*; Bib. *rab*; Pal. *rb*; Syr. *rabb-*; Mnd. *rba* // < Sem. **rabb-* (DLU 382–3).

(2) Hbr. *gādōl* // < Sem. **gVdVl-* (HAL 177; 179); Egyp. (MK) *ḏdʒ* 'fat' (adj.) if < **gdl* (v. EDE I 245) is related, going back to Afras. **gVdVl-* 'big, fat'.

(3) Bib. *ŝaggī(ʔ)* (syn.) // Aramaic only; the interpretation as 'big' is debatable.

(4) Qur. *kabīr-*; Leb. *kbeyr*; Mec. *kabiyr*; Mlt. *kbīr* // < Sem.: Akk. *kabāru* 'to become fat, heavy, thick, strong' (CAD *k* 4), Syr. *kbr* 'multus fuit' (Brock. 316), Sab. *kbr* 'great; richness, abun-dance' (SD 76), etc.

(5) Gez. *ʕabiy*; Tna. *ʕabiy*; Tgr. *ʕabi* // < Sem. **ʕ/ɣby* 'to be big, thick' (LGz 55).

(6) Amh. *təlləḳ* (< *tə-lləḳ*); Sod. *maläḳ*; Cha. *näḳ* // < Eth. **lhḳ* 'to grow, grow up' (LGz 309) < Sem., if Lelsau's interpretation of Soq. *di-leḳ* as 'which is numerous' (LS 129) is correct.

(7) Arg. *läham*, *näham* // < Sem. **lVhVm-*: Akk. *lim*, nom. *līmu* 'one thousand' (CAD *l* 194), Arb. *lahmūm-* 'grand nombre' (BK 2 1034); cf. also W. Chad. Hausa *lùmùmú* 'in quantity' (Barg. 732), E. Cush. Darasa *lumo* 'big' (Huds. 27) < Afras. **lV(H)m-* 'big quantity'?

(8) Gaf. *əmmuna*; Cha. *əmmiyä* (syn.); Jib. *ʔum*, Soq. *ʔä́m* (fem.) // likely < Sem. **ʔu/imm-* 'mother' (v. in LGz 22; cf. also LGur 49–50).[21]

---

[20] On *ŝₓ* v. SED I XLVIII–CV. The decision to separate this root (**ŝₓV(n)pVl-*) from **špl* 'to be low' was taken by the SED authors after a lot of discussion and hesitation; the fact that the two roots are usually represented as one is *not* what I call "mythetymology" (where the blunder usually lies on the surface, due to lack of professionalism in etymological technique, inertia, old stereotypes, overreverence toward one's scholarly ancestors, or sloth of mind) — this case is really very complicated, with the difference in consonantism being fairly subtle and very likely in-volving traces of contamination. The fact that this entangled situation keeps triggering fancy ideas is evidenced by the following comparison in EDE I 324 (note 11): "OEg. **sfʒ* [**sfl*] → MEg. *sfʒ* "to hate" = Soq. *špl* "to despise" < Sem. **špl* "to be low." One wonders how a word in one language can be equated with a semantically compatible word in another language, whose meaning ("despise") is, however, openly recognized to be secondary and derived from quite a different meaning ("be low", which is quite tenable)?

[21] For the semantic shift, cf. 'big' < 'father' below (#10). This case is similar to the one discussed in n. 18: it is hard to decide whether the shifts 'mother' > 'big' (in MSA, for nouns in the fem. gender and/or objects associated with the feminine as opposed to the 'father' > 'big' shift for nouns in the masc. gender and/or objects associated

(9) Sod. *gəddər* (syn.); Har. *g(i)dīr*, Wol. *gädärä* // in Wol. and Zway *gädärä* is 'to grow up (child), be big' compared in LGur 264 (with hesitation, but quite reasonably) to Amh. *(tä)gäddärä* 'to germinate' ("that is, grow" ibid.), further related to Arb. *ǯdr* 's'élever au-dessus du sol (se dit des plantes); se former (se dit des certain fruits)' (BK 1 263) < Eth.-Arb. **gdr* 'to grow, grow big'.[22]

(10) Hrs. *ŝoḫ*; Mhr. *ŝoḫ* (also 'old') // < Sem. **ŝyḫ-* 'to grow big or old': Akk. *šâḫu* (*šiāḫum*) 'to grow (in size or age)' (CAD š1 106), *šīḫu* 'tall, high, stately' (ibid. š2 418), Arb. *šayḫ-* 'vieillard; ancien, cheikh; maître' (BK 2 1296).

(11) Hrs *nyōb*, Mhr. *nōb* (fem.) // < Sem.: Arb. *nāb-*, pl. *ʔanyāb-* 'chief of a tribe', *nawb-* 'power' < Afras.: Egyp. (OK) *nb* 'lord, master' (< **nVb*, cf. Vyc. 138–9); E. Cush.: Afar *nabba* 'big' (see EDE I 107).

(12) Jib. *ʔeb* (syn.); Soq. *ʔeb, heb* (syn. 1) // likely < Sem. **ʔab-* 'father' (v. in DLU 2; LGz 2).

(13) Soq. *ʕeḳar* (syn. 2) // < *ʕéḳar* 'grandir' (LS 325) < Sem. (Arb.-MSA): Jib. *ʕɔ́ḳɔ́r* 'to grow up' (JJ 11), Arb. *ʕḳr* 'ê. grand, haut, d'une belle croissance (se dit des plantes)' (BK 2 315).

◊   Urm. *gūr* < Kurdish *gaur, gûr*.

→   **Common North and West Semitic**: **rabb-* (#1); cf. S. Omot.: Ongota *arba* 'big'.

   Common Semitic debatable (#8): **ʔu/imm-* 'big', i.e. 'mother' < Afras. **ʔVma* 'mother' represented in Sem., Chad. and Cush. (ADB).

6 BIRD:

(1) Akk. *iṣṣuru*;[23] Ugr. *ʕṣr, ʕuṣṣūru* (Huehn.) // < Sem. (Akk.-Ugr.) **ʕVṣṣūr-*.

(2) Hbr. *ṣippōr*; Pho. *ṣpr*; Bib. *ṣippar*; Pal. *ṣypr*; Syr. *ṣeppar-*; Mnd. *ṣipr-*; Urm. *sipr-*, Mlt. *(gh)asfūr* (< Arb. *ʕaṣfūr-*, with a secondary ʕ- perhaps influenced by Sem. **ʕVṣṣūr-*, or even a remnant of a composite form) // < Sem. **ṣVp(p)Vr-*.

(3) Syr. (syn. 1), Urm. (syn.) *ṭayr-*; Qur. *ṭayr-*; Leb. *ṭayr*; Mec. *ṭayr* // < Sem. **ṭayr-* (SED II No. 235).

(4) Syr. *pāraḫ-t-* (syn. 2) // < Sem. **parḫ-* 'chick, brood' (SED II No. 179).

(5) Gez. *ʕof*; Tna. *ʕuf*; Tgr. *ʕuf*; Amh. *wof*; Arg. *of, wof*; Gaf. *yəfʷä*; Sod. *wof, of*; Cha. *afʷ*; Har. *ūf*; Wol. *ũfʷ* // < Sem. *ʕawp-* 'bird' (SED II No. 48), related to **ʕwp* 'to fly', both < Afras. **ʕa(w)p-* 'bird; flying': Egyp. (late) *ʕpy* 'to fly'; S. Omot.: Ari *ʔafti, apti*, Dime *iftu*, Hamer *ap/fti* 'bird' (a generic term) < **ʔap-t-i* < **ʕap-* (ADB).[24]

---

with the masculine) took place independently in S. Eth. and MSA or the "potential" for this shift had already been there in the corresponding terms in Proto-Sem. — and the mentality of its speakers.

[22] Presumably, with fossilized suffixed **-r* < Sem **ga/idd-*: Arb. *ǯidd-* 'beaucoup, extrêmement' (ibid. 260), Sab. *gdd* 'great' (SD 49), Tgr. *gäddä* 'to be bigger, surpass' (LH 602; unless an Arabism) < Afras. **gVd(d)-*: Brb. C. Morocco *gudy* 'ê. nombreux, beaucoup, abonder', *sgudy* 'produire beaucoup, en grande quantité' (DRB 737–8 without specifying the language; cf. Ahaggar *egdeh*, Ayr *egdu* 'suffire' ibid. 727), W. Chad. Bolewa *gòdo* 'many' (Kr. I 87), N. Cush. Beja *gwud* 'many', E. Cush. Arbore *guudá* 'many', Dasenech *guddu* 'big' (Bla. Om. No. 5.2), Oromo *guddaa* 'big; greatly, very' (Gr. 184), S. Omot. Dime *gɛɛd* 'big' (Bnd Om. 205), Ongota *gadaḫ/hune, gaddahino* (Fl. Ong. 42), *gaddaʕuni*, pl. *giddeʕeta* 'big, old' (S-T 117). V. in Mil. RE.

[23] Certainly not < **ʕiṣpur-*, proposed by some Semitists and uncritically repeated by others — a typical example of what can be described by the oxymoron "scholarly folk etymology", by me called "mythetymology". See SED II LIV–LV for more details on this.

[24] Cf. also EDE I 67, where the S. Omot. forms are compared with Egyp. *ǯpd* 'bird', implying an irregular — and non-existent — sound correspondence Egyp. *d* ~ Omot. *t* (the note on the Omot. forms "assim. < **ʕVpd-*" is of no help, since no such process is attested in S. Omot. — otherwise it should have been demonstrated). Such forced "disposable" correspondences, "valid" only for one example (they occur in hundreds in Semitic and in thousands in Afrasian studies), are an insult to the comparative method — especially when they are proposed by one of the very few really professional adherents of this method in Afrasian linguistics.

(6) Tgr. *särerät* // < Sem. *šrr 'to fly, jump' (LGz 514).

(7) Hrs. *ʔaḵāb*; Mhr. *ʔaḵabit* (the other term for 'bird', *ṭeyrīt*, must be an Arabism); Jib. *ʕes̃yet* // < Sem. (Arb.-MSA, less likely, an Arabism in MSA because of difference in meaning): Arb. *ʕaḵāb-* 'eagle' (BK 2 310).

(8) Soq. *noyhir* (another term, *ʔəsféroh*, is more likely an Arabism) // < Sem. **nVšr-* 'eagle, vulture' (SED II No. 166).

◊ No term in Sab.

→ **Common Semitic**: **ʕVṣṣūr-* (#1), met. < Afras. **çirāʕ-*: S. Cush. Iraqw *çirʕi*, Alagwa *ciraʕa*, Burunge *čiraʕa*, Asa *širaʔa* 'bird'; E. Cush. Saho *čaráʕ, čarráʕ* 'Madenhacker, buphaga erythrorchynchus', etc. (v. in SED II No. 43).

  Common West Semitic 1: **ṣVp(p)Vr-* (#2) < Afras. **çi/apur-*: W. Chad. Mburku *čápùr*, C. Chad. Bura *càvuř*, Margi *càvuř* 'guinea fowl', Kilba *cívìr*, Hildi *civ̇rəw*, Wamdiu *civùř*, Mofu *càvár* id. (v. in SED II No. 212).

  Common West Semitic 2: **ṭayr-* (#3).[25]

7 BITE:

(1) Akk. *našāku*; Ugr. *ntk*; Hbr. *nšk*; Gez. *nsk*; Hrs. *neṯōk*; Mhr. *nəṯk* // < Sem. *nṯk (v. in LGz 402).

(2) Pal. Syr. Mnd. *nkt*; Tna. *näḵäsä*; Tgr. *näkša*; Amh. Arg. Gaf. *näkkäsä*; Sod. *näkkäsäm*; Cha. *näkäsäm*; Har. *näḵäsä*; Wol. *näkäsä* // < Sem. *nkṯ (cf. LGz 402).[26]

(3) Urm. *ḳrṭ* // < Sem. *ḳʷrṭ 'to cut, pinch': Arb. *ḳrṭ* 'to cut in pieces', Tna. *ḵʷärṭäṭä* 'to pinch, break off leaves', etc. (v. in LGz 444), further related with a fossilized -*m* suffix to Arb. *ḳrṭm* 'couper', Gez. *ḳarṭama* 'to munch, chew food that is hard', Soq. *ḳarṭem* 'to chew', etc. (LGz 445). Formally is also compatible with Arb. *ḳrẓ* 'couper' (BK 2 716).

(4) Qur. Leb. Mec. *ʕḍḍ* // < Sem. **ʕṣ̂ṣ̂*: Arb. *ʕḍḍ* 'mordre; ê. rusé, astucieux', *ʕiḍḍ* 'méchant, qui mord; homme d'un mauvais caractère' (BK 2 276), Gez. *ʕaḍḍa* 'to deprive, cause harm, affront, do wrong' (LGz 58), Soq. *ʕéḍ(ḍ)* 'traître durement' (LS 323). There are isolated Afras. parallels: W. Chad. Hausa *gàçā* (possibly < **ʕaç̂-*), N. Omot. Dizi *wâç*. Probably related is HEC **Hi(n)ç-* 'to chew': Darasa *inç-*, Hadiya *īçç-*, Kambatta *īṭ-*, Sidamo *hinç-* (Huds. 413).[27]

---

[25] The only Afras. parallel found so far is in N. Omot.: Manjo *ṭōro* 'vulture' (H. Fleming. Kefa (Gonga) Languages, *The Non-Semitic Languages of Ethiopia*, Mon. No. 5).

[26] After some hesitation, scored differently from *nṯk. I suggest two main criteria to allow variant roots, or root variants, to be scored as different lexemes in a lexicostatistical study (in a "normal" etymological entry it suffices to just describe the controversy without taking any dramatic decisions): (1) if the variant roots in question occur in the same language; (2) different sets of cognates in related languages (for which their origin should be traced to the deepest chronological/taxonomic level possible). It is according to these criteria that the difficult decision on the *nṯk/ *nkṯ case was made (counter to Kog. Eth. 373 averting: "the metathetic variation, well attested for this root within and outside Ethiopian, is intriguing, but can hardly be regarded as an obstacle for postulating an eventual etymological identity of both variants"). According to criterion (1), there are two cases where both roots co-exist in the same language, one being Gez. *nasaka* 'to bite', *ma-nsak* 'jaw, teeth' (ibid.) and *nakasa* 'to bite', marked in LGz 398 as an Amharism, but having a few derived forms including *ma-nkas* 'jaw, jawbone'; the other, semantically less reliable, Syr. *nəkat* 'momordit; offendit iram' (Brock. 430) and *nətak* 'damno affecit' (ibid. 452). Application of criterion (2) is not so simple, since, while *nkṯ has quite reliable matches in non-Semitic Afrasian, the parallels to *nṯk unearthed so far are much less convincing.

[27] Quite likely, derivable from Afras. **ʕaç̂(ʕaç)-* 'facial bone, lower cheekbone' (see **10 BONE** #3); the idea (in EDE II 574) that Arb. *ʕḍḍ* 'to bite' is related to Gez. *ʕəḍe* 'vermin, worm, moth, caterpillar', Tna. *ʕəṣe* 'larva' (sustained by a similar connection between Aram. *tōlēʕā* 'worm' vs. *mətalleʕōt* (pl.) 'teeth', but what is meant is perhaps Hbr. 'jaw-bones', v. SED I No. 177) implies some sort of association between 'worm, larva' and 'tooth' and, to me, looks funny in view of the fairly deep knowledge of animal anatomy by the ancient Semites, clearly reflected in their anatomic lexicon (SED I).

(5) Mlt. *gidem* // Either < **gdm* 'to cut' (v. in LGz 182) or <**kdm* (*g-* < **k-* by assimilation with **-d*) < Arb *kdm* 'mordre' (BK 2 875); I could find no other parallels in Sem.

(6) Jib. *ĉaʕar* // Perhaps metathetically related to Arb. *ʕirḍ-* 'nuée de sauterelles' [BK 2 220] (and *ʕrḍ* 'crever pour avoir mangé trop d'herbe (se dit des moutons)') and Tgr. *ʕarṣat, ʕarṣätit* (also *ʕarṣetet*) 'termites' [LH 463], going back to Sem. **ʕrŝ-* ~ **ṣʕr* 'to devour'.[28]

(7) Soq. *ḳárḍeb* // Related to Arb. *ḳrḍb* 'couper, consumer tout, manger' (BK 2 714).

◊ No terms in Pho., Bib. and Sab.

→ **Proto-Semitic**: **nṯk* (#1) < Afras. **(nV-)čVk-*: C. Chad.: Mofu *-čáčək-* 'goûter', Mada *áččaka* 'goûter', etc.; (?) Egyp. NK *ḥsk* 'essen von etw.' (EG III 169; if <**ḥ-čk* with a hypothetic verbal prefix **ḥ-)*.

Proto-West Semitic: **nkṯ* (#2) < Afras. **(nV-)kuč-* or **(nV-)kʷVč-* 'tooth, biting' (or 'a biting tooth'): Egyp. (MK) *ṯs* (<**kVs*) 'tooth'; Brb.: Ntifa *uḳs*, Zenaga *ukši* 'tooth', Ahaggar *əkš* 'manger, mordre'; C. Chad. Malgwa *kúča* 'to bite off';[29] N. Cush.: Beja *kōs* 'tooth, horn', E. Cush.: HEC: Sidamo *kis-* 'to bite', S. Cush.: Qwadza *koʔos-iko* 'molar tooth'.[30] (ADB; cf. also EDE I 239).

8 BLACK:

(1) Akk. *ṣalmu*; Sab. *ẓlm* (SD 172; debatable, v. discussion in Bulakh Dis.); Gez. *ṣallim*; Tna. *ṣällim*; Tgr. *ṣällim*; Gaf. *ṣälläma*; Harari *ṭäy*; Wol. *ṭem* // < Sem. **ẓlm* 'to be black' (v. in LGz 556; Bulakh 2003 5–6 and Bulakh Dis.).

(2) Hbr. *šāḥōr* // < Sem. **šḥr* 'to be black' (HAL 1465, 1466, 1457; Bulakh 2003 13–14).

(3) Pal. *ʔwkm, ʔkwm*; Syr. *ʔukkām-*; Mnd. *ʕkum-*; Urm. *kūm* // < Sem. (compared in Bulakh Dis.): Akk. *akāmu* 'cloud of dust, mist' (CAD *a*1 259), Hbr. pB. *ʔkm* 'to be sun-burnt, black, dark-colored' (Ja. 64) < Afras. **kVm-*: Egyp. (Pyr.) *km* 'black'; E. Cush.: Dullay: Gawwada *kummay*, Harso *kúmma*, Tsamay *guma* 'black', etc., Yaaku *kumpuʔ* id.[31]

(4) Syr. *ḳənāʔ* (syn.) // Akk. (from OB) *uḳnû* 'Lapislazuli, Lasurstein, Türkis; (grün)-blau; künstliche Lapislazuli, blaue Glasur' (AHw. 1426f.), Ugaritic *iḳnu* 1) "gem of lapis lazuli"; 2) "violet blue"; 3) "violet purple or violet textile" (DUL 93), (?) Pho. *ʔḳnʔ* (lapis lazuli/purple; Phoenician blue/purple?) (HJ 100), Arb. *ḳunuww-* 'couleur noir', *ḳān-in* 'très-rouge' (BK II 826), *ḳanaʔa* 'être rouge, être teint en rouge (se dit de la barbe teinte en rouge, des doigts teints en rouge ou rougís du suc des mûres', *taḳnīʔ-* 'teindre en rouge foncé (les doigts, la barbe); teindre en noir (la barbe)', *ʔaḳnaʔ-* 'rouge' (там же, 818). Cf. AA **ḳVn-* 'to (be) white, yellow' (ADB).

---

[28] A tentative parallel suggested in Kog. Ug. note 51 is Gez. *ṣaʕara* 'to cause pain, torment, vex, etc.', with cognates in other Ethiopian; this seems erroneous not only because of Gez. *ṣ* instead of the expected *ḍ* — that might be accounted for by the scribe confounding the two graphemes which happens in Geez texts — but mainly because of reliable Aramaic matches with *ṣ* instead of the expected *ʕ*, corresponding to Jib. *ĉ* (the voiceless emphatic lateral affricate pronounced by several of my Jibbali-speaking informants; rendered by Johnstone as *ẑ*; anyway, <**Sem. **ŝ*), quoted in LGz 544; all of these forms are probably related to the Common Sem. verb **ṣʕr* ~ **ṣɣr* 'to be small' with a meaning shift 'to be small' > 'to be despised, neglected, treated badly' > 'to torment, vex, etc.' (cf. HALOT 1043).

[29] Cf. also W. Chad.: Paʔa *kačì* 'to insult'; probably also related are W. Chad.: Buli *ngàs-*, Zaar *ŋgas*, C. Chad. Daba *ŋàč*, etc. 'to bite' (CLR II 24–5), which, according to Stolb. 2005 No. 445, may go back to **nkač-*, with voicing of the velar consonant.

[30] Cf. also the enigmatic Bilin (C. Cush.) form *näkät-*, the main term for 'bite' (besides Qemant *näkäs*, a regular-looking Ethiopism), with *-t* instead of *-s*, expected both in an Ethiopic loan and in an inherited term < **nkč* (cf. App. CDA 33).

[31] Likely also C. Chad. Buduma *kaimē* 'Schatten (eines Menschen)' (LBud. 108) and, perhaps, C. Cush. Aungi *kem* 'farsi sera', Qwara *kŭm* 'giungere a sera', N. Omot. Kullo *kamma* 'notte' (CR Aw 164).

(5) Qur. *ʔaswad-*; Leb. *ʔaswəd*; Mec. *ʔaswad*; Mlt. *ʔiswet* // Obviously comparable with Mhr. *sátwəd* 'to be disgraced, blackened' (JM 353), Jib. *essōd* 'to blacken, curse', *estēd* 'to turn black, be disgraced' (JJ 232); however, lack of a direct meaning 'black as color' everywhere outside Arb. makes one suspect these forms to be metaphoric loans from Arb. (cf. *swd* III 'parler bas à l'oreille de quelqu'un' and the expression *sawwada llāhu waʒhahu* 'qui Dieu rende son visage noir!' pour dire, 'que Dieu le damne!' BK 1 1161[32]). Cf. discussion in HALOT 1417 and especially 1418 (in connection with Arb. *ʔaswad-*) about such demon names as Akk. *šēdu*, Hbr. **šēd* etc., including Mnd. *šdum* (with *-m* suffixed?) 'a spirit of the darkness, one of those ruling the underworld'. Cf., finally, Akk. *sēdu(m)*, attested in a lexical list and tentatively translated in AHw. 1034 as 'rot' (CAD *s* 206 gives no meaning). Outside Sem. there is a possible parallel in Chad. **sVdH-*: C. Chad. Lame-Peve Mesme *soḍ* 'dirt', Zime-Batna *suḍo*, Masa *súdoy* 'faeces' (CLR II 129), E. Chad. Kera *sòḍì* 'Dreckigkeit' (Eb. 108), Mokilko *sììḍo* 'earth (soil)' (CLR II 117).

(6) Amh. *ṭək*ʷər*; Arg. Sod. Cha. *ṭəkur*;[33] Gaf. *ṭəḳurä* (syn.) // Eth. **ṭḳr* 'to be black', **ṭaḳar* 'soot' (LGz 596). The only Sem. parallel, problematic both phonetically and semantically, that can be tentatively suggested is the metathetic Sem. **ḳutr-* 'smoke, incense' (see LGz 452 and ADB).

(7) Hrs. *ḥéwer*; Mhr. *ḥōwər* (*ḥɔr*); Jib. *ḥɔ́r*; Soq. *ḥohar, ḥaur* // < Sem. **ḥwr* 'to be black and white': Hbr. *ḥwr* 'to grow pale', Syr. *ḥewwār-*, Mnd. *hiwar-* 'white', Arb. *ḥwr-* 'ê. d'un noir et d'un blanc bien prononcé' (BK 1 509) (cf. Bulakh 2004 273–4).[34]

◊   No terms in Ugr., Pho. and Bib.

→   **Common North and West Semitic**: **ṯlm* (#1) < Afras. **čilam-* "to be dark, black": W. Chad.: Karekare *čàlúm* 'shade, shadow', C. Chad.: Bura *cilim* 'black soil used as a dye-stuff', Buduma *čilim* 'dark', Makari *silim* 'black', etc. (claimed by some Chadicists to be a Kanuri loan, which is out of the question in the light of Afras. data), E. Chad.: Mawa *čilim* 'black, dark'; S. Cush.: Qwadza *calam-* 'green'; S. Omot.: Ari *čelmi* 'to be dark' (ADB).

9 BLOOD:

(1) Akk. *dāmu*; Ugr. *dm*; Hbr. *dām*; Pho. *dm*; Pal. *ʔădam, ʔydm* (<**ʔa-dam-*, with **ʔa-* prefixed); Syr. *dəm-*; Urm. *dim-*; Qur. *dam-*; Leb. Mec. *damm*; Mlt. *dem*; Sab. *dm*; Gez. Tna. Tgr. Amh. Arg. Sod. Cha. Har. Wol. *däm*; Gaf. *däm*ʷä* // < Sem. **dam-* (SED I No. 50).

(2) Mnd. *zma*[35] // < Sem. **zam-*: Arb. *zaʔama* 'presser une plaie de manière que le pus en sorte, le sang se dessèche et forme une croûte' (BK 1 967), Gez. *zam* 'blood' (LGz 638) < Afras. **ʒam(ʔ)-* 'blood': W. Chad. Galambu *ʒàamá* (*ʒ-* <**ʒy-*), Sha, Kulere *zòm* (cf. Stolb. 1987 190), S. Omot. Ari Hamar *zumʔ-i*, Dime *ʒum-u* (Bnd Om. 206), cf. SED I No. 296.[36]

---

[32] M. Bulakh regards the possibility of borrowing into MSA as "undoubtful" (Bulakh Dis.).

[33] The other term for 'black', *gämbäna*, is from HEC, cf. Qabenna *gamballa*, Tembaro *gämbälla* (LGur 281).

[34] Possibly matching Egyp. (OK) *ḫȝ.ty* 'Bleicher, Wäscher' (unless <* *ḥVl-*), v. EDE I 149.

[35] This word's identification as a strange phonetic variant of **dam-* (also reflected in Mnd. as the less common form *dma*) by practically all the authors is one more Semitic "mythetymology".

[36] Not to be confounded with another Afras. root, **ʒ/ʒVn-* 'blood': Egyp. Pyr. *znf* (presumably *zn-f* "his blood"), Brb. Ahaggar *a-hni* (<**-ʒ/ʒVni*), Ayr *a-zni*, etc., W. Chad. Hausa *ʒìní* (<**ʒ/ʒini*); N. Omot. Zaysse *zonn-e* 'pus' (Hay Om 265; for the semantic shift, cf. Sem.: Mhr. *dəm*, Jib. *dihm* 'pus' JM 71 < **dam-* 'blood', v. #1). The variant roots **ʒam(ʔ)-* and **ʒ/ʒVn-* must have existed as *different* roots (*contra* EDE I 183 and 289) as early as in Proto-Afrasian and must be separated as such (with cross-references, of course), although eventually they appear to be related — one "simple" root and one with fossilized suffixal **-b* (this segment is frequently encountered in quite a few anatomic and non-anatomic terms: see Mil. RE): C. Chad.: Bachama *zambe*, Bata *ʒambe* <**ʒam/nb/p-* 'blood'; S. Omot.: Hamar *zumbi, zömḅi*, Karo *zunḅi* 'animal blood'.

(3) Hrs. *ḏōreʔ*; Mhr. *ḏōr-əh*; Jib. *ḏohr*; Soq. *dör* // Generally regarded as derived from MSA **ḏVrr-*: Mehri *ḏər* 'to spread out; to spread (gravy, curry, seed)' (JM 47), Jibbali *ḏerr* 'to spread out' (JJ 47) < Sem. **ḏrr/y/w/ʔ* 'to scatter, spread (seed), disperse, winnow'[37] (cf. HA-LOT 280; LGz 644; Mil. Farm.): Akkadian *zarû* 'to sow seed; broadcast; scatter, sprinkle; winnow', Hebrew *zry* 'to scatter, winnow', Judaic Aramaic *dry, drʔ* 'to scatter, strew; winnow', Arabic *ḏry/w* 'vanner, nettoyer (le grain)' (BK 1 771).

◊ No term in Bib.

→ **Common North and West Semitic**: **dam-* (#1) < Afras. **dam-* id. (ADB; EDE I 240).

10 BONE:

(1) Akk. *eṣemtu*; Ugr. *ʕẓm*; Hbr. *ʕäṣäm*; Pho. *ʕṣm*; Qur. *ʕaẓm-*; Leb. *ʕaẓam*; Mec. *ʕaẓum*; Mlt. *(gh)adma*; Gez. *ʕaṣam*; Tna. *ʕaṣmi*; Tgr. *ʕäč̣äm*; Amh. *ʔaṭənt*; Arg. *haṭəm, aṭənt*; Gaf. *aṣmʷä*; Sod. Cha. Wol. *aṭəm*; Har. *āṭ* // < Sem. **ʕaṭm(-at)-* (SED I No. 25).

(2) Bib. *gəram*; Pal. *grm*; Syr. Urm. *garm-*; Mnd. *girm-* // < Sem. **gVrm-* 'body; bone': Hbr. *gäräm* 'bone', Arb. *ǯirm-*, etc. 'corps', Sab. *grm* 'body (of animal)' (SED I No. 94).

(3) Hrs. *ʔāẑayẑ*; Mehri *ʕaẑayẑ*; Jib. *ʕayĉeĉ* // < Arb.-MSA **ʕVẓā/īṣ̂-* (v. SED I No. 24): Soq. *ʕêḍ* 'noyau (substance)'[38] (LS 323), Arb. *ʕuḍāḍ-, ʕaḍḍāḍ-* 'le haut du nez' (BK 2 277), 'os, carti-lage' (Belot 501) < Afras. **ʕaĉ(ʕaĉ)-* 'a facial bone': E. Cush. **ʕaḍ-*: Afar *óḍe*, Konso *aḍa*, Gol-lango *ʕaḍo* 'cheek', Arbore *ʔaĉéĉ* 'lower jaw', S. Cush. Iraqw Gorowa Alagwa Burungue *ʕūnĉa* 'cheek' (cf. K-M 309), Maʕa *iʔóŝo* 'cheekbone'.[39]

(4) Soq. *ṣéḥloh* // The comparison (made with reservations) to Soq. *ḍalḥ* 'côté' in LS 347 (< Sem. **ṣ̂il(a)ʕ-* 'rib, side (of chest)', v. SED I No. 272) is possible only if the two forms in Soq. are to be treated as variant roots; the comparison with metathetic Sem. **ḥVlṣ-* 'loin, hip' with the same root consonants (Hbr. *ḥălāṣayim*, Gez. *ḥəlṣ* 'loin', etc., v. ibid. No. 118) seems more attractive.[40]

◊ No term in Sab.

→ **Common North and West Semitic**: **ʕaṭm(-at)-* (#1). No Afras. parallels that I could find.

11 BREAST:

(1) Akk. *irtu*; Ugr. *ʔirt* // < Sem. **ʔir(r)-at-* (rather 'chest' than 'breast', cf. Tgr. *ʔarra* 'milt, by-stomach (of cattle)' SED I No. 9; cf., with metathesis, Sem. **riʔ-at-* 'lung' with Afras. paral-lels and Afras. **warVy-* 'lungs' ADB).[41]

(2) Ugr. *ṯd*; Hbr. *šōd*; Pal. *td*; Syr. *təd-*; Hrs. *ṯōdi*; Mhr. *ṯodi*; Jib. *ṯodɛʔ*; Soq. *todi* // < Sem. **ṯVdy-* '(woman's) breast' (SED I No. 280).

(3) Bib. *ḥădē*; Syr. *ḥady-* (syn.); Mnd. *hady* // < Hbr.-Arb. **ḥaḏ(V)y-* 'breast' (with plausible wider Sem. connections, v. SED I No. 112).

---

[37] The meaning shift seems uncommon unless we suppose an intermediate stage: 'to spread out' > '*to (let) flow' > 'to bleed/blood'. Cf. the shift from 'to flow' to 'blood' in Arb. *drr* IV 'laisser couler en abondance', *dirrat-* 'abondance (de lait, de la pluie)' and 'sang' (BK 1 681–2). Cf. verbal forms of the same root as 'blood' in MSA: Mhr. *ḏátri* (-*t*- stem) '(blood) to flow' (JM 81), Jib. *eḏrέʔ* 'to let an animal blood run over an invalid' (JJ 47), the latter verb pointing to a magic ritual which may account for the semantic evolution 'to flow' > 'blood'.

[38] For the semantic development cf. Russian *косточка* 'fruit-stone', literally 'little bone'.

[39] Cf. EDE I 299, comparing the Cush. forms with Arb., but not MSA, and tentatively with Egyp. *ḥḏ-wy* 'Kinnbacken', comparable only as a variant root, since Egyp. *ḥ* in no way corresponds to Afras. **ʕ*.

[40] Alternatively cf., with metathesis, Arb. *ḫlṣ* 'to be fractured (bone)' and *ḫaṣīl-* 'tail'.

[41] The interpretation of Akk. *irtu* as a reflexation of Sem. **ḥaḏ(a)y-* 'breast' proposed by some authors (e. g. Holma) is but another case of "mythetymology" in Semitic linguistics.

(4) Qur. *ṣadr-*; Leb. *sidr-*; Mec. *ṣadr*; Mlt. *sidēr* // No clear cognates outside Arb.[42]

(5) Gez. *ʔəngədʔā* or *ʔəngədʕā* // < Sem.: Tna. *ʔəngədʕā* (LGz 29), Arb. *naӡd-* 'mamelle', Hbr. *nägäd* 'in front of' (SED I No. 195)[43] < Afras. *nag(ʷ)V(H)d-* 'breast with neck' (W. Chad.: Kirfi *ngìḍò*, Galambu *ngìryà*, Diri *ŋgwáḍù* 'neck'), perhaps with *n-* prefixed < Afras. *gaʔid-* 'upper part of breast with the neck': Sem.: Arb. *ӡīd-* '*cou long et gracieux*' [BK 1 361]; E. Cush.: LEC: Somali *gaaddo* 'breast' (ADB).

(6) Tna. *ṭub*; Tgr. *ṭəb*; Amh. Arg. *ṭut*; Gaf. *ṭŭwwä*; Sod. *ṭəbuyyä*; Cha. *ṭu*; Har. *ṭōt*; Wol. *ṭub* // < Arb.-Eth.-MSA *ṭVb-* 'teat' (SED I No. 277).

(7) Jib. *géhɛʔ* (syn.); Soq. *gehe* (syn. 1) // < Sem. *gaw(w)iʔ-* '(front part of) body; chest, belly; interior' (SED I No. 99).

(8) Soq. *bérak* (syn. 2) // < Sem. *barak-* 'chest, thorax' (SED I No. 38).

◊ No terms in Pho. and Sab.

→ **Common South and West Semitic**: *ṭVdy-* (#2); no Afras. parallels.

    Common Semitic: *ʔir(r)-at-* (#1)[44] < Afras. *ʔVr(a)r-* chest and belly': Brb: E. Tawllemmet *a-hǎror* 'poitrine'; C. Chad.: Padokwo *arwa* 'chest', E. Chad.: Jegu *ʔurre* 'navel'; E. Cush.: LEC: Somali *ùur-*, pl. *úurár* 'stomach', Rendille *ûr*, pl. *urʔár* 'belly, abdomen', HEC: Burji *ír-a* 'stomach', Yaaku *irêh* 'belly'; N. Omot.: Mao *ʔaare* 'breast'.

12 BURN (tr.):

(1) Akk. *šarāpu*; Ugr. *šrp*; Hbr. *ŝrp* // < Sem. *ŝrp* (HAL 1358).

(2) Ugr. *ḥrr* (syn.); Gez. *ʔaḥrara*; Tna. *ḥarärä, ʔaḥrärä*; Tgr. *ḥarärä* // < Sem. *ḥrr* (HAL 357, LGz 243).

(3) Bib. Pal. *yḳd*; Syr. *ʔ-yḳd*; Urm. *ḳwd* (met.) // < Sem. *y/wḳd* (HALOT 430).

(4) Mnd. *ḵla* // < Sem. *ḵlw* (v. in LGz 431; cf. also EDE III 645).

(5) Qur. *ḥrḳ* VIII; Leb. *ḥarraʔ*; Mec. *ḥaraḳ* // No Sem. parallels that I know of. Related to Afras.: Brb. *HVrḳ* 'to burn': Ghadames *ărɣ*, Ghat *ərɣ*, Rif *arɣ* 'brûler', Ahaggar *ərəɣ* 'ê. enflammé', etc. (Kossm. 213), Egyp. Pyr. *rḳḥ* (met.; also *rkḥ* — a variant root with *k* vs. *ḳ*?) 'Feuer anfachen, verbrennen' (EG II 457–8).

(6) Mlt. *ṭabbat* // No straight parallels. To be tentatively compared either to Arb. *ṭbb* 'exercer la médecine' (BK 2 51; < Sem. *ṭbb* 'to know, be wise, treat medically', v. LGz 585) implying the semantic shift 'to cure' > 'to cure by cautery, cauterize' > 'to burn'; or to Arb. *ṭūb-* 'brique cuite' (BK 2 116; related to or borrowed into Eth., v. LGz 585).[45]

(7) Sab. *wfṭ*; Gez. *wafaṭa* (syn. 1) // Cf. also derived nouns: Gez. *mafaṭ, mafṭ, mofṭ* 'oven, furnace, pit for firing pottery', Tna. *mofṭ-i* 'firing of pottery' (borrowed from Gez.?). Seems to be an Eth.-Sab. root with no parallels in other Sem. (v. LGz 607).[46]

---

[42] Cf. Arb. *ṣidār-* 'chemise court, qui ne couvre que la poitrine, le thorax', *ṣadriyyat-* 'veste, gilet; chemisette' (BK 1 1319) apparently derived from *ṣadr-* 'poitrine' (ibid.) and Jud. *ṣədār-, ṣərād-* (met.) 'coarse web (of hemp), rough cloth' (Ja. 1264; 1299), cautiously compared in LS 346 with Soq. *miṣdéreh* 'tapis, vêtement en poil, sac'. Cf. also Mhr. *ṣēdər* 'stem, bow, prow (of a ship)' (JM 358), *ṣadēr-* 'Vorderseite' (ibid. after Jahn), Jib. *ṣédɛ́r* 'prow of a boat' (JJ 235), which are obviously borrowed from Arb. *ṣadr-* 'proue (d'un vasseau)'. Finally, cf. Syr. *ṣūdār-* 'crapula, nausea' (Brock. 622); the sensation caused by crapulence, hangover, or nausea may, in principle, be associated with 'breast'.

[43] Leslau quotes the Arb. and Hbr. forms yet considers neither of them satisfactory, obviously, for phonetic reasons; I, however, see no problem at all if we assume a prefixal *ʔV-*; as for the Auslaut, cf. Gez. *sanbuʔ, sambuʕ* 'lung' vs. Akk. *sinib/ptu* 'part of sheep's lung' (SED I No. 235) and similar examples (v. Mil. RE).

[44] One of the few exclusively Akk.-Ugr. isoglosses on the 100-word list, a remarkable fact discussed in Kog. Ug. 464., which, however, in no way implies any particular genetic closeness.

[45] Cf. also Eth. *ṭbs* 'to roast' ibid. 586, perhaps representing a relict causative with *-s* suffixed from *ṭb* 'to burn' with the meaning shift 'to bake/burn bricks' > 'to burn'.

[46] The comparison with Egyp. *wbd* 'brûler', mentioned in DRS 584 and strangely referred to in EDE I 285 as "not excluded", *is* excluded, since Egyp. *b* does not correspond to Sem. *p*. There are, however, two other possi-

(8) Gez. *ʔandada* (syn. 2); Tna. *ʔanäddädä* (syn. 1); Sod. *änäddädä* // Likely metathetically related to Arb. *ndʔ* 'faire un petit creux dans les cendres chaudes pour y mettre le pain, etc., qu'on veut faire cuire' (BK 2 1224); cf. also Hbr. *nad* (Is. 17:11) translated by Driver as 'to burn up' (quoted in LGz 385; not in HALOT).

(9) Gez. *ʔawʕaya* (syn. 3) // < Eth. **wʕy* 'to burn, be hot' (LGz 603: perhaps Arb. *wʕy* 'to stir up a riot'; semantically vague).

(10) Tna. *ʔaḳḳaṣälä* (syn. 2); Amh. *aḳaṭṭälä*; Arg. *əḳḳaṭṭäla*; Gaf. *(tä)ḳaṭṭälä* // No parallels that I could find in or outside Sem.[47]

(11) Cha. *mäk̇ärä* // < Gur. **mäġġärä*, derived with *m-* prefixed from Gur. **ġirġir balä* 'to blaze, flicker, burn in a bright and wavy way, burn easily (dry wood)' (ibid. 310). Related to Sem.: Amh. *gärrärä* 'spark (fire)' (ibid.), Akk. *girru* 'fire' OB on (CAD *g* 93). Perhaps to be further compared to Akk. *agurru* 'kiln-fired brick', according to Kauf. 33, a term of unknown etymology borrowed into Syr. *ʔgwrʔ*, whence into Arb. (*ʔaǯur-* 'brique cuite au feu' BK 1 13), but, anyway, rather related than not to the present root.[48]

(12) Har. *māgäda*; Wol. *magäda* // only Eth.; the comparison in LGur. 393–4 with Sem. **w/yḳd* is phonetically untenable.

(13) Mhr. *hə-nḥū*; Jib. *e-nḥé*; Soq. *ə-nḥi* // Comparable as forms containing the fossilized prefix *n-* to Gez. *ḥaw (haw)* and Tna. *ḥawwi* 'fire' (v. **FIRE** No. 3). Another parallel, semantically questionable, is Arb. *nāwaḥa* 'souffler du côté opposé à l'autre (se dit d'un vent)' (BK 2 1363) with the common underlying meaning 'to blow up fire'.

◊ Hrs. *ḥrōḳ* is very likely a lw. from Arb.; no term in Pho.

→ **Common North and West Semitic**: **ŝrp* (#1); no Afras. parallels found.
Common West Semitic: **ḥrr.*

13 CLAW (NAIL):

(1) Akk. *ṣupru*; Hbr. *ṣippōrän*; Bib. *ṭapar*; Pal. *ṭpr*; Syr. *ṭepr-*; Mnd. *ṭupr-*; Urm. *ṭarp-* (met.); Qur. *ẓufr-*; Leb. *ẓafir*; Mec. *ẓafr*; Mlt. *dufrēy*; Gez. *ṣəfər*; Tna. *čəfri*; Tgr. *čəfər*; Amh. Sod. Cha. Wol. *ṭəfər*; Arg. *čəfər*; Gaf. *ṣəfrä*; Har. *ṭifir*; Mhr. *ḏfēr*; Soq. *ṭifer* // < Sem. **ṭip(V)r-* (SED I No. 285).

(2) Hrs. *kef*; Mhr. *kaf* (syn.); Jib. *kéf* (also 'palm of the hand, paw') // < Sem. **kapp-* 'palm, flat of hand or foot' (SED I No. 148).

◊ No terms in Ugr., Pho. and Sab.

→ **Proto-Semitic**: **ṭip(V)r-* (#1) < Afras. **čipar-* ~ **čarap-* 'fingernail' (not quite reliable): C. Cush.: Qwara *ǯarfā*, Khamta *zafér* 'finger' (App. CDA 67)[49], E. Cush.: HEC: Burgi

---

bilities: (1) Egyp. (Med.) *wft* 'durchbohren' (EG I 307), which fits in well phonetically (Egyp. *-t* reflects Afras. **ṭ* in quite a number of cases, cf. EDE I 231–4) and is semantically tenable (for the "isosemantic string (or series or row)" 'to burn' → 'to drill' see Maizel 206–7), referring to a special technique of drilling (may eventually be akin to possibility #2); (2) Egyp. (NK) *fty* 'von der Bearbeitung von Metallwaffen', commented upon in EDE II 593 "The OEg. root, however, might have certainly been **fd*" (*d* being the most regular reflex of Afras. **ṭ*) and compared with Chad.: C. Chad. **vVḍ-* (< **fVḍ-*: Gisiga (Dogba) *vúḍ* 'to forge', Mada *váḍ* 'to forge', *ávàḍ* 'to heat, forge, pierce', Mafa *viḍ-* 'to forge, fabricate', E. Chad.: E. Dangla *pâḍé* 'to hammer the brand iron, to forge hot' (the Egyp. and Chad. forms are compared in EDE II 593). The resulting N. Afrasian root would be **fVṭ-* ~ **wVfVṭ-* 'fabricate, process by heating', perhaps (if Egyp. *wft* 'to drill' belongs here, and considering the meaning 'pierce' in Mada *ávàḍ*) 'and drilling'.

[47] Cf. C. Cush. Khamir *ḳaṭəls*, Kunfäl *kanṣäls* 'to burn', considered an Amharism in App. CDA 39; could it be the other way round, i. e. an Agaw loan in Eth.?

[48] Note, however, a related root in E. Cush. (e. g. Sidamo Hadiya *giir-* 'to burn' Huds.) which, in principle, could be a source for the Gurage forms, if they are borrowed. For broad Afras. connections, see EDE III 678–680.

[49] While Bilin *čəfər* and Qwara *ṭeffer* 'claw' (App. CDA 45) look like normal Ethiopisms, *z* in Xamtanga and *ǯ* with metathesis in Qwara ("the somewhat anomalous initial *j-* of the Qu. form" App. CDA 67), if these forms are related, rather speak against borrowing from Eth.

zurup-mata 'fingers' (pl.),[50] S. Cush: Alagwa, Burunge *čarafu* 'fingernail' (this metathetic form can hardly be a loan from Amh.).

14 CLOUD:

(1) Akk. *erpetu*; Ugr. *ʕrp-t* // < Sem. **ʕVrp-*: Hbr. *ʕrp* 'to drip', Arb. *ʕarf-at-* 'wind' (cf. HALOT 887; EDE I 296).

(2) Bib. *ʕānān*; Pal. *ʕănan*; Syr. *ʕənān-*; Mnd. *anan-*; Urm. *(ʕ)nān-* // < Sem.: Arb. *ɣayn-* 'nuage qui couvre et assombrit le ciel' (BK 2 527; *ʕannat-* and *ʕanān-* 'nuage' ibid. 377 may be borrowed from Syr.)[51].

(3) Hbr. *ʕāb*; Urm. *ʕayb-* // < Sem.**ɣayb-* (HAL 773).

(4) Qur. *saḥāb-;* Mec. *siḥāb*; Mlt. *sḥāp* // Apparenly < Arb. *sḥb* 'traîner par terre' (BK 1 1957, cf. *saḥāb-* 'nuage (surtout quand poussé par le vent il est en mouvement)' ibid.) < Sem. **sḥb* 'to drag, pull' (LGz 492–3; HAL 749; LS 284).

(5) Leb. *ɣeym* // < Arb.-Arm.: Arb. *ɣaym-*, Syr. *ʕaym-* 'nebula' (Brock. 522).

(6) Gez. *dammanā*; Tna. *däbäna, dämmäna*; Amh. Gaf. *dämmäna*; Arg. *dammäna, dona*; Sod. *dämmäna, dabäna*; Cha. *dabāra*; Har. *dāna*; Wol. *däbäna* // < Eth. **daman-* (with a variant root **daban-* in Mod. Eth. accounted for by **-m-* dissimilated from *-n-* into *-b-*) < Sem. **da/im(m)-*: Syr. *dīmətā də-ṭallā* 'nebula tenuis' (lit. 'fog of dew'), Arb. *damm-* 'nuage qui ne donne pas de pluie', *dimām-* 'nuage sans eau' (BK 1 728). The obvious connection with C. Cush. (Bilin *dɛmna*, Khamir *dəməna*, Kemant Qwara *dämäna*, Aungi *dammini* 'cloud' App. CDA 46) and E. Cush. (LEC: Oromo *dūman-sa*, Bayso *dumbo*, HEC: Burji *dumman-ci*, Darasa *duuman-ca*, Hadiya *duuba* id.) forms would suggest a Cush. borrowing into Eth., if not for the Syr. and Arb. cognates; Ethiopisms in Cush. are hardly likely either (v. the Hadiya form), though certain influence in both directions is possible.[52] I am inclined to regard the Sem. and Cush. forms, with some irrelevant exceptions, perhaps, as continuing Afras. **da/im(-an)-*, also including W. Chad. Tangale *hadam* 'rain', Hausa *dàmunā*, Ngizim *dòmán* 'rainy season', Bade *demanu* 'rain', *dàmànə́n* 'rainy season' and C. Chad. Logone *də́man* id. (ADB).[53]

(7) Tgr. *gimät* // < Eth. **gim-*: Gez. *gime* 'fog, cloud, dampness, mist, vapor', etc. (contra LGz 193, not related to Arb. *ɣaym-*, Syr. *ʕaym-*). No reliable Sem. parallels. One wonders whether it could be related to or borrowed from (or to?) Cush.: N. Cush: Beja *gīm, gēm*, E. Cush.: HEC: Sidamo *goma* 'cloud', *gomi-ččo* 'fog', S. Cush.: Dahalo *ŋgúmine* (also N. Omot.: Wolayta *guma* id., admittedly borrowed from Cush. or Amh.) (ADB).

(8) Hrs. *ʔāfor*; Mhr. *ʔafur*; Jib. *ʕafor* // Perhaps a meaning shift from 'dust cloud' (cf. Hrs. *ʔāfor* 'cloud, dust wind' JH 6) < Sem. **ʕapar-* 'dust' (DLU 85; HALOT 861–2); less likely, metathetically related to (or influenced by) Sem. **ʕrp* (v. #1). The most tenable comparanda, however, are in ESA: Sab. *ʕpr* 'sowing (land) before rain' (SD 13–14) and forms adduced in EDE II 389, under the discussion of possible various parallels to Egyp. (Pyr.) *pꜣʕ.t* 'irrigable land', all of them fitting into Afras. **ʕapur-* '(rainy) cloud, rain, rain-watered or irrigated

---

[50] Though the initial consonants in both Qwara/Xamtanga and Burji are irregular and hard to explain, they are hardly unrelated to the present root.

[51] Cf. HALOT 857–8, comparing Hbr. *ʕānān* 'clouds' and the Arm. forms with just one word which is not quite clearly quoted as "Arb. *ʕanna*, or a primary noun".

[52] Also N. Omot.: Koyra *dūma* 'cloud'.

[53] EDE III 603 quotes the Agaw and Koyra examples meaning 'cloud', comparing them directly with various Afras. forms meaning 'darkness', 'black' and 'night'. While the eventual kinship between the latter forms and the quoted group of terms meaning 'cloud' is not to be ruled out (the connection with 'rain' seems to me a stronger possibility), it would be methodically more correct to juxtapose the two groups taken separately, instead of mixing some of the terms from one group with the whole set of terms from the other.

area' (the MSA words meaning 'cloud' are unusually overlooked by a generally Argus-eyed Takács): Brb.: Ahaggar *a-fara*, pl. *i-ferw-ân* 'lieu couvert de végétation persistante'; W. Chad.: Dera *àpɔre* 'to shed, pour out', C. Chad.: Zime-Dari *pùwōr* 'pluie', E. Chad.: Kera *pɔ́ɔrú* 'Regenzeit'; N. Cush.: Beja *afra* 'Wolke'.

(9) Soq. *ḥeyhor* // < *ḥohar* 'black' (v. BLACK No. 7).

(10) Soq. *ʕalíloh* (syn.) // < MSA **γVlVl-*: Mhr. *γɘllēt*, Jib. *γíẑ́ɔ�̇t* 'mist' (JM 136). Compared in LS 310–11 with Arb. *taʕlūl-* 'masse de nuages formée par l'amoncellement des uns sur les autres' (BK 2 336), which, however, may go back to the verb *ʕll* 'tenir lieu d'une autre chose' (ibid. 334), thus having nothing to do with the present term; cf. also Arb. *γalal-* 'eau stagnante qui couvre pendant quelque temps la surface du sol et disparaît ensuite' (ibid. 488).

◊   No terms in Pho. and Sab.

→   **Common Semitic**: **ʕVrp-* (#1)[54] with isolated parallels in E. Chad.: Jegu *nyúrắpè* 'cloud' (with prefixed *n-*), Mogum (Jegu) *yurupe* 'cloud' (ADB).

15 COLD:

(1) Akk. *kaṣû*; Mlt. *kiesaḥ* // The two forms, if indeed related, may be traced to the phonetically immaculate Proto-Semitic form **kVṣaḥ-*.

(2) Hbr. *ḳar*; Pal. *ḳryr*; Syr. *ḳarīr-*; Mnd. *ḳarir-*; Urm. *ḳayr*; Gez. *ḳʷärir*; Tna. *ḳ̣ʷärri*; Tgr. *ḳɘrur* // < Sem. **ḳ̣ʷrr* 'to be cold' (v. in LGz 443; cf. **ḳurr-* 'freddo (s.)' Fron. 147).

(3) Pal. *ṣᵉnin* (syn.) // < Hbr.-Arm. **ṣinn-*: Jud. *ṣinnɘt-* 'cold', Hbr. **ṣinnā* id. (v. in HALOT 1037; no reliable parallels in other Sem.).

(4) Mnd. *ḳaruš-* (syn.) // < Sem. **ḳrš* 'to be frozen': Syr. *ḳrš* 'refrigeratus est' (Brock. 701), Pho. *ḳrš* 'to become frozen' (Tomb. 294), Arb. *ḳrs* 'ê. très-rigoureux (se ditdu froid); geler (se dit del'eau)' (BK 2 710).

(5) Qur. *bārid-*; Leb. *berid*; Mec. *bārid*; Gez. *bɘrud* (syn. 1), Tgr. *bɘrud* (syn. 1); Tna. *bärid* (syn. 1); Amh. *bärid, bärrad*; Sod. Wol. *bɘrd*; Har. *bäräd* // < Sem. **barad-* 'hail; cold', **brd* 'to be cold' (LGz 103).[55]

(6) Tna. *zɘḥul* (syn. 2) // < Eth.: Gez. *zɘḥla* 'to cool down' (LGz 634), Wol. *zul*, Selti *zūl* 'wind with cold'. No parallels outside Eth.[56]

(7) Amh. *ḳäzḳazza* (syn.); Arg. *ḳäzḳazza* // < Eth.: Gez. *ḳzz* 'to cool (off)', etc. (v. in LGz 457) with parallels in C. Cush. (Khamta *qazqɘz-äw*, Aungi *kezkazz-* considered loans from Amh. in App. CDA 46–47), N. Omot. (Dizi *ḳɛž-* 'wet, cold' Bnd Om. 220) and S. Omot. (Ari *qáž-í*, Dime *ḳ̌ɛ̌ž-in*, Hamar *ḳaž-* 'cold' Bnd Om. 47) — loans of Amh. *ḳaz-*, according to Bnd Om. 207. Cf. also W. Chad. Gwandara *àkúšúka*, E. Chad. Ubi *keckeci*, Munjile *kɔ̄súk* 'cold', Mubi *kùsúk* 'cold wind' (ADB).

(8) Hrs. *ḥebūr*, Mhr. *ḥɘbūr*; Jib. *ḥōr*; Soq. *ḥebhor* // < MSA **ḥVbūr*; the only parallels I can suggest is metathetic Arb. *bāriḥ-* 'hot wind' and Chad.: W.: Kirfi *bùrá* 'harmattan', C.: Mbara *bàràwáy*, Munjuk *ḅaray* 'tornado', Musgu *bɘrbɘr* 'cold (of wind)' (sic!), E.: Bidiya *ʔàbar* 'to blow (wind)', Kwang *ká-bār* 'wind'. If all these forms are related, Afras. **ḥVbūr- ~ *bāriḥ-* '(cold or hot) wind' can be reconstructed.

(9) Hrs. *ḳaṣm* (syn.); Mhr. *ḳāṣɘm* (syn. 1); JIb. *ḳéṣm* (syn. 1) // < MSA **ḳaṣm-*. No straight parallels in Sem. For possible Afras. matches cf. C. Cush.: Bilin *ḳɘšḳaš*, Khamir *ḫäšäš* 'cool',[57]

---

[54] This case is very similar to 11 BREAST #1, representing an exclusive Akk.-Ugr. isogloss (with some — if little — evidence from other Sem.); see note 44.

[55] EDE II 269 quotes a certain EEWC (I was unable to find this reference in any list of abbreviations in all three volumes of EDE) wherein this Sem. root is compared with Egyp. (NK) *brd* 'to be stark, stiff'; this is quite tenable.

[56] Leslau's suggestion (in LGur 707) "probably from Cushitic: Darasa *didallo* 'wind'" does not look tenable.

[57] According to App. CDA 47, Bilin *ḳɘšḳaš*, Khamta *qazqɘz-äw*, Khamir *ḫäšäš*, Aungi *kezkazz-* "are all clearly cognate though the variation in the sibilants especially prevents reduction to a common proto-form. The root oc-

E. Cush. Oromo *qaçaç-* ' to drizzle for many hours', S. Cush. Alagwa *qanca* 'rainy season' (Ehr PCR No. 147) and N. Omot. Dizi *ḳeč-*, Sheko *ḳeṭns* (Bnd Om 207), Janjero *koču* (ibid. 161) 'cold', supposedly < Afras. **ḳVç/s-* (then *-m* in the MSA forms is to be regarded as a fossilized prefix).

(10) Mhr. *ĉabil*; Jib. *ĉall* (both syn. 2) // No parallels found.

(11) Soq. *šeḳaḳ* (syn.) // Obviously to be connected with Har. *šiḳāḳ* 'a cold' (compared in LHar 146) with no other visible parallels in Sem.; cf., however, Brb. Siwa *šqi* 'froid' (Lao. 242) < Afras. **sVḳ(Vḳ)-*?

◊ No terms in Ugr., Bib., Pho., Sab. Cha. *ziza* (only in Gur. LGur 724) is likely a borrowing from Omot., cf. Sheko *záazzə* 'cold' (Bnd Om 207), Ari *zá(a)z-* id. (Bnd Ar 147).

→ **Common Semitic** (if the comparison in #1 is valid): **kVṣaḥ-*.

Common West Semitic 1: **ḳʷrr* 'to be cold' (#2) with parallels in E. Cush, if the latter are not loanwords from Amh.: LEC: Oromo *qorra* 'intense cold', HEC: Sidamo *qorre* 'cold'. Possibly related to Afras. **ḳVr-* 'dry': Sem.: Akk. *ḳarūru* 'drying', Urmian Arm.: *ḳayr-* 'dry'; Brb. **ḳʷar-* 'be dry'; C. Chad.: Mbara *kìwírì* 'dry season', E. Chad.: Bidiya *karay* 'make dry (cereals, land)'; C. Cush.: Khamir *xírə* 'dry' (<*ḳir-*), E. Cush.: LEC: Oromo *qōrā* 'dry'.

Common West Semitic 2: **barad-* 'hail; cold', **brd* 'to be cold' (#5).

16 COME:

(1) Akk. *alāku* // < Sem. **hlk* (v. in DLU 165).

(2) Ugr. *mɣy* (DLU 265; Kog. Ug.); Gez. *mṣ'*; Tna. *mäṣ'e*; Tgr. *mäṣ'a*; Amh. Arg. *mäṭṭa*; Sod. *mäṭṭa*; Wol *mäṭä*; // < Sem. **mṭ'* 'to reach, arrive' (v. in LGz 369–70; DLU 311; EDE III 877).

(3) Pho. *'t'*; Bib. *'ty/'*; Pal. Syr. Urm. *'ty*; Mnd. *ata*; Qur. *'ty*; Sab. *'tw*, Gez. *'atawa* // < Sem. **'ty/w* (v. in LGz 46–7).

(4) Hbr. *bw'* // < Sem. **bw'* (v. in HALOT 108; LGz 114–5; DLU 98).[58]

(5) Qur. *ǯy'* (syn.); Leb. *'əža* (met.); Mec. *žā'*; Mlt. *aža* (met.) // < **gy'*, likely related to Sem. **gʷVy(')* : Gez. *gʷayya* 'to run, flee', Soq. *ge* 'to flee, hurry', etc. (in LGz 209 the Arb. verb is not compared; cf. also DRS 107) and its reduplicated variant **gʷV'gʷV'-*: Gez. *gʷagʷə'a* 'to hurry, rush, flee', etc., compared with Arb. *(ta)žaža'a* 'to flee' in LGz 184. Arb. **gy'* 'to come' has solid Afras. parallels in W. Chad. Kanakuru *gài*, C. Chad. Kilba *gwà-*, Margi *gwa*, Masa *gɔio* 'to enter', Glavda *gwìyà* 'to return', E. Chad. Kabalai *gìyə* 'to come' (ADB); E. Cush. Afar *gay-* 'kommen' (RAf 853), Oromo *gaya* (Gr. 171), Darasa *ge-* (Huds. 21) 'to arrive'.

---

curs in Amh. *ḳäzäḳḳäzzä*, etc., and there has evidently been some cross interference; only Aungi and Khamta are obviously directly from Amh." I am somewhat confused about this assertion: if all the above Agaw forms are "clearly cognate", how come two of them are "directly from Amh." and the other two are not (and cannot be, judging by their form)?

[58] With numerous Afras. parallels (ADB), some of them adduced in EDE II 81. Proposing Proto-Cush. **baḥ-* 'to go out', based on E. Cush. **baḥ-* 'to go out', Takács also quotes Agaw **ba-t-* 'to leave' and further extends the comparison to N. Omot. forms (like Wolamo *bā-*, Yemsa *be'-*) and W. and C. Chad. forms (like Miya *bō-*, Margi *ba*). Since all the quoted C. Cush. (Agaw), N. Omot. and Chad. forms do not preserve either **ḥ* nor **'*, or are expected to preserve some traces of **ḥ* (but not **'*) which are obviously not there, I marvel at the author's knowledge inaccessible to me when he asserts: "The common LECu. -NWOmt. root (**baḥ-*; I wonder how it is known that the Omot. root is < **baḥ-* with *ḥ*? — AM) is often mistakenly (sic! — AM) equated with Bed. *bāy* "to go", Agaw **fi-* "to go out" [GT]...and Sem. **bw'* "to enter"." [GT] stands for Gábor Takács, and it is hard to understand whether "mistakenly" refers to the author as well (which would be correct in the case of Agaw **fi-* that has nothing to do with the Afras. root in **b-*), or only to his unnamed opponents. Anyway, except for E. Cush. **baḥ-*, I cannot find any criteria to discern between the two roots, which, I am afraid, seriously endangers my professional reputation.

(6) Gaf. *sällä* // S. Eth. only (LGur 542).

(7) Har. *diǧa* // According to LHar 55 and LGur 315, borrowed, together with other S.-E. Eth. (Wol. *žeže*, Selti *žēže*, Zway *žīžī* 'to arrive, reach') from HEC Darasa *dáge*, Sidamo *dayi*. More likely, however, <*dida*, with *d > ž* (v. LHar 7 and 9; LGur XLIV) < Sem. **dydy* 'to arrive, come, walk' (v. in HALOT 214 and DRS 223).

(8) Cha. *čänä-m* // Only Gur., according to LGur 174; likely represents **tan-* (on *č < *t* in Gur. v. ibid. LXII), comparable with Arb. *tnʔ* 's'arrêter et séjourner dans un endroit' (BK 1 208).

(9) Hrs. *nōka*; Mhr. *nūkaʕ*; Jib. *nikaʕ*; Soq. *nkʕ* // < Sem. (Arb.-MSA; the difference in meaning rather rules out Arb. borrowing into MSA): Arb. *nkʕ* 'partir, s'en aller, s'eloigner' (BK 2 1343); unconvincingly compared in LS 267 with Arb. *nkḥ* 'cohabiter avec une femme'. Cf. a possible, if isolated, parallel in C. Chad. Mofu *-nakwá-* 'aller, marcher' etc. (Stolb. 2005 230).

(10) Jib. *zaḥám* (syn.) // Likely a meaning shift from 'to push one's way in the crowd', cf. *zaḥmέt* 'crowd', *sá-zéḥəm* 'to jostle in a crowd', *zaḥmún* 'arrival; one who pushes' (JJ 318) < Sem. (Arb.-MSA; unless an Arabism in MSA): Arb. *zḥm* 'serrer, reserrer (dans un espace droit)', *zaḥm-* 'foule qui se presse dans un espace étroit' (BK 1 979).

(11) Soq. *ʔéraḥ* (syn.) // < Sem. **ʔurḫ-* 'way, road' (v. ROAD No. 1; HALOT 86).

→ **Proto-West Semitic**: **ʔty/w* (#3) < Afras. **ʔa/it-* 'walk, come and go': W. Chad.: Bokkos *ʔat* 'travel', E. Chad.: Mokilko *ʔétté* 'to go, leave, come', Dangla *ə̀tɛ* 'to arrive'; N. Cush.: Beja *ʔat* 'tread, march', C. Cush. **ʔənt-(ät-)* 'to come' (CDA), E. Cush.: LEC: Arbore *ʔiʔit-*, Elmolo *iit* 'to walk, go'; S. Omot.: Ari *aata* 'to come' (ADB).

Common West Semitic: **mṭʔ* (#2).

17 DIE:

(1) Akkadian *muātu*; Ugr. Pho. *mt*; Hbr. Pal. Qur. Sab. Gez. *mwt*; Syr. Urm. *myt*; Mnd. *mit*; Leb. *mət*; Mec. *māt*; Mlt. *mīt*; Tna. *motä*; Tgr. Amh. Wol. *motä*; Arg. *moda*; Sod. *motäm*; Cha. *mʷätäm*; Har. *mōta*; Hrs. Mhr. *mōt* // < Sem. **mwt* (v. in LGz 375–6).

(2) Gaf. *fättärä* // < S. Eth.: Amh. *a-fättärä* 'faire mourir subitement' (LGaf 199 after Guidi), Endegeñ *(a)fettärä* 'to hit someone so as to nearly kill him'.[59] Cf. also Arb. *ftr* 'tomber dans la longueur, faiblir après un effort' (BK 2 534). One wonders whether these forms could be related, assuming a fossilized suffixal *-r*, to MSA nouns (Hrs. *fyet*, Mhr. *fōtēt*, Jib. *fétét* 'carcass of an unslaughtered animal') and verbs: Mhr. *fōt*, *ftōt* '(animal) to die unslaughtered' (JM109), Jib. *fēt* 'to die without being slaughtered' (JJ 67), related, in turn, to Arb. *fwt* 'mourir', unless the latter is a secondary semantic shift from the other meaning of *fwt*, 'passer' (see BK 2 642), in which case the MSA forms should rather be treated as Arabisms. The only isolated form can be found in E. Chad.: Mokilko *púutè* 'cadavre' (compared with the MSA forms, but not with the Arabic one, in EDE II 540).

(3) Hrs. *ɣāb* (syn.) // Meaning shift from 'to faint' (*ɣeyōb* JH 48), cf. Mhr. *ɣəyōb*, E. Jib. *ɣáb* ' to faint, be absent' (JM 146). Related to Arb. *ɣyb* 'ê. absent, caché, disparaître' rather than borrowed from it (cf. the expression in Arb. *ɣuyyabatu ɣuyyābutan* 'il est mort') BK 2 521. Note an isolated parallel in E. Chad.: Mokilko *goʔobè* 'dead, corps' (ADB).

(4) Mhr. *ɣəzōl* (syn.) // An unusual meaning shift (rather than an homonym) from the other meaning of this verb — 'to spin' (JM 148); cf. Jib. *ɣɔ́zɔ́l* 'to spin; to fall down in a swoon, to be on the point of dying' (JJ 92), Soq. *ʔɔ́zɔl* 'to spin' (JM 148) < Arb.-MSA (perhaps an Arabism in MSA): Arb. *ɣzl* 'filer (le lin, etc.)' (BK 464).

---

[59] In LGur 248 compared with hesitation to Cha. *(a)fätärä* and the like 'to finish (up)', but, strangely, not compared with the Gaf. and Amh. forms.

(5) Jib. *ḫáróg* // < MSA: Mhr. *ḫrūg* 'to take out, draw out, pull out' (JM 447), Soq. *ḫrg* 'cesser, ê. defendu' (LS 188) < Sem. (Arb.-MSA):[60] Arb. *ḫrž* 'sortir, quitter un endroit; paraître au dehors' (BK 2 554).

(6) Jib. *enúsum, əntsím* (syn.) // Also 'to breathe one's last' < 'to breathe': Hrs. *ansōm* 'to breathe', *šenésem* 'to sigh' (JH 97), Mhr. *hənsōm* 'to breathe' (JM 300) < Sem. *nšm* 'to breathe' (SED I Verb No. 50).

◊ Soq. *ṣame* is likely a loan of Arb *ṣmy* 'tomber roide mort, ê. tué sur place' (BK 2 1373).

→ **Proto-Semitic**: **mwt* (#1) < Afras. **mawVt-* 'die': Brb. **immut*; W. Chad. **mawut-*, C. Chad. **mVtV-*, E. Chad. **mawut-*; E. Cush.: LEC: Somali *mōd/t* 'death', Oromo *a-mutaa* 'mourning', Rendille *-mut-* 'to die', Gidole *muut-* 'become very weak and close to death' (ADB; EDE III 683–690).

18 DOG:

(1) Akk. *kalbu*; Ugr. Pho. Pal. *klb*; Hbr. *käläb*; Syr. Mnd. Urm. Qur. *kalb-*; Leb. *kaləb*; Mec. *kalb*; Mlt. *kelp*; Gez. Tgr. *käləb*; Tna. *kälbi*; Jib. *kob*; Soq. *kalb* (viewed by some authors as an Arabism, their argument being that there originally were no dogs in the island of Soqotra) // < Sem. **kalb-* (v. in DLU 214; LGz 282).

(2) Hrs. Mhr. *mābayl* // lit. 'owned', cf. Jib. *baʕál* 'to own' (JJ 22) < Sem. **bʕl* 'to own' (v. in HALOT 142–3).

◊ Amh. *wušša, wəšša*, Arg. *wəšša*, Gaf. *wəššä*, Sod. *wəssa* are < HEC (Sidamo *woši-ččo*, etc. LGz 667); Cha. Wol. *bučo*, Har. *buči* are < Oromo *buči* (LGz 130). No terms in Bib. and Sab.

→ **Proto-Semitic**: **kalb-* (#1), perhaps continues, with **-b* suffixed (see Mil. RE) Afras. **kʷVl-* 'dog, wolf': (?) Sem.: Gez. *kʷähila* 'fox-like animal'; Brb.: Ahaggar *ă-kûlen* 'loup, loup peint (lycaon)' (non us. dans l'Ah.) F. 799; (?) C. Chad.: Logone *kəle*, Buduma *kelī* 'dog' (otherwise <*kVr-*); C. Cush.: Waag *kuli* 'dog' (ADB).

19 DRINK:

(1) Akk. Ugr. Hbr. Pal. Syr. Urm. *šty*; Bib. *štʔ*; Mnd. *šta*; Gez. *satya*, Tna. *sätäyä*; Tgr. *säta*; Arg. *šäčča*; Har. *säča*; Wol. *säče* (*-č-<*t*) // < Sem. **šty* (v. in DLU 458; LGz 516).

(2) Qur. *šrb*; Leb. *šərəb*; Mec. *širib*; Mlt. *šōrop*; Gez. *ŝ/saraba* // < Sem. **ŝrṗ*:[61] 'to drink, swallow, suck': Akk. *sarāpu* 'to sip (?)' (CAD *s* 172), Hbr. pB. *ŝrp* 'to absorb, quaff, sip, suck' (Ja. 1632), Jud. id. (ibid.), Syr. *srp* 'suxit; sorpsit' (Brock. 500), Gez. *saraṗa* 'to celebrate Mass, bless an object, sip (the sipping of the blessed wine being a part of the Mass)' (LGz 513), Tna. *s/šäräbä* 'to approach (rain), condense (gas to liquid), etc. (Kane T 674), etc. (v. in LGz 533).

(3) Sod. *säččäm*; Cha. *säčäm* (*-č < *ḳ*); Hrs. *teḳ* (*heḳō* caus.; *h-* < **š-*); Mhr. *hutḳi* <s-t-ḳy>; Jib. *šuṣ~i* // < Sem. **šḳy* 'to give to drink; to water, irrigate' (v. in LGz 511).

(4) Soq. *re* // < Sem. **rwy* 'to drink one's fill': Hbr. *rwy* 'to quench thirst, drink to saturation', etc. (v. in LGz 478).

◊ Amh. *ṭäṭṭa* and Gaf. *ṭiṭṭä*, to which no parallels in Sem. seem to exist, are considered with hesitation in LGaf 242 to be loanwords from Oromo *ḍuḍan*, although the similarity is not overwhelming. No terms in Pho. and Sab.

→ **Common North and West Semitic**: **šty* (#1) with a C. Chad. parallel: Bura *sata* 'to drink', Matakam *sawat-* 'to be, make thirsty'.

---

[60] Because of the difference in meaning, borrowing from Arb. is less likely.

[61] On Sem. **ṗ* v. SED I CV–CXVI and SED II LX–LXI.

20 DRY:

(1) Akk. *šābulu* // < *abālu* 'to dry up, dry out' (CAD *a*1 29) < Sem. **ʔbl*; Hbr. *ʔbl* 'to dry up', Arb. *ʔubullat-* 'dried figs' (v. in HALOT 7).

(2) Hbr. *yābēš*; Pal. *ybyš*; Syr. *yabbīš-*; Mnd. *yabuš-*; Qur. *yābis-*; Sab. *ybs₁*; Gez. Tgr. *yəbus* // < Sem. **ybš* 'to be dry' (v. in LGz 626).

(3) Leb. *nešif*; Mec. *naššaf*; Mlt. *nīšef* // I have not been able to find any parallels.

(4) Tna. *nəḵus* // Eth.: Gez. *naḵsa* 'to dry up, be exhausted, be split, etc.', Tgr. *näḵsa* 'to become weak', reasonably compared in LGz 400 with Arb. *nḵṣ* 'to diminish, wane', Sab. *h-nḵṣ* 'to diminish' (after Biella; in SD 98, *hnḵṣn* and *hḵṣn* 'to cede, concede', *mḵṣ-m* 'loss, damage'), Mnd. *nḵṣ* 'to decrease'.

(5) Amh. Arg. Sod. Har. Wol. *däräḵ*; Cha. *ṭäräḵ* // Eth.: Tna. *däräḵ* (rare, according to my informants, unless an Amharism). Compared in DRS 318 with Arb. *darḵ-* 'dur'. Cf. C. Cush.: Kemant *dərḵ* and Aungi *dərk* 'draught', considered by Appleyard loans from Amh.

(6) Hrs. *ḵôša*; Mhr. *ḵayŝaʕ*; Jib. *ḵəŝaʕun*; Soq. *ḵeŝaʕ* // < Sem. **ḵaŝVʕ-?* (Arb.-MSA or an Arabism in MSA): Arb. *ḵašiʕ-* 'sec, desséché' (BK 2 743); compared in LS 389. Cf. W. Chad.: Hausa *ḵèḵasà* 'to dry (soil, clothes)', possibly <*ḵVĉḵaĉ-*, perhaps implying Afras. **ḵaĉ(ʕ)-* 'dry'.

◊ Urm. *bārūz-* has no parallels outside Neo-Aramaic and has to be treated as a loan-word. No terms in Ugr. Pho. Bib. and Gaf.

→ **Proto-West Semitic**: **ybš* (#2), perhaps < Afras. **bVs-*: (?) Egyp. (Coptic): "Subahmimic" *bōsʕt*, Sahidic *bosʕt*, *bast* (derived verbal forms);[62] W. Chad.: Hausa *búšè* 'to be dry, dry up', (?) Dera *bášà* 'harvest season' (from 'dry season, season with no rain'?).

21 EAR:

(1) Akk. *uznu*; Ugr. *ʔudn*; Hbr. *ʔōzän*; Pal. *ʔdn*; Syr. *ʔedn-*; Mnd. *ʕudn-*; Qur. *ʔuḏn-*; Leb. *ʔəḏən*; Mec. *ʔiḏin*; Mlt. *widna*; Gez. Tgr. *ʔəzən*; Tna. *ʔəzni*; Arg. *izin*, *əzən*; Gaf. *əznä*; Sod. *ənzən*; Cha. *ənzər*; Har. *uzun*; Wol. *əzən*; Hrs *ḥeyḏēn*; Mhr. *hayḏin*; Jib. *ʔiḏεn*; Soq. *idihən* // < Sem. **ʔu/iḏn-* (SED I No. 4).

(2) Urm. *nāt-* // L. Kogan (oral communication) thinks that it can hardly be separated from **ʔu/iḏn-*, but I cannot imagine such a phonetic development. The only suggestion, though semantically rather weak, that occurs to me is to compare it (as a jargonism? borrowed from an Arb. dialect?) with Arb. *nātⁱⁿ* 'enflé (membre du corps); saillant, protuberant' (BK 2 1195), *nātiʔ-* 'qui est en sallie' (ibid. 1191) or *nyt* 'ê. très-faible au point de ne pas pouvoir se tenir solidement et au point de pencher d'un côté ou de l'autre' (ibid. 1375). Otherwise, to be treated as a loan from an unidentified source.

(3) Hrs. *mēšmēʔ* (syn.) // < Sem. **šmʕ* 'to hear' (v. in LGz 501–2).

◊ Amh. *ǯoro* is borrowed from Oromo *gurra* (Gr. 188); on Amh. *ǯ* < **g* v. SED I LXIX; LXXXII–LXXXV. No terms in Pho., Bib. and Sab.

→ **Proto-Semitic**: **ʔu/iḏn-* < Afras. **ʔi/uǯ-n- ~ ʔi/udn-* 'ear': Egyp. *îdn*, phonetic value of the 'ear' hieroglyph determinative;[63] E. Chad.: **ʔudu/in-* 'ear':[64] Dangla *ḍéŋgei*, Jegu *ʔúdúŋê*,

---

[62] According to Takács, who, in EDE II 318–19, compares the Coptic forms with W. Chad. and Sem. ones (and adduces some more fairly tenable Sem. examples, besides those <*ybš*, proposed by A. Zaborski and A. Belova), "the Egyp. root is undoubtedly related to AA (Afras. — AM) **b-s* 'dry'" (ibid. 318). Except for the adverb "undoubtedly", I am inclined to accept this comparison as plausible.

[63] Egyp. *d* < Afras. *ǯ* is rare but confirmed by a few irrefutable examples, *îdn* being one of them, cf. EDE I 317–18.

[64] It is hard to imagine that the E.Chad. forms are not related to Egyp. *îdn* and, hence, to the entire Afras. root, though *d-* < **ǯ-* looks somewhat strange; perhaps, *d-* < **ǯ-* in both Egyp. and E. Chad. reflects some unexplained

Birgit *ʔúdúŋì*; C. Cush. **waʒ-* 'to hear': Bilin *was*, Khamir *waz/ʒ*, Khamta *waš* (App. CDA 82); N. Omot. **waʒ-*: Male *wayz* 'to hear', *woyzi*, Chara *wááza* 'ear', etc. (ADB; cf. also EDE I 83).

22 EARTH:

(1) Akk. *erṣetu*; Ugr. *ʔarṣ, ʔarṣu*; Hbr. *ʔäräṣ*; Pho. *ʔrṣ*; Bib. *ʔăraʕ*; Pal. *ʔrʕ*; Syr. Urm. *ʔarʕ-*; Mnd. *arḳ-*; Qur. *ʔarḍ-*; Leb. *ʔaraḍ*; Mec. *ʔarḍ*; Mlt. *art*; Sab. *ʔrḍ*; Jib. *ʔɛrĉ* // < Sem. **ʔarṣ̂-* (v. in. DLU 51).

(2) Gez. *mədr*; Tna. *mədri*; Tgr. Amh. Arg. Sod. *mədər* // < Sem. **midr-* (v. in. LGz 330; Kog. Eth. 378; EDE III 786).[65]

(3) Gez. *maret*, Tna. Amh. Arg. *märet* (syn.) // < Eth. **mar-(V)t-*, probably also Sab. *mrt-n* 'limestone?' (SD 86; compared in LGz 361 where the Sab. form is quoted as *mrt-m*)[66]; with reliable Afras. parallels: Brb. Ghadames *ta-mmur-t* 'terre, sol' (Lan. 215), Rif *ṯa-mur-ṯ* 'pays, contrée, territoire', Shawiya *ṯa-mur-ṯ* 'terrains propres à la culture' (MCB 258), etc.; Egyp. OK *mr* 'Viehweide' (EG II 97); E. Chad. Sokoro *māro* 'feuchte Erde' (LZS 42).

(4) Gaf. *afärä*; Cha. Har. Wol. *afär* // Either < Sem. **ʔapar-* 'dust, soil; ashes' (Hbr. *ʔēpär* 'loose soil crumbling into dust; ashes' HALOT 80, Gez. *ʔāfar* 'dust, soil'[67]) or < Sem. **ʕapar-* 'dust, soil' (HALOT 861–2 erroneously includes "Eth. *ʔafer*"; should add Tgr. *ʕafär* 'dust; desert' LH 492).

(5) Jib. *gədrét* (syn.) // Compared in JJ 71 with Soq. *gədhar* 'reddish-brown' (not in LS). Probably to be compared (as a form with fossilized suffixal *-r*) to Arb. *ʒadad-* 'terrain uni et dur' (BK 1 260), having Afras. parallels in C. Chad. Masa *nàgàdà* 'earth' (CLR II 117), E. Chad. Sokoro *gédē* 'fruchtbar Erde' (LZS 43) and S. Cush. Dahalo *gudde* 'land' (EEN 32).

(6) Hrs. *ḥōhi*; Soq. *ḥohi (ḥoihe)* // < Sem. **ḥašaw/y-*: Jib. *ḥáši* 'soil', *aḥšé* 'to play with dust' (JJ 118), Tna. *ḥašăwa*, Amh. *aššäwa*, Arg. *hašawa*, Wol. *ašawa* 'sand' (LGur 102).

(7) Mhr. *ḳāʕ* // Same as Hrs. *ḳā* 'land, ground' connected with Arb. *ḳāʕ-* (*<ḳwʕ*) 'plaine, terrain plat; terraine bas où l'eau demeure stagnante' (BK 2 835);[68] perhaps further related to Egyp. (MK) *kȝḥ*[69] 'Erdreich; Nilerde' (EG V 12) and C. Chad. Musgu *káikai*, Mulwi *kàykày*, Munjuk *kaykay* 'sand' (ADB).

◊ Wol. *däčče* (syn.) is borrowed from E. Cush.: Oromo *dačči*, Hadiya *däččeʔe* (LGur 198).

→ **Proto-Semitic**: **ʔarṣ̂-* (#1) < Afras. **ʔariĉ-* 'earth': Egyp. (MK) 'bewässertes land' (EG I 168);[70] W. Chad.: Paʔa *riṣa*, Siri *rəṣu*, etc. 'earth', E. Chad.: Bidiya *ʔírǎḍyà* 'valley' (ADB).

---

secondary phonetic process, common of Egyp. and Chadic (making, together with Berber, the African North Afrasian subbranch of Afrasian, in my classification).

[65] Cf. Egyp. (Med.) *mȝd* 'ein mineralischer Stoff', compared in EDE III 127, among other things, with ESA-Ethiopian root for 'earth, soil, clay (or limestone)' (**mVr-t-*, see #3). Though phonetically unacceptable (with a meaningless comment: "perhaps an irregular (Eg. *d-* vs. Sem **-t*)" ibid. 128), this comparison leaves open the possibility of comparing the meaning of the Egyp. word with 'earth', in which case it is a potential match with Sem. **midr-* (through metathesis). See the discussion on some other possible connections of the Sem. term in EDE III 786–7.

[66] See a more detailed discussion in EDE III 128–9.

[67] In LGz 10, related to the S. Eth. forms and provided with the following comment: "Dillmann 808 considers G. an Amharic loanword, unless it is to be identified with Heb. *ʔēpär*".

[68] Borrowing from Arb. into MSA cannot be ruled out.

[69] With a peculiar phonetic development, due to the vicinity of *ȝ* (< **ʔ*) and *ʕ* in one root?

[70] In EDE I 258 the unexpected *ʕ-* (< **ʔ*) is tentatively explained as "interchange of *j* (which I prefer to render as *y-* so as not to confound it with *j*, often inconsistently rendering [ʒ] in Afrasian studies — A.M.) ~ *ʕ* in the proximity of *ḍ* in Eg.". I tend to explain it out of **ʔVrḍ* (*<*ʔVrĉ*), with the guttural or uvular or "burring" [R] (rendered in Egyp. in this case, like in many others, by *ȝ*), which assimilated the glottal stop in the Anlaut. Cf. a similar process

23 EAT:

(1) Akk. *akālu*; Hbr. Pho. Bib. Pal. Syr. Urm. Qur. *ʔkl*; Mnd. *akal*; Leb. *ʔakəl*; Mec. *ʔakal*; Mlt. *kiel* // < Sem. **ʔkl* (v. in DLU 21; LGz 15).

(2) Ugr. *lḥm* // < Sem. **laḥm-* 'food (bread or meat)' (v. in DLU 243; HALOT 500; Kog. DD).

(3) Pal. *ṭʕm* (syn.) // < Sem. **ṭʕm* 'to taste' (v. in LGz 583).

(4) Gez. *blʕ*; Tna. *bälʕe*; Tgr. *bälʕa*; Amh. *bälla*; Arg. *bälla, əlʔa*; Gaf. *bällä*; Sod. *bällam*; Cha. *bänam*; Har. *bäläʔa*; Wol. *bälä* // < Sem. **blʕ* 'to swallow, eat' (LGz 94–5).

(5) Hrs. *tewō*; Mhr. *tu*; Jib. *te*; Soq. *té* // < Sem. **tʔw/y*: Akk. *taʔû* 'essen, weiden' (AHw 1341; no MSA parallels quoted) < Afras. **tiʔw-*[71]: Brb.: Ayr *ăttyu*, Ahaggar *tətt*, Ghadames *tätt*, etc. (habitative) 'to eat'; W. Chad.: Hausa *čí*, Dera *twi/a*, Siri *tuu*, Daffo-Budura *čuh*, etc. 'to eat (soft things)', C. Chad.: Lame *-tí-*, etc. 'to eat', E. Chad.: Migama *tíyáw*, Birgit *túwà* 'to eat soft things'; N. Cush.: Beja *tiyu* 'to eat' (ADB).

◊ No term. in Sab.

→ **Common North and West Semitic**: **ʔkl* (#1), cf. W. Chad.: Hausa *kằlằ-čī* 'food' (ADB).

24 EGG:

(1) Akk. *pelû* // Related either to Sem. **pūl-* 'bean' (Hbr. *pōl*, Arb. *fūl-* HALOT 918) or, more likely, to Afras. **p̣il(?)- ~ *p̣ulp̣ul-*: W. Chad. Ngamo *ḅila* 'egg', C. Chad. Banana *bọ̀lọ́ʔá* 'egg-shell', E. Cush. Burji *bulbul-é, bubul-é* (treated by Sasse as N. Omot. loan), Yaaku *bolbŏlî̂ʔ*, N. Omot. Male *ḅūla*, Wolayta *p̣up̣uliya*, etc., S. Omot. Hamar *ḅūla* 'egg'.[72]

(2) Hbr. *bēyṣā*; Pal. *byʕh*; Syr. *bēʕt-*; Mnd. *bit-*; Urm. *biyy-*; Qur. *bayḍat-*; Leb. Mec. *bayḍa*; Mlt. *bayda* // < Sem. **bayṣ̂-at-* (SED I No. 43).

(3) Syr. *bar-t-* (syn.) // Presumably < Sem. **barr-* 'wheat' (v. in HALOT 153; Mil. Farm. 138) with a meaning shift 'corn' > 'egg'.

(4) Gez. *ʔanḳoḳəho*; Tna. *ʔənk̇ʷak̇ʷəho*; Tgr. *ʔənḳoḳho*; Gaf. *anḳʷä*; Sod. *anḳo*; Har. *aḳuḥ*; Wol. *ənḳaḳot* // Supposedly < Sem. **ḳʷaḳʷay-* 'egg' (cf. SED I No. 160) with **ʔan-* prefixed and *-ḥ* explained as the result of contamination with Mod. Eth. **ʔVn-ḳulaliḥ-* (v. below). However, it must be somehow connected with Cush.: Beja *kŭáhi* (RBeḍ 137–8; <*k̇ʷaḥ-*), Saho *unqōqahó* (ibid.), Oromo *hanqāqū* (Gr.; < **ḥanḳaḳ-*, with metathesis?), Dasenech *ġonġono* (Tos. Das. 543), Hadiya *ḳunḳa* (Huds.), Maʔa *ikokoha* (HRSC 386; <**ʔi-ḳVḳVḥ-?*), Iraqw *qânḥi* (ibid.; <**ḳanḥ-*) 'egg', while neither Eth. nor Cush. forms look like loanwords from each other (perhaps except Saho).

(5) Amh. Arg. *ənḳulal*; Cha. *ənḳura*; Mhr. *k̇áwḥəl*; Jib. *ḳeḥẑin*; Soq. *ḳ̇holhin* // < Sem. **ḳa(w)ḥil-* (cf. SED I No. 170). Relations with a Cush.-Omot. term (e.g. C. Cush.: Bilin *käḥaluna, käḥala*, Khamir *ḳäluna*, Khamta *qululūna* App. CDA 59–60; N. Omot.: Wolayta *ḳuḳulliya* Lmb-Sot 430[73]) are not quite clear; as for C. Cush. Khamta *enqulal*, Aungi *ənkʷlal*, Appleyard regards them as Amharisms, which is possible, but the rest of the Agaw forms require us to explain how Amh. *ən-* could become lost in the process of borrowing. For a possibility of a common Afras. root, cf. C. Chad.: Bata *kwal* 'egg'.

---

in Egypt. (OK) *ʕm* 'Asians', rendering, in my opinion, *ʔārammī* 'Arameans' (very likely, the common ethnonym for speakers of Proto-Canaanite-Aramaic, or, in my classification, Proto-South Levantine) and several other cases that deserve a separate study.

[71] Judging by the Afrasian *comparanda*, Sem. **tʔw/y* might have been the original verbal root for 'to eat', later substituted in North and West Sem. by **ʔkl*.

[72] Compared in EDE II 68, but with the following comment: "... Ometo **ḅ* ... is difficult to explain from AA **b*", Afras. **p̣* not admitted. Perhaps related to Egyp. *py.w* (pl., grain determinative), probably 'small round object' (EDE II 68–9; 413), if < **pVl-*; Afras. **p̣* yields Egyp. *p*.

[73] With many fantastic comparisons.

(6) Hrs. *beḵelēt* // < Sem. **baḵʷal-* 'plant, vegetation' (v. in LGz 100).

(7) Hrs. *bēḏeh*, Mhr. *bēḏáyt*, JIb. *béḏ* (all syn.) // < Sem. (Arb.-MSA; because of the serious difference in meaning cannot be suspected to represent an Arabism in MSA) **bayṭ-at-*: Arb. *bayẓat-* 'oeufs des fourmi' (v. SED I No. 43 note). A variant root of **bayŝ-at-*, to be scored differently.

◊ No terms in Ugr., Pho., Bib. and Sab.

→ **Common West Semitic**: **bayŝ-at-* (#2) < Afras. **bayĉ-*: W. Chad. **(m-)bwiĉ-* 'egg': Geji *mbúŝī*, Zaar *buùŝ*, Sayanchi *mbúŝ*, Zul *mbúŝe*, etc. (ADB); cf. also discussion in EDE II 363–4.

25 EYE:

(1) Akk. *īnu*; Ugr. Pho. *ʕn*; Hbr. *ʕayin*; Pal. *ʕyyn*; Syr. Urm. Qur.*ʕayn-*; Mnd. *ayn-*; Leb. Mec. *ʕayn*; Mlt. *(gh)ayn*; Sab. *ʕyn*; Gez. *ʕayn*; Tna. *ʕayni*; Tgr. *ʕən*; Amh. *ayn*; Arg. Cha. *en*; Gaf. *inä*; Sod. Wol. *in*; Har. *īn*; Hrs. *ʔāyn*; Mhr. Soq. *ʕayn*; Jib. *ʕihn* // < Sem. **ʕayn-* (SED I No. 28).

→ **Proto-Semitic**: **ʕayn-* < Afras. **ʕayVn-* 'eye; to see': Egyp. *ʕn*, *ʕyn*, hieroglyph determinative sign for 'eye'; Brb. **HVnVy ~ nVHVy* 'to see': Ayr *ənəy*, Taneslemt *ənh*, Adghaq *ənhi*, Izayan *anni*, etc.; W. Chad. **HayVn-* 'to see': Bolewa *ʔinn-*, Polchi *yeni*, Paa *ḥan*, Tule *yāni*, Fyer *yaána*, Daffo-Butura *yen*, etc., C. Chad.: Gaanda *ànnì*, Gerka *anana* 'to see, find', (?) E. Chad.: Jegu *ʔinn-* 'to know'; S. Cush: Dahalo *ʕeen-aaḏ* 'to see from afar'; (?) N. Omot.: Gimirra *an* 'eye' (ADB; Cf. EDE I 125–6, where this root is confounded with Afras. **ʔi(n)ṭ-* 'eye').[74]

26 FAT (n.):

(1) Akk. *lūpû* (*lipiu*, *lī/ēpu*) // < Sem. **li/apiʔ-* 'fatty, fleshy tissue' (cf. SED I 180).

(2) Ugr. *šmt*, *šmn* // < Sem. **šam(-an)-* 'fat, oil' (cf. SED I 248) < Afras. **sim-an- ~ *sin-am-* [75] 'oil, fat, (fat) milk': Brb.: Ghat *isim* 'graisse (de tout animal)', *ésim* 'graisse fondu', Qabyle *ṭa-ssəm-ṯ* 'graisse animal', Canarian (Ferro) *achemen* 'milk' (<*a- *šVmVn*); Egyp. (Med.) *smy* 'fat milk, cream'; W. Chad.: Jimi *sin*, Diri *sinama* 'oil', E. Chad.: Somrai *swānī*, Kera *sɔn*, Migama *séwén*, Sokoro *súnu* 'oil'; N. Cush.: Beja *símma* 'fat' (n.), C. Cush.: Bilin, Khamir, Qemant *səna*, Aungi *səni* 'butter', E. Cush.: HEC: Gollango *šiinan-ko* 'fat', Gawwada (Dalpena) *šiinán-ko*, pl. *šiinam-aane* 'butter', S. Cush.: Qwadza *sum-* 'to milk'.

(3) Hbr. *ḥēläb*; Pho. *ḥlb* // < Sem. **ḥilb-* 'fatty tissue covering internal organ; caul' (v. SED I No. 131) or **ḥa/ilVb-* 'milk, fat' (cf. LGz 229).

(4) Pal. *trb*; Syr. *terb-*; Mnd. *tirb-*; Urm. *tarb-* // < Sem. **ṯarb-* (SED I No. 283).

(5) Urm. *šahr-* (syn.) // The only if problematic parallels I could find are either Zway *šāra* 'sediment after butter has been melted' (in LGur 584 quoted as a loan from E. Cush.: Hadiya Oromo *šāra* id.) or Muher *šärrä*, Wol. *sore*, etc. 'to feed well a sick person', Chaha, Muher, etc. *šärät* 'food', Har. *sōr* 'food offered to a group of people on a special occasion' (according to LGur. 584, a loan from E. Cush.: Oromo *sor*, Somali *sōr*, etc. 'food').[76]

(6) Qur. *šaḥm-*; Mec. *šaḥam*; Mlt. *šaḥām* // < Sem. **ŝaḥm-* (SED I No. 263). Obviously matching E. Cush.: LEC: SAM: Rendille *siḥím-e* 'butter', Somali *siḥin-* 'curds' making Afras. **ĉaḥim-*.[77]

---

[74] Some of the above forms from languages, wherein ʕ is not preserved or distinctly reflected, may alternatively belong to other Afras. roots, cf., for example, **naʔ/w/y-* 'to see' (attested in Egyp. and Chad., see EDE I 126).

[75] Should perhaps be divided into two metathetic variant roots — **sim(-an)-* and **sin(-am)-*.

[76] Cf. the idea of 'fat food' as 'good food' and of 'fat person' as 'healthy person' in MSA *ṣáyleḥ* below.

[77] Often included into Afras. **sim-an-* (cf., e. g., EDE I 192), but better fits in with Sem. **ŝaḥm-*, requiring no explanation of what -ḥ- is doing in **sim-an-*, and, if it is a hypothetic suffix (after Takács), why it is found in the medial position; as for reflexes of Afras. **s-* and **ĉ-*, they seem to have merged into *s-* in the SAM languages. A natural guess that the SAM word could be an Arabism (there are plenty of them in Somali) is contradicted by its

(7) Leb. *dihn* // < Sem. **duhn-* (v. in SED I No. 48).

(8) Gez. *ŝəbḥ*; Tna. *səbḥi*; Tgr. *šəbeḥ*; Amh. *səb*; Cha. *səwä*; Har. *säbaḥ*; Hrs. Jib. *ŝabḥ*; Mhr. *ŝabaḥ* // < Sem. **ŝabḥ-* (SED I No. 261).

(9) Amh. *mora* (syn.: 'animal fat, suet'); Sod. *mora*; Wol. *morä* // It is hard to say whether these terms are borrowed from Oromo *moora* (Gr. 291) or, vice versa, it is the latter that is an Amharism, borrowed by other Cush. and Omot. languages either directly or through Oromo mediation: C. Cush.: Aungi *morí*. E. Cush.: LEC: Arbore *moora*, etc., HEC: Qabenna, Sidamo *mōra*, etc., Dullay: Tsamay *mooru*, etc.; N. Omot.: Zaise, Yemsa *mōra*, S. Omot.: Ongota *mōra* (SLLE 6), etc. At least part of these forms may continue Afras. **marVʔ/y-* 'fat, oil': Sem.: Akk. *marû* 'to fatten' (CAD *m*1 307), Ugr. *mrʔ* 'to fatten' (DUL 570), Hbr. *mrʔ* 'to feed on the fat of the land, graze' (HALOT 630), ESA: Sab. *mrʔm* 'Mastvieh'; Arb. *mrʔ* 'trouver un aliment sain, bon'; W. Chad. Sura *mwɔɔr*, Bolewa *mor*, Barawa *miyir*, Kulere *màr*, C. Chad Tera *mar* 'oil', Nzangi *mare*, Bachama *marəy* 'fat', etc. (ADB; EDE III 431).

(10) Gaf. *buššəra* // Most likely a metathesis from **ṯarb-* (v. above).[78] Tentatively compared in EDE II 321 with an obscure Egyp. term *bšʔ*, probably 'oil' (<*bŝr*?), and several Chad. forms of the **bVs-* type meaning 'fat' and 'oil' (other quoted Chad. and C. Cush. terms of the **bVz-* type are too distant phonetically), implying a fossilized *-r* in Gaf. (cf. Mil. RE).

(11) Hrs. Mhr. *ṣáyleḥ* (both syn.) // Cf. other meanings: Mhr. *ṣáyleḥ* 'to be fat' and *haṣlēḥ* 'to improve in health, change for the better' (JM 363) < Sem. **ṣlḥ* 'to be or do well, be successful' (v. in HALOT 1026).

(12) Jib. *ʕəĉ* (syn.) // The only phonetically acceptable parallel with the same meaning that I could find is Egyp. OK *ʕḏ* 'Fett' (EG I 239), possibly < Afras. **ʕVĉ-*. One wonders if it is comparable semantically with the phonetically impeccable MSA-Arb. **ʕVŝā/īŝ-* 'bone, cartilage' (v. **bone** No. 3), if so, with the primary meaning 'bone with fat on it'.

(13) Soq. *ʕínat* // As suggested to me by L. Kogan, tentatively compared to Hbr. *ʕōnā*, probably meaning 'oil, oinment' (corresponding to Akk. *piššatu*, v. HALOT 855).

◊ Arg. *čoma* seems borrowed from Oromo id. (Gr. 85), probably via Amh. id. (C. Cush. Kemant *čoma* and Aungi *čūmí* are regarded as Amharisms by Appleyard). No terms in Bib. and Sab.

→ **Common South and West Semitic**: **ŝabḥ-* (#8) < Afras. **ĉabḥ-* 'fat, butter': W. Chad.: Diri *ŝəḇu* 'fat'; E. Cush.: Saho *subaḥ* 'clarified butter', Afar *sebaaḥ, subaḥ* 'butter'; LEC: Somali *subag* (with irregular *-g* instead of the expected *-ḥ*) 'clarified butter or animal fat', Rendille *subaḥ* 'butter; clarified animal fat', Baiso *suba* 'butter' (ADB).[79]

27 feather:

(1) Akk. *nāṣu*; Hbr. *nōṣā* (both meaning 'feathers') // < Sem. **nāṣ(y)-* (SED I No. 202).

(2) Syr. *merṭ-* // < Sem. **mrṭ* 'to pluck, pull out hair':[80] Hbr. *mrṭ* 'to pull out hair, depilate' (HAL 635), Arb. *mrṭ* 'arracher le poil' (BK 2 1092; cf. *marīṭ-* 'qui n'est pas encore garni de

---

[78] Less probably < Sem. **bi/aŝar-* 'flesh, (human) body' (SED I No. 41), since we also have Gafat *bäsärä* (with *-s-*!) 'meat', directly continuing Sem. **bi/aŝar-* id.

[79] The E. Cush. terms may in principle have been borrowed from Gez. or Amh., but the difference in vocalism would rather testify against this. That the Eth. forms could be loans from E. Cush., as asserted by some authors, is unlikely in view of the MSA cognates.

[80] The original meaning of Syr. *merṭ-* must have been something like 'hair that is easily plucked/pulled out'.

plumes (flèche)' ibid.), perhaps also Akk. *marāṭu* 'to rub, scratch' (CAD *m* 276) with a meaning shift.[81]

(3) Syr. *ʔebr-* (syn.) // < Sem. **ʔa/ibr-* 'pinion, wing' (SED I No. 1).

(4) Mnd. *guspart-*; Urm. *par-*; Hrs. *ferfayr* // In Mnd., *guspart-* also means 'crest (of bird), comb (of cock)', related in DM to Syr. *gespār-* 'pinna (piscis)' (Brock. 127); both are likely compounds consisting of **gis-* 'side'[82] and **par-* 'feather', preserved in the Urm. term (otherwise < Persian, according to Tser. 0167) and, in a geminated variant, in Hrs. One wonders whether it is possible to trace this back to something like Sem. **par(par)-?*

(5) Leb. *r-ši*; Mec. *riyša*; Mlt. *r-š* // No parallels that I know of.

(6) Tna. *kəntit*; Har. *kät* // LH 95 does not quote the Tna. form as a parallel to Har., implying that the latter is probably from Oromo *kočo* 'wing'. Unclear if the Tna. term[83] (and the Har. one with loss of *-n-*, if related) is connected to Tgr. *känta* 'to cut off (branches), to pluck off' (LH 417).[84]

(7) Cha. *zoyä* // < Gur.: Gyeto *zäwyä*, etc. (LGur 718). According to Leslau (ibid.), either "to be identified with *zorro* with palatalization of *r* to *y*" (v. Wol. below) or to be connected with Amh. *zəyy* 'kind of bird' (ibid. 719). The latter opportunity seems more attractive; Amh. 'kind of bird' must go back to 'goose' (cf. Gez. *zəy* 'goose' regarded in LGz 646 as an Amharism), very likely related, with metathesis, to Sem. **ʔa/iw(a)z- ~ *waz(z)-* 'goose'.

(8) Mhr. *ŝíf(f)*; Jib. *ŝəf ḏ-ʕeṣfɔ̄r*; Soq. *ŝéf(f)* // All in Nak.; the orig. meaning is 'hair' (the Jib. form lit. means 'hair of bird'), v. **HAIR**.

(9) Mhr. *ḳəṭfīf*; Jib. *ḳaṭaf* (both syn.) // < Sem. **ḳVṭVp-* 'pluck (leaves, fruit)' (see LGz 453). On the meaning shift see #2 and 6.

(10) Soq. *milyaṭ* (syn.) // According to LS 233, probably comparable with Arb. *līṭ-* 'peau'. I would rather compare it to Arb. *malīṭ-* 'qui n'est pas encore garni de plumes (flèche); qui n'a pas encore de poil (foetus avorté)' (BK 2 1149).[85]

◊ Tgr. *čəgär* (quoted by an informant as 'feather', but in LH 630 said to mean only 'hair, fibre') is a common Eth. loan from Cush. 'hair';[86] Amh. *laba, läboba*, Arg. *laba* are from Oromo *laboba* (LGur 373); Sod. *balle* is from E. Cush.: Oromo *balli*, Sidamo *bāla*, Somali *bāl*, etc. (ibid. 138); Wol. *zorro* is from HEC: Qabenna *zōrú-ta*, Alaba *zōrʔu-ta* (ibid. 714). No terms in Ugr., Pho., Bib., Pal., Qur., Sab., Gez. and Gaf.

→ **Common Semitic 1**: **nāṣ(y)-* (#1).
  Common Semitic 2 (debatable): **par(par)-* < Afras. **Parw-*: Brb.: Ahaggar *a-fraw* 'plume' (F. 336), Ayr *afrut* 'aile' (Aloj. 42), etc.).

28 FIRE:

(1) Akk. *išātu*; Ugr. *ʔiš-t, ʔišī-tu*; Hbr. *ʔēš*; Pho. *ʔš*; Pal. *ʔyšh, ʔäššā*; Gez. *ʔəsāt*; Tgr. *ʔəsat*; Amh. Arg. Cha. *əsat*; Gaf. *əsatä*; Sod. *äsat*; Har. *isāt* // < Sem. **ʔiš-āt-* (v. in LGz 44).

(2) Pal. *nūr* (syn.); Syr. Mnd. Urm. *nūr-*; Qur. *nār-*; Leb. Mec. Mlt. *nār-* // < Sem. **nū/ār-* (< **nawr-?* Cf. **nawir-* 'luminoso' Fron. 144) 'fire; light' (v. in HALOT 683; 696; 723; DLU 331).

---

[81] The Hbr., Arb. and Akk. forms are erroneously derived in HALOT 635 from **mrẓ* (**mrṯ̣*, in our rendering).

[82] In Syr. represented by *gess-* 'coxa, latus' ibid. 126, v. also SED I No. 97; in Mnd. **-i- > -u-* with accomodation to *-p-*.

[83] Cf. N. Omot.: Mao (Diddesa) *kwīnte* 'hair'. An accidental look-alike?

[84] On the meaning shift 'to pluck' > 'feather' v. Syr. *merṭ-* above and #9.

[85] For a somewhat paradoxical semantic connection between 'feather' and 'an arrow *not* yet furnished with feathers', cf. Arb. *marīṭ-* in #2 above. Another possibility is that the two lexemes represent variant roots with *l* vs. *r* and, as such, could have influenced one another.

[86] Cf. C. Cush. Bilin *šagar*, Qwara *ṭagur*, E. Cush. Somali *ḍagur* (LGz 550).

(3) Gez. *ḥaw, haw* (syn.); Tna. *ḥawwi* // Eth. (LGz 248; for its presumed connections with Gez. *ḥə/awāy* 'evening, the red glow of the evening sky' and further with Arb. *ʔiḥwawā* 'to become red inclining to blackness' v. ibid. 250; cf. discussion in Bulakh Dis.) with a debatable parallel in MSA **nḥy/w* 'to burn' (v. BURN No. 13). There are, however, clear cognates in Chad. (W.: Warji *ḥwa-* and, with metathesis, Sha *hwoḥ* 'to burn'; C.: Kilba *húʔù*, Mbara *hū́* 'fire'; E.: Mokilko *ʔùwwó* id., etc.). Cf. also C. Cush.: Bilin *ḥaws* 'to burn', Khamir *háu-y*, Khamta *ḥawš* 'to heat' (all trans.), considered in App. CDA 39 borrowings in Agaw from Eth. (because of the presence of *ḥ*).

(4) Hrs. *ẑawt* // MSA: Hrs. *ẑaw*, Mhr. *ẑaw* 'light' (JM 478) < Sem. **ṣ̂aw/ʔ-* (Arb.-MSA; unless an Arabism in MSA): Arb. *ḍawʔ-* 'lumière, clarté', *ḍwʔ* 'briller, luire (se dit du feu, etc.)' (BK 2 44).[87]

(5) Hrs. *ŝēwēṭ* (syn.); Mhr. *ŝīwōṭ*; Jib. *ŝoṭ*; Soq. *ŝiáṭ* (*ŝeyɔ́ṭ*) // In LS 424 compared with Syr. *swṭ* 'ê. enflammé' ('to burn, consume away' CSD 364) and Arb. variant roots *šiwāẓ-*[88] and *šiwāṭ-* 'flamme pure, sans fumée' (BK 1288), which may be cognate to the MSA terms (all < Sem. **ŝiwāṭ-*), unless the latter ones are Arabisms.

◊ Wol. *žirä* is from E. Cush.: Sidamo **žira*, Saho Afar *girā* 'fire' (LGur. 319). No term in Sab.

→ **North and West Semitic**: **ʔiš-āt-* (#1) < Afras. **ʔis-* 'fire': Brb.: Ahaggar *a-həs* 'big fire'; W. Chad. **ʔyas-* 'fire': Montol *ʔús*, Ngamo *yàsì*, Geruntum *iši*, etc., E. Chad.: Dangla *ɔ́ɔ́sɛ́* 'to make fire', Migama *ʔis* 'warm', Bidiya *ʔɔ̀s*, Birgit *ʔìssí* 'to burn' (ADB).

Common West Semitic: **nū/ār-* < Afras. **nur-* 'fire; coal, ashes': W. Chad.: Boghom *nwur-əŋ*, Kiir *ŋúr-əŋ* 'ash', C. Chad.: Gude *ŋira-ḍu* 'ash', Logone *nur* 'coal', E. Chad.: Somrai *nyùrīŋ*, Ndam *nùrē* 'coal'.

29 FISH:

(1) Akk. *nūnu*; Pal. *nūn*; Syr. Mnd. Urm. *nun-* // It is hard to decide if Sem. **nūn-* (including Hbr. *nūn* and Arb. *nūn-at-* 'un gros poisson' BK 2 1373) is reconstructible (v. in HALOT 681) or if we deal with a chain of borrowings from an unknown source > Akk. > Arm./Hbr. > Arb.[89]

(2) Ugr. *dg*; Hbr. *dāg* // DRS 216 also quotes Hbr. *dūgāh* 'pêche', Pal. *dəgōgīt-* 'barque de pêcheur' and, with a question mark, Amh. *žuž* 'paquet des poissons liés ensemble'.[90]

(3) Qur. *ḥūt-*; Mlt. *ḥūta* // Only in Arb. Possibly < **ḥaw-t-*; in this case possible Afras. parallels are: Egyp. (OK) *mḥy-t* 'fish' (<**mV-ḥVy-t*?), W. Chad.: Tsagu *ḥāti* id., **HVyw-* 'to fish': Bolewa *ʔyuw-*, Tangale *oi*, C. Chad.: Bura *yiha* 'to fish in shallow water', Mofu-Gudur *áwèt* 'fish', Gude (met.?) *tàhwá* 'type of fish' (ADB); the Tsagu and Mofu-Gudur forms can theoretically represent Arabisms, but the rest of the quoted Chad. forms certainly cannot.

(4) Mec. *samak* // Only Arb., with no parallels whatsoever, except for Gez. *samak*, which is clearly an Arabism (LGz 502).

(5) Har. *tuläm* // No parallels at all.[91]

---

[87] Note the comparison of Hbr. *mēṣīṣ* (<**ṣwṣ*, hif. 'to gaze'), Jud. *ṣwṣ* 'to look, see' and Mnd. *ṣuṣ*, *ṣiṣ* 'to look or shine (of the eyes)' with the Arb. forms in HALOT 1013–14, which is vague semantically and hardly acceptable phonetically.

[88] In fact, Soq. *ṭ* may correspond to Arb. *ẓ* < Sem. **ṭ*, but this is not the case as other MSA parallels clearly point to **-ṭ*.

[89] A common opinion, shared by my coauthor L. Kogan who insisted upon not including this root into SED II; for me, it remains baseless until the source of this presumed borrowing is presented (note, however, Uralic **ńowŋa* 'salmon').

[90] Note Indo-European **dhǵhū-* 'fish'.

[91] Littman's idea about the connection with Somali *kallun*, Afar *kullun*, with alternance *k : t*, which is regarded as possible in LH 149, does not really hold water.

(6) Hrs. *ṣayd*; Mhr. *çayd*; Jib. *çod*; Soq. *ṣode* // < Sem. **ṣyd* 'to fish, hunt'.[92]

◊ Leb. *nun* seems more likely < Arm. than inherited < Arb. ancestor. Gez. *ʕāŝā*, Tna. Tgr. *ʕasa*; Amh. Arg. Sod. Cha. *asa*; Gaf. Wol. *asä* are, according to LGz 73, from C. Cush., while Appleyard qualifies the Eth. forms as "clearly of non-Semitic, and probably Agaw origin"[93] (App. CDA 68). The term is also attested in N. Cush.: Beja *āši*, E. Cush. Saho *ʕaasa*, and Omot. **Haš-* (Kafa *haašoo*, Bworo *aso*, Anfillo *haašo*, Nao *aša*) id. (cf. Bla. Fau. 237). No term in Sab.

→ (?) **Common North and West Semitic**: **nūn-* (#1), if not borrowed.

30 FLY (v.):

(1) Akk. *naprušu* // According to AHw 740, related to Sem. **prŝ* 'to spread out' (v. in HALOT 975).

(2) Ugr. *ʕp*; Hbr. *ʕwp* // < Sem. **ʕwp* 'to fly' (v. in HALOT 800; LGz. 78), related to *ʕawp-* 'bird' (SED II No. 48).

(3) Pal. Syr. Urm. *prḥ*; Mnd. *phr* (met.) // < Sem. **prḥ* 'to fly' (v. in HALOT 966), related to **parḫ-* 'chick, brood' (SED II No. 179).

(4) Pal. *ṭws*; Mnd. *ṭus* (both syn.) // < Sem. **ṭwŝ ~ *ṭŝŝ* 'to flutter, jump': Hbr. *ṭwŝ* 'to flutter (on the ground)' (HALOT 373), Jib. *ṭŝŝ* 'to jump up' (JJ 280), etc.

(5) Qur. *ṭyr*; Leb. Mec. *ṭār*; Mlt. *tār* // As a verb only in Arb.;[94] related to Sem. **ṭayr-* 'bird; divination from birds, augury' (cf. SED II No. 235).

(6) Gez. *s/ŝarara* // Also 'to flee, leap up in the air, etc.' < Eth.: Tna. *särärä* 'to mount', Tgr. *särra* 'to jump, fly', *särerät* 'bird', Amh. *särrärä* 'jump, mount, climb', etc.[95] Cf. Mnd. *si/ara* 'flock of birds, swarm' (DM 329) and Sem. **ŝVrŝVr-*: Akk. *suššuru* (*šūšuru*) 'a dove', Arb. *šaršūr-* 'petit oiseau' (SED II No. 216).

(7) Gez. *barra*; Tgr. *bärra*; Amh. Sod. *bärrärä*; Arg. *bärrära*; Cha. *bänärä*; Harari *bärära*; Wol. *bärärä* // A root of debatable origin (cf. SED I No. 1). Contra LGz 107 and many others, not related to Sem. **prr* (quoted as *frr* ibid.).[96] Likely related to Sem. **ʔa/ibr-* 'pinion, wing', less likely borrowed from Cush. N. (Beja *bīr* RBeḍ 50), C. (Khamir *bir-*, Kemant *bärär*, Aungi *berer-*, considered in App. CDA 70 borrowings from Amh.) or E. (Saho *-ibrir*, Oromo *barar-*, Burji *burr-*, Kambatta *burri y-*, Darasa *birret-*).[97] Most likely, all the above Sem. and Cush. forms go back to Afras. **bi/arr-* 'to fly, jump', also including N. Omot. (Male *bar-an* 'to fly'), C. Chad. Mulwi *bìrì* 'to fly', Musgoy *mbìr* 'to jump', Musgu *bára* 'to fly, jump', E. Chad. Kwang *bre*, Birgit *bèrí* 'to fly' (ADB; cf. also EDE II 274).

(8) Tna. *näfärä*; Tgr. *näfra* (syn.) // Eventually, undoubtedly < **n-pr*, with a fossilized *n-* prefix (v. Mil. RE). A debatable issue is whether the N. Eth. forms should be scored with the MSA ones (v. # 9 below). After much hesitation, I choose to follow L. Kogan's advice and score them differently.

---

[92] See HALOT 1010, where no MSA terms are quoted, and LS 349, where the MSA terms for 'fish' are justly related to the verb 'fish, hunt' in other Sem.

[93] However, upon quoting Bilin *ʕasa* and Khamir *ḥaza* (together with Bilin *ʔasa*, Kemant *asa*, Aungi *asi*), he points out to "the influence of N. Eth., as neither *ʕ* nor *ḥ* occur in purely Agaw roots".

[94] Gez. *ṭayyara* 'to fly' and similar Eth. forms are regarded in LGz. 601 as Arabisms.

[95] Some of the Sem. parallels quoted in LGz 514 look questionable.

[96] Another mythetymology, extremely popular among Semitists and quite tenacious.

[97] According to Appleyard, "some of these may be borrowed from or influenced by EthSem., others may represent an original Cushitic form of this AA root", which he (erroneously, after LGz 107), equals with Afras. **pVr-* 'to fly' (another popular mythetymology).

(9) Hrs. Soq. *fer*; Mhr. *farr*, Jib. *ferr* // For broader Sem. connections v. LS 342, 341 and 107. Goes back to Afras. *pi/ar- 'to fly' (clearly different from *bi/arr- 'to fly, jump' above): Egyp. *pȝ* 'fliegen' (EG I 494), W. Chad. Hausa *fírà* 'to soar into the air', Angas *piir* 'to stretch the wings', C. Chad. Mafa *párr*, *pérr* 'bird's flight', Gude *pər*, *fər* 'to fly away (bird)'; N. Cush.: Beja *fīr* 'to fly', S. Cush.: Maʕa *púru* id. (ADB; see also EDE I 55).

(10) Mhr. (syn.) *agōnəḥ* // Jib. *génaḥ* 'wing' (an Arabism, according to JJ 77), Soq. *ganḥ* 'devant, milieu de la poitrine' (Noged dialect *génⁿaḥ* 'sternum'), Arb. *žanāḥ-* 'bras (chez l'homme); aisselle; aile (chez les oiseaux, les insectes, etc.)', *žāniḥ-at-* 'côte, surtout cette partie qui est du côté de la poitrine' (BK 1 338).[98] See SED I.

◊ No terms in Pho., Bib., Sab. and Gaf.

→ **Common West Semitic**: *ʕwp (#2) < Afras. *ʕVp- 'bird; to fly': (?) Egyp. (late) *ʕpy* 'to fly' (perhaps a Semitism); S. Omot. *HVp/f-t- 'bird': Dime *ipt*, *ift*, Ari *apti*, (?)*aft-i*, Hamer *apt-i*, *aft-i* (SED II; ADB).

31 FOOT:

(1) Akk. *šēpu*; Soq. *ŝab*, *ŝaf* // < Sem. *ŝayṗ- 'foot, sole of foot; shoe' (SED I No. 269).

(2) Ugr. *pʕn*; Pho. *pʕm*; Mhr. *fēm*; Jib. *faʕm* // < Sem. *paʕm/n- (SED I No. 207).

(3) Hbr. Bib. *rägäl*; Plm. *rgl*; Syr. *regl-*; Mnd. *ligr-* (met.); Qur. *rižl-*; Sab. *rgl* // < Sem. *rigl- (SED I No. 228), with semantically diverse but undoubted Afras. parallels (see below).

(4) Mnd. *kraia* (syn.) // < Sem. *kʷirāʕ- 'knee and shin-bone; lower leg' (SED I No. 157).

(5) Urm. *ʔaḵl-* // < Arm.: Syr. *ʔaḵlān-* 'armilla, brachiale' (Brock. 44), probably related to Arb. *wḵl* 'lever un pied en l'air en posant l'autre sur le sol' (BK 2 1591).

(6) Leb. *ʔəžər*; Gez. Tgr. *ʔəgər*; Tna. *ʔəgri*; Amh. *əgər*; Arg. *ingir*, *ägər*; Gaf. *əgʷrä*; Sod. Cha. *ägər*; Har. *ingir*; Wol. *əngər* // < Sem. (Arb.-Eth.) *ʔi(n)gi/ur- (SED I No. 7). Continues, with a fossilized prefixed ʔV-, Afras. *gVr- (see below).[99]

(7) Mec. *gadam* // More likely <*ḳadam: Arb. *ḳadam-* 'le premier pas; pas; pied' (BK 2 691; the original meaning must be 'front leg of an animal') < Sem. *ḳdm 'to precede, be in front' (v. in LGz 421; cf. *ḳudm- 'parte anteriore' Fron. 265). Less likely, though not impossible, < *gad-am-, with a fossilized suffix -m, < Sem. *gVd-at- '(part or bone of the) leg of animal': Akk. *gudgudātu* 'part of the lower leg of a quadruped', Gez. *ʔagadā* 'thighbone, shinbone, leg, large bone of the leg, shoulder of animal', etc. (v. SED I No. 71).

(8) Mlt. *siʔ* // <*siḳ: Arb. *sāḳ-* 'jambe, tibia' < Sem. *šāḳ- 'thigh, leg' (SED I No. 241).

(9) Tgr. *ʕəḳəb* (syn.) // < Sem. *ʕaḳib-, *ʕiḵb- 'heel' (SED I No. 14).

(10) Hrs. *gedel*; Mhr. *gēdel* (syn.); Jib. *gédəl* (syn.) // < Sem. *gVd(V)l- 'limb' (SED I No. 73).

(11) Soq. *suḳal* (syn.) // < Sem. *š/suḳl- 'leg, thigh; elbow' (SED I No. 242); derivation, with a fossilized prefixed -l, from Sem. *šāḵ- 'thigh, leg' (No. 8, above) is possible.

→ **Common Semitic 1**: *ŝayṗ- (#1) < Afras. *ĉayṗ- 'foot, sole of foot; shoe': Egyp. (Gr) *šp* 'hoof'; N. Cush.: Beja *šib* 'to shoe', *šab* 'to be shod, put on one's footgear'; C. Cush.: *šanp/b- 'foot, heel': Bilin *šaanfi*, Qwara *šaanpaa*, Dembea *šanfa*, Qemant *šaambaa*, S. Cush.: Asa *išiba* 'sandal' (ADB).

Common Semitic 2: *paʕm/n- (#2) < Afras. *Paʕun/m- 'leg, foot': W. Chad.: Fyer *fŭŋ* (< *funH-), C. Chad.: Fali Kiria *pùnuʔ* 'thigh', Zime-Batna *fun* 'buttocks', E. Chad.: Soko-

---

[98] This root is attested only in MSA and Arabic, which always causes suspicions of an Arabism in MSA; this is hardly the case, however, since the root has reliable Afras. cognates and the primary meaning seems to have been 'wing': Egyp. (Pyr.) *ḏnḥ* (<*gnḥ) 'wing'; (?) C. Chad.: Mbara *gàŋ-làŋ* 'wing' (the second element is not clear); E. Cush.: HEC: Kambatta *gonna-ta* id.; N. Omot.: Dizi (Maji) *gaŋg* 'to fly'.

[99] Its widely accepted and much-discussed cognation with *rigl- is yet another mythetymology among Semitists.

ro *offen*, *òpán* (<*ʾVfyan-*) 'foot, leg'; E. Cush.: Oromo *fana* 'trace' (n.), S. Cush.: Qwadza *paʾam-uko* 'foot' (ADB).

Common West Semitic 1: **rigl-* (#3) < Afras. **riga/ul-* 'limb, leg': Brb. Zenaga *tə-rgəl*, Ghadames *ta-rză/el-t* 'plume', *tū-rəgl-a* 'gros pédoncule du régime de dattes', Ayr *a-rgul* 'penis', Ahaggar *â-r̄gal* 'queue'; W. Chad. Tangale *argil* 'inner side of upper thigh' (Stolb 2005 233), etc. Cf. also Egyp. (Pyr.) *ʕg.t* 'hoof (of cow and ass)', (NK) *ʕgʔ.t* 'claw (of lion and bird)' (dissimilation <**ʔgʔ* <**rgl*?) (ADB).

Common West Semitic 2: **ʔi(n)gi/ur-* (#6) < Afras. **(ʔi-n)gur-* 'leg, knee': W. Chad. Warji *ŋgarái*, Miya *âgár* 'leg', Mbara *mù-gùrí*, E. Chad. Kera *gə̀gə̀r* 'knee', Sokoro *gorun-gorun-du* 'foot'; S. Cush. **gurun-guda* 'knee': Iraqw, Gorowa *gurūngura*, etc. (cf. K-M 122), N. Omot.: Mocha *gurāto*, Gimirra (She) *gurät* 'knee' (LMč 33; comparison with Amh. *gulbät*, etc. is wrong) (ADB).[100]

32 FULL:

(1) Akk. *malû*; Ugr. *mlʔ*; Hbr. *malǟʔ*; Pal. *mly*; Syr. *məlē*; Mnd. Urm. *mily-*; Qur. *malʔān-*; Leb. *məlīn*; Mec. *malʔān*; Mlt. *memli*; Gez. *məlu*ʾ; Tna. *mulu*ʾ; Tgr. *malu*ʾ; Amh. *məlu*; Arg. *muli*; Sod. *mulä*; Cha. *mura*; Har. *mullu*ʾ; Wol. *mulli*; Mhr. *mīla*ʾ; Jib. *miẑi*ʾ; Soq. *mili* // < Sem. **ml*ʾ 'to fill, be full' (v. in LGz 342).

◊ No terms in Pho., Bib., Sab., Gaf. and Hrs.

→ **Proto-Semitic**: **ml*ʾ < Afras. **ml*ʾ ' be full, filled': Brb.: Ahaggar *amâli* 'tout, entier', Ayr *măllu* 'ê. rempli entièrement, pleinement', *məluməl* 'ê. complètement rempli'; W. Chad.: Hausa *màlā́là* 'to flow out, into; pervade entirely', *màlàlà* 'abundantly', (?) C. Chad.: Mada *məla-kiya* 'full moon' (*kiya* 'moon'); S. Cush.: Iraqw *milalāʕ* 'to fill to the brim' (with an ʕ of unclear origin); (?) N. Omot.: Wolamo *mūliya* 'totality, wholeness (?)' (cf. *ḳamma mūliya* 'the whole night') (EDE III 413; ADB).

33 GIVE:

(1) Akk. *nadānu* (other verbal forms include *tadānu* and *idinu*) // The comparison, as an *n*-prefixed verb, with Arb. *dyn* 'prêter; rétribuer' (BK 1 757),[101] corroborated by Egyp. (Pyr.) *wdn* 'opfern' (EG I 391) and W. Chad.: Angas *tūn* (*t-* can reflect **d-*) 'tuwo as an offering'[102] seems more tenable than the widely quoted equation with Sem. **n/ytn* (so in AHw 701; HALOT 733; DLU 543; EDE I 241; cf. a detailed discussion in EDE III 764),[103] where Akk. *d* vs. Sem. **t* is impossible to explain other than by some kind of contamination of the two roots.

(2) Ugr. *ytn*; Hbr. Bib. Pal. *ntn*; Pho. *ytn* (*n-tn*) // < Sem. **ytn ~ *ntn* (v. in HALOT 733).

(3) Bib. Pal. *yhb* (both syn.); Syr. *y(h)b*; Mnd. *ahb* (syn.); Urm. *yhb*; Sab. Gez. *whb*; Tna. *habä*; Tgr. *haba*; Arg. *hawa*; Gaf. *wabä*; Sod. *abä*; Wol. *wabä* // < Sem. **whb* (LGz 609).

---

[100] Very likely related to North African Afras. **(nV-)g^war-* 'hand; holding, catching': Egyp. (Pyr.) *ḏʔ.t, ḏr.t* 'hand' (cf. *nḏry* 'to catch', *nḏr* 'to seize'; W. Chad.: Bolewa *gòru* 'to snatch', *ngàru* 'to pinch and hold tight', Tangale *kɔɔrɛ* 'to hold tightly', C. Chad.: Musgoy *ŋgər* 'hand', Gude *ngərə* 'to pick up, lift' , Musgu *ŋgraŋge* 'arm' (ADB; cf. EDE I, 321). The eventual kinship of the two roots, one meaning 'leg, knee', the other meaning 'hand', can be accounted for by assuming a common Pre-Proto-Afrasian ancestor lexeme meaning 'limb, leg (of animal)', with an eventual bifurcation.

[101] Cf. also MSA: Mhr. *adyēn* 'to lend (money, supplies), to give credit' JM 78, Jib. *edyín* id. JJ 44, Soq. **šédyen* caus.-refl. 's'emprunter' LS 127 (unless all from Arb.)

[102] Most likely continuing N. Afras. **dVw/yVn-* 'offering' (note 'to offer a sacrifice' as one of the meanings of Akk. *nadānu* CAD *n* 42).

[103] Apparently another mythetymology, though more subtle in this case.

(4)   Qur. *ʕṯy* IV; Leb. *ʔaʕṭa*; Mec. *ʔaʕṭā*; Mlt. *ta* // Only Arb.

(5)   Amh. *sättä*; Har. *säṭa* // < Sem. **(y/w)šṭw* 'to hold out, give/take' (v. in LGz 520).

(6)   Hrs. *wezōm*; Mhr. *wəzōm*; Jib. *əzōm*; Soq. *ʔézom* // Also 'to lend'; related to Arb. *wzm* 'payer, acquitter (la dette)', *wazima* 'éprouver quelques pertes dans son avoir' (BK 2 1529). Likely cognates are: W. Chad. Hausa *zā́mā* 'to defraud a person of his proper share', C. Chad. Logone *zum* 'plündern', Buduma *ham* id., Gude *zəmə* 'to cheat, neglect to pay debt; extract payment from so.' (ADB).

(7)   Soq. *mnḥ* (syn.) // < Sem. **mnḥ*: Ugr. *mnḥ* 'entregar', Hbr. *minḥā* 'gift, offering', Arb. *mnḥ* 'donner, offrir; donner à quelqu'un l'usufruit des bestiaux' (BK 2 1156), Gez. (metathetic) *maḥana* 'to pay hommage, make a gift', etc. (v. in LS 246; DLU 282–283; HALOT 601; see also EDE III 306–7 for possible Cush. and Chad. parallels)[104].

→   **Common West Semitic 1**: **whb* (#3) < Afras. **wahab-* ~ **hVwab-* 'to bring, give, take': Brb. **Hubay-*: Ghadames *əbbi*, Siwa *əbba* 'to drive, bring', Ahaggar, Taneslemt *hub-ət* 'to drag'; W. Chad.: Sura *hwɔ́p*, Ankwe *waap* 'to borrow', Pero *wábà* 'offering'; E. Cush.: Saho-Afar *ab-it-* 'to take for oneself', HEC: Sidamo *ab-*, Hadiya *ēb-*, Alaba *ib-* 'to bring', S. Cush.: Alagwa *hub-* 'to bring' (ADB; EDE I 72–3)[105].

Common West Semitic 2: **ytn* ~ **ntn* (#2).

34 GOOD:

(1)   Akk. *ṭābu*; Pal. *ṭb*; Hbr. *ṭōb*; Bib. *ṭāb*; Syr. Mnd. *ṭāb-*; Qur. *ṭayyib-*; Mec. *ṭayyib*; Mlt. *tayyip* // < Sem. **ṭayVb-* (v. in DLU 479; HALOT 370).

(2)   Akk. *damḳ-* (syn.) // < Sem. **dmḳ* 'to be pleasing, good, beautiful' (v. in DRS 276; LGz 135).

(3)   Ugr. Pho. *nʕm* // < Sem. **nʕm* 'to be pleasant' (v. in HALOT 705) < Afras. **nʕm* 'to be sweet (of honey)': W. Chad.: Dera *nəmnəm* (redupl.) 'sweet', C. Chad. **nVm-* 'sweet, honey' (CLR II 549); S. Cush.: Qwadza *naʔam-uko* 'honey-comb', Maʔa *naʔá* 'honey'.[106]

(4)   Syr. *šappīr-* (syn.); Urm. *šapīr-* // < Sem. **špr* 'to be beautiful, clean; to shine' (v. in HALOT 1635).

(5)   Qur. *ḥasan-* (syn.) // Translated as 'beau, joli; bon, excellent' in BK 1 428. Obviously connected with Tgr. *ḥasna* 'to talk and do good' (LH 73), Mhr. *ḥássən* 'to improve in health' (JM 189; marked as Arabism), Jib. *aḥsín* 'to be kind to so.', *aḥtsín* 'to improve', *s̃ḥɛsín* 'to think so. or st. good' (JJ 116); all these forms, however, may well be Arabisms. On one hand, likely related to Hbr. *ḥăsīn* 'strong' (HALOT 338), Syr. *ḥasīn-* 'firmus, robustus', *ḥsn* (etpe.) 'superatus est' (Brock. 248), all < Sem. **ḥsn*. On the other hand, cf. Arb. *ḥisnat-* 'salaire, prix du travail' (BK 1 428) cognate to Jud. *ḥsn* (Itpa.) 'to be fully compensated' (Ja. 489) and E. Jib. Mhr. *ḥəsənēt* 'heavenly reward' (JM 189; an Arabism?), also < **ḥsn*. The question is whether these are two homonymous roots or just one, with polysemy.

---

[104] Tgr. *männäḥa* 'to let a cow (as a loan) in usufruct' (LH 127, compared in HALOT 601 without any comments), is certainly an Arabism.

[105] Contrary to the established opinion (e.g. in EDE I 72–3), Egyp. (Pyr) *hꜣb* 'to send (a letter or message inter alia), to write a letter', (MK-NK) 'letter, message' is not related, since ꜣ renders here **-r-*, i. e. the implied Egyptian form would be **hVr(V)b-*. This is demonstrated by forms in languages that have borrowed the Egyp. term in the meaning 'to write', namely Chad.: Hausa *rubutā*, Buduma *rebōde*, Afade *ohárbotù* (the latter word perfectly conveys the consonant root composition of the Egyp. word) and Brb: Lybian (East Numidian) *tt-rb, t-rb-thn*, Ghadames *ūrəb*, etc. < Brb. **Harəb*, a conspicuous case of **b̠* (> *b̠* in Ghadames and Audjila) < **b* with a laryngeal in the same root (for details, see Mil. Tuar. 200).

[106] Cf. EDE I 261, comparing the Sem. and S. Cush. forms with Egyp. *nḏm* 'sweet, pleasant', where *ḏ* is impossible to justify, since the correspondence Egyp. *ḏ* ~ Sem. *ʕ* does not exist.

(6) Leb. *mləḥ* (*mnəḥ*) // Arb. *mlḥ* 'ê. beau ou bon' (BK 2 1144), related to Ugr. *mlḥ* 'hermosura' (DLU 274; quoted with a question mark). The meaning 'good' is presumably derived from 'salt, salty' (< Sem. *milḥ-* 'salt', v. in LGz 343; this semantic shift is attested in several other roots, cf. Bulakh 2005), cf. Arb. *milḥ-* 'sel; ľesprit, le piquant', *mlḥ* 'saler; ê. salé' (BK ibid.). Cf. also Gez. *malḥ*, *məlḥ*, *məlḥā* 'salt, taste, savor, common sense' (LGz 343) and the comments by Leslau on *malḥa* 'to do, work': "possibly 'do good work', salt being the symbol of good deeds" (ibid.).[107]

(7) Sab. *ṣdḳ* // The meaning 'good' is debatable ('right; justice; justification; truth; *that which is good, proper, satisfactory*' SD 141) < Sem. *ṣdḳ* 'to be just, true' (v. in HALOT 1003; LGz 548).

(8) Gez. *ŝannāy*; Tgr. *sänni* // Common Eth.[108] External parallels, adduced in LGz 532, are not very convincing, except for Mhr. *meŝnâ* 'fitness, efficiency', quoted after Bittner, but having a different meaning in JM.

(9) Tna. *ṣəbbuḳ* // To compare with Arb. *sbḳ* 'devancer, arriver le premier', *ʔasbaḳ-* 'qui devance les autres et arrive le premier; supérieur, excellent' (BK 1 1046); perhaps an Arabism.

(10) Amh. *ṭəru* // Several etymological hypotheses may be proposed in the absence of direct parallels. Either we should derive it from *ṭärra* 'to be pure, clear' < Eth. *ṣry* id. (v. in LGz 564), or identify it, as a metathesis-enhanced *ṭrw*, with S. Eth. *ṭwr* 'to do things well, arrange well' (v. in LGur 637), or with Sod. *ṭiräňňe* 'to be strong, powerful, courageous' and similar S. Eth. forms (v. in LGur 631–2). In any case, no clear parallels outside Eth.

(11) Arg. *dəmma* // Though no etymology is offered in LArg 198, likely related to Mod. Eth. *dämam* 'attractive, pretty', derived in LGur 209 from *däm* 'blood'. The meaning shift 'blood' > 'good, attractive' is not self-evident and needs more data to be convincing. An alternative semantic shift, although also debatable, is 'attractive' < 'red'[109] (Amh. *addäma*

---

[107] Otherwise, to be compared with Brb. *-mallay* 'good, beautiful' (Qabyle *a-mellay* 'good, merciful', Ayr *mol-ăn* 'good', etc.); C. Chad.: Kotoko *màlà* 'sweet, pleasant'; N. Cush.: Bilin *milmil-í* 'beautiful, graceful' (see EDE III 242), in which case we are setting up a different etymology, apparently not connected with 'salt'; the quoted Brb., N. Cush. and C. Chad. forms are not expected to reflect Sem. *-ḥ* and are thus comparable with the Sem. root; if, however, they are related to Egyp. (late NK) *mn.t* (if <*ml-t*) 'happy state of being' (ibid. 241), bearing no traces of *ḥ*, the comparison with Arb. (and possibly Ugr.) *mlḥ* 'to be good' should be disregarded, which again returns us to the 'salt' version. Another much quoted parallel with Arb. *mlḥ* is Egyp. (Pyr.) *mnḫ* 'richtig, trefflich' (EG II 84), *s-mnḫ* 'gut machen' (ibid. IV 136), possibly <*mlḫ* (cf. EDE III 313–16; note, however, Sem. variant forms with *-n-*: the Leb. variant root *mnəḥ*, Syrian Arabic *mnīḥ* 'nice' and MSA: Mhr. *menaḥ* 'nice', Soq. *ménaḥ* 'beau'); the latter parallel is much less tenable, since Egyp. *ḫ* vs. Sem. *ḥ* is not regular (cf. discussion in EDE ibid.).

[108] Tna. *sännay* is rare, according to my informants.

[109] Curiously, the two roots with the same consonantal composition and obviously associable meanings 'red' and 'blood' make up two different lexemes on the Proto-Afrasian level and thus should be qualified as homonymous on that level. Cf. Afras. *(ʔa-)dVm-* ~ *diʔm-* 'red': Sem.: Akkadian (OAkk. on) *adamu* (*adammu*, *adumu*) 'a red garment' (CAD *a*1 95) (cf. metathetic Standard Bab.) *daʔmu* 'dark-colored, dark-red' ibid. *d* 74), Ugr. *ʔadmānu* 'red (earth)' (Huhn., 104), Hbr. *ʔādōm* 'reddish(-brown), of blood, grape-juice, lentils, cow, horse, skin' [HALOT 15] (cf. also its reduplicated stem variant *ʔădamdām* 'right red, reddish' ibid.), Arb. *ʔudmat-* 'red color' (DAF 64), *dmm* 'to paint red' (BK 1 728), Gez. *ʔadama* 'be red', *ʔaddāmāwi* 'red' (LGz 8), Amh. *addämä* 'to be blood-red' ibid. (otherwise < 'blood'), *dama* 'brown (mule, horse), reddish' (LGur 207) Gurage (all dialects) *dama* 'brown (mule, horse), reddish' (ibid.), Masqan *dämyä* 'red (maize)', Cha., Muher, etc. *dämyät* 'red, reddish' (ibid. 210); Egyp. (OK) *îdmy* 'red cloth'; Brb.: Tashelhit *adəmmani*, Tamazight *adəmman* 'brown, bronze coloured', Qabyle *ddəmdəm* 'violet'; C. Cush.: Aungi *dámmá*, Kunfäl *demé* (App. CDA 114), E. Cush.: Saho *duma*, Oromo *dímā*, Konso *tīm-*, Sidamo *duuʔmo* 'red', Darasa *diimma* 'to become red', S. Cush.: Qwadza *dimayi-* 'red'; Omot.: Kafa *damme* 'red', Ongota *damaʔtə* 'yellow' (ADB).

'to be blood-red', *dama* 'dark-red', Chaha *dämyät* 'red', Harari *dāma* 'brown, dark-skinned', etc. < Sem. **ʔadam- ~ *dVʔVm-* 'red').[110]

(12) Gaf. *gunnä* // Likely < Eth. **gnn* 'to become important, abundant, numerous, strong; to exceed' (LGz 198), Tna. *gänänä* 'to be lucky, fortunate' (Kane T 2316), related to Arb. *ǯnn* 'grandir et se développer dans une riche végétation, être abondant et touffu (se dit des plantes, des herbes)' (BK 1 331–2); cf. DRS 147.

(13) Cha. *wäḵe* // Controversially commented upon in LGur 650 as "perhaps a phonetic variant of *wäge*", in its turn commented upon as "perhaps a phonetic variant of *wäḵe...* coming from **wäke*" (ibid. 646). Perhaps related to Gez. *wākaya* 'to shine, be brilliant', etc. (LGz 612), reasonably compared ibid. with Akk. *akukūtu* 'red glow in the sky' (also 'flame, blaze' CAD *a*1 285).[111] Cf. interesting, though isolated, parallels in Chad.: W.: Hausa *kyâu* 'goodness, beauty' (Abr. Hs. 602) and C.: Gisiga *kuwi* 'good', *kuko* 'beautiful' (Sk. Hs. 164) and E. Cush.: Yaaku *-ɛɛko* 'good' (Hei. Ya. 126).

(14) Har. *ṭoññam* // < *ṭōña* 'to exceed, excel' < Eth. **ṣnʕ* 'to be strong' < Sem. **ṣnʕ* 'to make, act skillfully' (v. in LGz 559).

(15) Wol. *bēzzä* // According to LGur 168, from Kambatta *bīzza* 'generous'. Otherwise from Common Eth. and Sem. **bzḫ* 'to abound, be abundant, become more', etc. (v. in LGur 168).

(16) Jib. *fəḳŝ-ún* // Placed in JJ 56 under the same root as *fḳɔ̂ŝ* 'to tap st. until it breaks (as., e.g., an egg)'. If this similarity is not the result of homonymy, but represents a very specific meaning shift, then the forms are related to Mhr. *fəḳáwŝ* id. (JM 92) and Arb. *fḳš* 'casser avec la main (un oeuf)' (BK 2 621; cf. ibid. variant roots *fḳs* and *fḳṣ* with close meanings).[112]

(17) Soq. *díye* // The only possible match that I could find is Hbr. **day* 'sufficiency, what is required, enough' (HALOT 219), but there are several tenable parallels in other Afras. with the meaning 'good': E. Chad. Tumak *ed*; N. Cush. Beja *day*, E. Cush. Bayso *ka-iida* (<**yid-*), Sidamo *aada*; N. Omot. Shinasha *doʔa*, Kafa *deʔō* making Afras. **dVy/ʔ- ~ *yVd-* 'good' (Mil. 2004 317–18; ADB).

◊ Sod. *fäyya* is < E. Cush.: Oromo *fayya* 'to be in good health', Sidamo *fayyi* 'to feel better', etc. (LGur 252); Wol. *bēzzä* is, according to LGur 168, from Kambatta *bīzza* 'generous'[113]; Hrs. *ged*, Mhr. *gīd* must be borrowed from Arb. *ǯayyid-* 'excellent, bon' (BK 1 351).

→ **Common North and West Semitic**: **ṭayVb-*.

Literature

Abr. Hs. — ABRAHAM, R., 1965. *Dictionary of the Hausa Language.* London.

ADB — *Afrasian Data Base* ((*http://starling.rinet.ru* and *http://ehl.santafe.edu*).

AHw — SODEN, W. VON., 1965–1981. *Akkadisches Handwörterbuch.* Wiesbaden.

Aloj. — ALOJALY, Gh., 1980. *Lexique touareg-français.* Copenhague.

---

[110] For the "isosemantic row" demonstrating the same meaning shift, cf. C. Cush. Kemant *säraɣ*, Qwara *sära*, both meaning 'red' and 'beautiful' (App. CDA 30), and Russian *красный* 'red' and the derived form *красивый* 'beautiful' .

[111] And unconvincingly with Soq. *ʔughe* 'to become reddish', since Soq. *g* does not correspond to Sem. **k*.

[112] One wonders if the MSA verb could be an Arabism. Similarity of a very specific meaning testifies in favor of borrowing; verbs, however, are not prone to borrowing in Semitic, and the MSA languages, full of nouns borrowed from Arabic, are not an exception in this matter.

[113] Though an origin from Common Eth. and Sem. **bzḫ* 'to abound, be abundant, become more', etc. (v. in LGur 168), cannot be entirely ruled out.

App CDA — APPLEYARD, D. A., 2006. Comparative Dictionary of the Agaw Languages. *Kuschitische Sprachstudien / Cushitic Language Studies*, Band 24. Köln.

Baet. — BAETEMAN, J., 1929. *Dictionnaire amarigna — français suivi d'un vocabulaire français- amarigna*. Dire-Daoua (Ethiopie).

Barg. — BARGERY, G., 1934. *A Hausa-English dictionary and English-Hausa vocabulary*. London.

Belot — BELOT, J. B., 1929. *Vocabulaire arabe-français à l'usage des étudiants*. Beyrouth.

Biella — BIELLA, J.C. *Dictionary of Old South Arabic*. Chico, 1982.

BK — BIBERSTEIN-KAZIMIRSKI, A. DE., 1860. *Dictionnaire arabe-français*. Paris.

Bla. Fau. — BLAŽEK, V., 2003. Fauna in Beja Lexicon. *Studia Semitica* (FS A. Militarev). Moscow, 230–294.

Bnd Om. — BENDER, M.L., 2003. *Omotic lexicon and Phonology*. Carbondale.

Brock. — BROCKELMANN, C., 1928. *Lexicon Syriacum*. Halle.

Bulakh 2003 — BULAKH, M., 2003. Etymological Notes on the Akkadian Colour Terms. *Studia Semitica* (FS A. Militarev). Moscow, 3–17.

Bulakh 2004 — BULAKH, M., 2004. Color Terms of Modern South Arabian Languages: A Diachronic Approach. *Babel und Bibel* I. Moscow, 269–82.

Bulakh 2005 — BULAKH, M., 2005. Semantic notes in the lexical field of taste in Geez. *Varia Aethiopica. In Memory of Sevir B. Chernetsov (1943–2005)*. SPb, 325–353.

Bulakh Dis. — BULAKH, M., 2005. *Цветообозначение семитских языков в этимологическом аспекте (Color nomination in Semitic in the etymological aspect)*. Doctorate dissertation. Russian State University in the Humanities, Moscow.

CAD — OPPENHEIM, L., E. REINER & M.T. ROTH (ed.), 1956–. *The Assyrian Dictionary of the Oriental Institute, the University of Chicago*. Chicago.

CSD — A Compendious Syriac Dictionary, ed. J. Payne SMITH. Oxford 1957

CLR — JUNGRAITHMAYR, H. & D. IBRISZIMOW, 1994. *Chadic Lexical Roots* I–II. Berlin.

DLU — DEL OLMO LETE, G. & J. SANMARTÍN, 1996–2000. *Diccionario de la lengua ugarítica* I–II. Barcelona.

DM — DROWER, E.S. & R. MACUCH, 1963. *A Mandaic Dictionary*. Oxford.

DRS — COHEN, D., 1970–. *Dictionnaire des racines sémitiques ou attestées dans les langues sémitiques*. La Haye.

DUL — OLMO LETE, G. & J. SANMARTÍN. A Dictionary of the Ugaritic Language in the Alphabetic Tradition. 2003. Leiden-Boston.

Eb. — Ebert, K., 1976. *Sprache und Tradition der Kera (Tschad)*. Teil II. *Lexikon*. Berlin.

EDE I — Takács, G., 1999. *Etymological Dictionary of Egyptian*, Volume One: A Phonological Introduction. Leiden-Boston- Köln.

EDE II — TAKÁCS, G., 2001. *Etymological Dictionary of Egyptian*, Volume Two: *b-, p-, f-*. Leiden-Boston-Köln.

EDE III — TAKÁCS, G., 2008, *Etymological Dictionary of Egyptian*, volume Three. Leiden-Boston.

EEN — EHRET, C., E. ELDERKIN & D. NURSE, 1989. Dahalo lexis and its sources. *Afrikanistische Arbeitspapiere* 18, 1–49.

EG — ERMAN, A. & H. GRAPOW, 1957–71. *Wörterbuch der aegyptischen Sprache*, I–VII. Berlin.

F. — FOUCAULD, le père C. de., 1951–1952. *Dictionnaire touareg-français* 1–4. Paris.

Fron. — FRONZAROLI, P., 1965. *Studi sul lessico comune semitico*. III. I fenomini naturali. Rendiconti delle Sedute dell'Accademia Nazionale dei Lincei. Classe di scienze morali, storiche e filologiche. Ser. VIII, vol. XX, fasc. 3–4, pp. 135–150.

Gr. — GRAGG, G.. 1982. *Oromo Dictionary*. East Lansing.

HALOT — KOEHLER, L. & W. BAUMGARTNER, 1994–1996, 1999–2000. *The Hebrew and Aramaic Lexicon of the Old Testament* I–III. Leiden, New York & Köln. IV–V. Leiden, Boston & Köln.

Hei. Ya. — HEINE, B., 1975. Notes on the Yaaku Language (Kenia). *Afrika und Übersee* 58/2. 119–38.

HJ — HOFTIJZER, J. and K. JONGELING, 1995. *Dictionary of the North-West Semitic Inscriptions*. Leiden — New York — Köln.

HRSC — EHRET, C., 1980. *The Historical Reconstruction of Southern Cushitic Phonology and Vocabulary*. Berlin.

HSED — OREL, V. and O. STOLBOVA. 1995. *Hamito-Semitic Etymological Dictionary. Materials for a Reconstruction*. Leiden-New York-Köln.

Huds. — HUDSON, G., 1989. *Highland East Cushitic Dictionary*. Hamburg.

Huehn. — HUEHNERGARD, J. *Ugaritic Vocabulary in Syllabic Transcription*. Atlanta, 1987.

Ja. — JASTROW, M., 1996. *A Dictionary of the Targumim, the Talmud Babli and Yerushalmi, and the Midrashic Literature*. New York.

Jahn — JAHN, A., 1902. *Die Mehri-Sprache in Südarabien. Texte und Wörterbuch.* Wien.

JH — JOHNSTONE, T.M., 1977. *Ḥarsūsi Lexicon.* New York — Toronto.

JJ — JOHNSTONE, T.M., 1981. *Jibbāli Lexicon.* New York (NY).

JM — JOHNSTONE, T.M., 1987. *Mehri Lexicon.* London.

Kane A — KANE, T.L., 1990. *Amharic-English Dictionary.* Wiesbaden.

Kane T — KANE, T.L., 2000. *Tigrinya-English Dictionary.* Vol. I–II. Springfield.

Kauf. — KAUFMAN, S. A., 1974. *The Akkadian Influences on Aramaic.* Chicago-London.

K-M — KIESSLING, R. and M. MOUS, 2003. *The Lexical Reconstruction of West-Rift Southern Cushitic.* Köln.

Kog. DD — KOGAN, L., 2006. On Proto-Semitic Deverbal Derivation. Paper read at II Workshop on Comparative Semitic (Sitges, Spain)

Kog. Eth. — KOGAN, L., 2005. Common Origin of Ethiopian Semitic: the Lexical Dimension. *Scrinium. T. 1: Varia Aethiopica. In Memory of Sevir B. Chernetsov (1943–2005).*

Kog. Ug. — KOGAN, L., 2005. *Between Akkadian and Cannanite: Lexical Evidence and the Genealogical Position of Ugaritic* (manuscript).

Kossm. — KOSSMANN, M., 1999. *Essai sur la phonologie du proto-berbère.* Köln.

Lan. — LANFRY, J., 1973. *Ghadamès.* Vol. II. *Glossaire (Parler des Ayt Waziten).* Alger.

LArg — LESLAU, W., 1997. *Ethiopic Documents: Argobba.* Wiesbaden.

Lao. — LAOUST, E., 1932. *Siwa: son parler.* Paris.

LGaf — LESLAU, W., 1956. *Étude descriptive et comparative du Gafat (Éthiopien méridional).* Paris.

LGur — LESLAU, W., 1979. *Etymological Dictionary of Gurage (Ethiopic).* Vol. III. Wiesbaden.

LGz — LESLAU, W., 1987. *Comparative Dictionary of Geʕez (Classical Ethiopic).* Wiesbaden.

LH — LITTMANN, E. and M. HÖFNER, 1956. *Wörterbuch der Tigre-Sprache. Tigre-deutsch-englisch.* Wiesbaden.

LHar. — LESLAU, W., 1963. *Etymological Dictionary of Harari.* Berkeley and Los Angeles.

Lmb-Sot — LAMBERTI, M. and R. SOTTILE, 1997. The Wolaytta Language. *Studia Linguarum Africae Orientalis*, Bd. 6. Köln.

LMč — LESLAU, W., 1959. *A Dictionary of Moča (Southwestern Ethiopia).* Berkeley & Los Angeles.

LS — LESLAU, W., 1938. *Lexique Soqoṭri (Sudarabique moderne) avec comparaisons et explications étymologiques.* Paris.

Maizel — МАЙЗЕЛЬ, С. С., 1983. *Пути развития корневого фонда семитских языков.* Москва. [MAIZEL, S. S., 1983. *Ways of Root Derivation in Semitic.* Moscow.]

MCB — LAOUST, E., 1920. *Mots et choses berbères.* Paris.

Mil. Farm. — MILITAREV, A., 2002. The Prehistory of a Dispersal: the Proto-Afrasian (Afroasiatic) Farming Lexicon. *Examining the Farming/Language Dispersal Hypothesis,* eds. P. Bellwood & C.Renfrew. McDonald Institute Monographs. Cambridge, 135–50.

Mil. RE — MILITAREV, A., 2005. Root extension and root formation in Semitic and Afrasian. *Proceedings of the Barcelona Symposium on comparative Semitic,* 19–20/11/2004, *Aula Orientalis* 23/1–2, p. 83–130.

Mil. 2000 — MILITAREV, A., 2000. Towards the chronology of Afrasian (Afroasiatic) and its daughter families. *Time Depth in Historical Linguistics. Volume 1.* Ed. by C.Renfrew, A.McMahon & L.Trask. The McDonald Institute for Archaeological Research. Cambridge, 267–307.

Mil. 2004 — MILITAREV, A., 2004. Another Step towards the Chronology of Afrasian (I). *Orientalia et Classica. Papers of the Institute of Oriental and Classical Studies. Issue V. Babel and Bibel 1. Ancient Near East, Old Testament and Semitic Studies.* Moscow, 282–333.

Mil. 2007 — MILITAREV, A., 2007. Toward a Complete Etymology-Based Hundred Word List of Semitic. Irems 1–34 (First Third). *Proceedings of the 7th International Semito-Hamitic Congress (Berlin, Sept.13–15, 2004).* Aachen, p. 71–102.

Mil. 2008 — MILITAREV, A., 2008. Toward a Complete Etymology-based One Hundred Wordlist of Semitic: Items 34–66 (Second Third). *FS H. Jungraithmayr and A. Dolgopolsky.* Ed. by G. Takacs, p. 194–222.

Nak. — NAKANO, A. 1986. *Comparative Vocabulary of Southern Arabic.* Tokyo.

Pen. — PENRICE, J., 1873. *A Dictionary and Glossary of the Koran.* London.

RBeḍ — REINISCH, L., 1895. *Wörterbuch der Beḍauye-Sprache.* Wien.

Ricks Qat. — RICKS, S. D., 1982. A Lexicon of Epigraphic Qatabanian. PhD dissertation. Berkeley.

Sarg. — САРГИС (архимандрит), 1909. *Русско-сирский лексиконъ.* Урмия. (SARGIS, archimandrite, 1909. *Russian-Syriac Lexicon.* Urmia.)

Sat. — Satterthweit, A. C., 1960. Rate of Morphemic Decay in Meccan Arabic. *International Journal of American Linguistics*. 26/3, 254–261.

SD — Beeston, A. F. L., M. A. Ghul, W. W. Müller & J. Ryckmans, 1982. *Sabaic Dictionary (English-French-Arabic)*. Louvain-la-Neuve.

SED I — Militarev, A. & L. Kogan, 2000. *Semitic Etymogical Dictionary.* vol. 1: *Anatomy of Man and Animals*. Münster.

SED II — Militarev, A. & L. Kogan, 2005. *Semitic Etymogical Dictionary.* Vol. II: *Animal Names*. Münster.

Sk. Hs. — Skinner, N., 1996. Hausa Comparative Dictionary. Köln.

SLLE — D. Kusia and R. Siebert, 1994. Wordlists of Arbore (Irbore), Birayle (Ongota), Tsamai (Tsamaho). *Survey of Little-known Languages of Ethiopia. Linguistic Report No. 20*. Addis-Ababa. 1–12.

Sok. — Sokoloff, M., 1990. *A Dictionary of Jewish Palestinian Aramaic*. Jerusalem.

Star. — Starostin, S., 2000. Comparative-historical linguistics and lexicostatistics, in *Time Depth in Historical Linguistics*, vol. 1, eds. C. Renfrew, A. McMahon & L. Trask. (Papers in the Prehistory of Languages.) Cambridge: The McDonald Institute for Archaeological Research, 223–265.

Stolb. 1987 — Столбова, О., 1987. Сравнительно-историческая фонетика и словарь западночадских языков. *Африканское историческое языкознание*. Москва, 30–268. (Stolbova, O., 1987. West Chadic Comparative-Historical Phonetics and Vocabulary. *African Historical Linguistics*. Moscow.)

Stolb. 2005 — Stolbova, O., 2005. *Chadic lexical database. Issue I (L, N, Ny, R)*. Kaluga.

Sw. 1952 — Swadesh, M., 1952. Lexico-statistical dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proceedings of the American Philosophical Society* 96, 452–463.

Sw. 1955 — Swadesh, M., 1955. Toward greater accuracy in lexicostatistical dating. *International Journal of American Linguistics* 21, 121–137.

Tomb. — Tomback, R. S. *A Comparative Semitic Lexicon of the Phoenician and Punic Languages.* Ann Arbor.

Tos. Das. — Tosco, M., 2001. *The Dhaasanac Language. Grammar, Texts, Vocabulary of a Cushitic Language of Ethiopia*. Köln.

Tser. — Церетели, К. Г., 1958. *Хрестоматия современного ассирийского языка со словарем*. Тбилиси. (Tsereteli, K. G., 1958. *A Neo-Assyrian Anthology (with glossary)*. Tbilisi).

Vyc — Vycichl, W., 1983. *Dictionnaire étymologique de la langue copte*. Leuven.

Статья посвящена детальному этимологическому разбору первых 34 элементов из 100-словного списка М. Сводеша для подавляющего большинства живых и вымерших языков семитской семьи. Основная цель автора — максимально точная лексическая реконструкция соответствующих понятий на прасемитском, а также на промежуточных (западно-семитский, южно-семитский и т. п.) уровнях. Каждая этимология сопровождается подробным обсуждением альтернативных вариантов реконструкции и перечнем наиболее вероятных внешних параллелей в других языках афразийской макросемьи. В ряде случаев приводятся также общие соображения относительно методологии проведения лексикостатистических подсчетов.

George Starostin
Russian State University for the Humanities

Preliminary lexicostatistics as a basis for language classification: A new approach[1]

The article discusses the basic methodology that underlies the construction of a global lexicostatistical database for all of the world's languages, currently one of the main tasks of the Evolution of Human Languages project at the Santa Fe Institute. The author presents several important modifications of the traditional lexicostatistical procedure, such as: replacing the traditional 100-item wordlist with a more compact list of 50 "ultra-stable" items; use of low-level protolanguage reconstructions as primary construction nodes; a combination of the comparative-historical method and principles of phonetic similarity as the basis for the cognate scoring procedure; and, most importantly, a heavy emphasis on semantic precision and severe restrictions on the use of synonyms.

Keywords: lexicostatistics, taxonomy, comparative method, language relationship, semantic reconstruction, Swadesh wordlist.

1. The issue: how to set up the proper criteria for judging language relationship

For over a decade now, the author of this paper has been involved in the long-term scientific project of establishing an up-to-date classification of the world's languages and understanding how far back in linguistic prehistory it is possible to penetrate by using the comparative method — first, within the framework of the Moscow-based "Tower of Babel" project, later, within the broader "Evolution of Human Language" project, centered around the Santa Fe Institute; the major results and conclusions of EHL have been recently summarized in [Gell-Mann, Peiros, Starostin 2009].

At the moment, these results remain largely unendorsed by what may be tentatively called "Western mainstream linguistics" (tentatively, since the very notion of "mainstream lin-

guistics" eludes any precise definition), mainly due to the current trend in thought that tends to emphasize the importance of language contact and areal convergence over that of genetic relationship (for a solid overview of the interaction between the two in different regions of the world, see, e.g., [Aikhenvald & Dixon 2001]). It may, in fact, be noted that the old distinction between the so-called "lumpers" (i. e. those who believe in the historical reality and demonstrability of linguistic macrofamilies) and "splitters" (those in firm opposition to at least the idea of demonstrability of such macrofamilies) can, today, be all but reinterpreted as a distinction between "heritagists" and "arealists". Macrofamily hypotheses such as Altaic, Nostratic, Austric, Amerind, Khoisan, etc., are nowadays most commonly declined by their opponents not so much because the similarities between their members are perceived as random (this factor is still frequently wielded as a counterargument; however, the more rigorous work is being done on these hypotheses, the more it recedes into the background), but primarily because their proponents — so we are told — lack the proper means of separating true traces of common genealogical descent from the effects of "horizontal transmission".

This problem — the difficulty of differentiating between cognate and contact — is, of course, not restricted to hypotheses on long-range comparison; it regularly manifests itself in just about every branch of historical linguistics, which has so far been unable to offer it a uniform, objective solution or set of solutions — or, at least, to set up a certain number of strict "rules of conduct" that all historical linguists would agree to obey when dealing with the issue.

Thorough analysis of available data (first and foremost, Indo-European, later augmented by data from other well-studied families) has shown that, in any comparison of two or more related languages, the best way to distinguish between inherited and borrowed lexical strata is to set up two subsystems of phonetic correspondences — one, reflecting the older inherited layer, will inevitably be more complex and difficult to establish, the other one, representing borrowed items, will be more immediately obvious and consist of generally simpler rules. In this way, it has become possible, for instance, to distinguish between the old layer, inherited from Proto-Indo-European, and the new layer, borrowed from Iranian, in the Armenian language [Hübschmann 1875]; in the same way we distinguish between the "colloquial" — inherited — and "literary" — borrowed — readings of Chinese characters in Sinitic languages (see, e. g., [Starostin 1989: 61–65] for the description of such a differentiation within the Mǐn dialect group).

This criterion, however, is unusable in many types of situations — for instance, when the historical phonetic distance between the languages in question is too small to allow us to distinguish between phonetic laws responsible for vertical transmission and those governing horizontal one; such is the case with, e. g., certain non-literary Dravidian languages (such as Kolami or Gondi), where it is frequently impossible to determine whether a certain item has been retained from the Proto-Dravidian state or borrowed from Telugu. An even more typical situation concerns language families that have not been studied well enough for scholars to arrive at a definitive list of phonetic correspondences, so that distinguishing between any possible layers of lexical interrelation is out of the question. A good example of this is the Khoisan language grouping, where linguistic standards for identifying borrowings are generally substituted for sociological ones [Sands 2001; Güldemann 2006] — i. e., similarity between non-closely related languages is *a priori* attributed to areal diffusion and contact because (a) areal diffusion as such is known to occur in that region and (b) specialists (for now, at least) are unable to explain it properly in accordance with the canon of comparative-historical linguistics.

Even when rules *are* set up to differentiate between "old" and "new" correspondences, this does not serve as a guarantee that the "old" layer will be recognized as representing verti-

cal transmission. For instance, the necessity of disentangling the layers of cross-borrowings within Altaic languages has always been recognized by Altaicists as a *sine qua non* of their field of study, and it is hardly a coincidence that the most recent serious compendium of Altaic etymology [EDAL] attempts to deal with this problem in the very first chapter of its lengthy introduction (pp. 13–21), even before going into the general description of the phonological system of Proto-Altaic itself. In this chapter, the authors take on the issue of two of the most troublesome types of convergence between the descendants of Proto-Altaic (Turco-Mongolic and Mongolic-Tungusic contacts) and, in accordance with the above-mentioned principle, point out the differences between phonetic correspondences that reflect relatively recent contact and those that should rather be interpreted as reflecting relationship; e. g., Middle Mongolian *ašuɣ 'fang', corresponding to Proto-Turkic *aŕɨg, is a borrowing from some form of Old Turkic (azɨɣ), whereas Mongolian araɣa id. is a genetically related form, reflecting the regular correspondence "Turkic *ŕ : Mongolian r" [EDAL: 16].

This argument *per se*, however, does not appear sufficiently convincing to many specialists, who put forward the alternate hypothesis — namely, that this different set of correspondences merely reflects areal contacts that belong to an earlier layer; a particularly appealing theory here is that of a series of "Mongolo-Bulgar" relations during which many Turkic words in a specifically "Bulgar-like" shape must have penetrated the direct ancestor of all medieval and modern Mongolic languages [Georg 1999/2000]. Although non-linguistic evidence to support such a claim seems to be lacking, and a systematic linguistic scenario is hard to construct, theoretically, no matter how many different layers of phonetic correspondences we succeed in establishing, nothing prevents us from simply assigning each of them to a different layer of contact relationships, going back as deep in time as it suits our imagination. (The idea that it must take exactly the same amount of "rigorous proof" to justify a situation of historical contact as it takes to justify a theory of genetic relationship, for some reason, is usually missing in works critical of long-range relationship hypotheses — as if there were something wrong with the idea!)

It seems, therefore, reasonable to assert that, in differentiating between inherited and borrowed lexical layers in the language, we cannot rely on "mechanistic" phonetic criteria alone; each situation of alleged "contact" must also be subject to additional scrutiny, conducted from a statistical ("how much has been borrowed?"), sociolinguistic ("what exactly has been borrowed and why?"), and typological ("how often does this kind of borrowings happen?") points of view. Yet it is precisely these points, particularly the last one, that still remain rather obscure in today's work on language contacts.

The situation has, perhaps, been best summarized in a frequently quoted passage from a paper by Werner Winter: "the inspection of a wide array of observations… leads to the conclusion that in this field nearly everything can be shown to be possible, but… not much progress has been made toward determining what is probable" [Winter 1973: 135]. The quotation is now more than thirty years old, yet, despite the huge rise of interest in contact linguistics, its intonations still ring true; every now and then, we learn something new about the possibilities of borrowing, but we still have no idea of how to estimate the probability of borrowing on a reconstructed, pre-historic level, because there exists nothing like a general typological framework of contact situations to help us with this task.

Should this, however, mean that, simply because we do not have a fully operational model, the linguist should be prohibited from a genetic interpretation of the facts as the likeliest one, and should such a "ban" be equated with scientific caution and healthy skepticism, or would it rather represent an unnecessary hyper-reaction, inhibiting real progress in historical linguistics? I would say that it depends significantly on the situation, and that it

is our duty to learn to distinguish, as objectively as possible, between different types of situations.

A crucial component of the language on which it is reasonable to base our decisions is, of course, the **basic lexicon**, and more or less every serious linguist recognizes that the best place to look for non-contact-induced, non-chance similarities is somewhere in and around the Swadesh wordlist. On practice, however, the "skepticists" never fail to remind that the basic lexicon is only more *rarely* borrowed than the cultural one, and that it is fallacious to automatically count every non-chance similarity on the Swadesh list as reflecting genetic relationship; the very fact that we know for certain that English *mountain* < French *montagne* or that Japanese *niku* < Middle Chinese *n̪i̯uk* should be enough to keep us wary whenever we spot any similarities in the basic lexicon. It is, however, never stated precisely just *how* wary one should be, and what is the "breaking point" at which these similarities should become universally convincing as indications of relationship. Judging by such recent publications as [Yeon-Ju & Sagart 2008], in which it is argued that the Bai language in Yunnan has borrowed as much as 47% of the lexicon from Hàn-era Chinese (unconvincingly, in this author's opinion), one should be wary just about always, but surely this is a rather unsatisfying conclusion, were it to be judged as final.

Another equally unsettling problem, but this time coming from the other side — long-rangers' own elaboration of their hypotheses — is the issue of evaluating competing hypotheses and determining *degrees* of relationship rather than the simple *fact* of relationship. Certain evidence exists, as stated in one of our previous publications on the subject ([Gell-Mann, Peiros, Starostin 2009]; the evidence in question is available at http://starling.rinet.ru, the "Global etymologies" database), that suggests deep-reaching genetic relationship between all major macrofamilies of at least Eurasia, and possibly much of Africa and America as well. Within that scenario, supposing it were true (whether it *is* true does not matter for now), how do we find the means to set up internal subclassification? And how do we choose between mutually contradicting hypotheses, such as, e. g., Starostin's Sino-Caucasian [Starostin 1984] and Sagart's Sino-Austronesian [Sagart 2005], or multiple different models of Nostratic/Eurasiatic?

These and certain other issues can all, in fact, be reduced to a single one — the quest for the Holy Grail of historical linguistics: a set of stable, rock-solid "genetic markers", ones that would be generally stable and guaranteed against the pressures of both internal (ultra-slow rate) and external (resistance to borrowing) change. Since such a set would only make sense if all, or most, of its elements were applicable to all of the world's languages, it is clear that morphological markers and paradigms, one of the most popular types of data in establishing genetic relationship, cannot be part of it.

The typological approach, such as, for instance, is advocated for in [Nichols 1992] and is currently gaining more popularity in diachronic typology, certainly has this advantage of universal application: languages around the world may lack synthetic morphological markers, but no language is known to lack grammatical meaning as such. Nevertheless, it will probably take a lot more time before historical linguists learn to properly rely on typological data as serious argumentation supporting genetic relationship. For now, we have literally heaps of evidence from all the levels of language — phonology, morphology, syntax, semantics — showing how quickly a genetically non-related language can shift its typology once locked in a *Sprachbund* with languages from other families.

To quote but one example, it is rather hard to locate a significant number of typological features that would easily separate Modern Chinese in its Beijing form from the Thai language; the reconstructed Proto-Sino-Tibetan, from which Modern Chinese is unquestionably descended, however, looks seriously different from Proto-Zhuang-Tai in many more respects.

Perhaps some time in the future our understanding of linguistic typology and the mechanisms of its evolution will reach such heights that the "inherent" Sino-Tibetan traits of Modern Chinese will become easily detachable from its areal innovations, but for now it is safe to say that not only do we lack a strict set of rules to separate the wheat of genetically significant typological isoglosses from the chaff of typological diffusion, we do not even know where to begin in order to establish them.

2. Some basic thoughts on lexicostatistics

Coming around full circle, it can be seen that, for the moment at least, we still do not have any serious alternative to basic lexicon when it comes to issues of external relationship and internal classification that involve significant time depths. Discarding lexically based classification as such simply because it runs into certain problems will leave us with either classification methods that are even more questionable, or with no classification methods at all. A far more productive approach would be to tackle these problems head-on in an attempt to minimize their negative effects.

The main goal of this paper is to advocate, once more, the use of the *lexicostatistical* method in both testing hypotheses of relationship and establishing the internal classification of well-demonstrated taxa. In general, I propose nothing new: ever since the popularization of lexicostatistics by Morris Swadesh in the 1950s, it has been used for these purposes over and over again, in many different ways and with widely varying results. The Moscow school of comparative linguistics, in particular, has embraced it as the primary tool due to the works and influence of S. A. Starostin [Starostin 2000, 2007a, etc.][2], and, in recent years, Vaclav Blažek, working in close association with the Moscow school, has initiated a continuing series of papers [2006, 2008a, 2008b and others] that consistently apply Starostin's modified formula of "glottochronological decay" to various language families of Eurasia and Africa, with generally credible results.

(It should be quite specifically stressed at this point that I see it fit to distinguish between *lexicostatistics*, as a procedure that builds genealogical trees based on percentages of cognates on the Swadesh wordlist, and *glottochronology*, as an "add-on" to lexicostatistics that assigns absolute dates to nodes of separation. I am sympathetic to and, with some technical reservations, generally endorse glottochronology, but my primary concern in this paper and the intended follow-ups is with relative, rather than absolute, chronology, and the use of cognate matching in assessing the chances of genetic relationship. Glottochronological dates will be given from time to time merely for the sake of convenience; they are of no crucial importance for the method I am describing.)

Alternate methods and models of classification using the basic lexicon have recently been suggested by non-linguists based partially on their experience in other branches of science, such as Russell Gray [Gray & Atkinson 2003] and Mark Pagel [Pagel et al. 2008]. All of this means that lexicostatistics is still an active field of study, maybe even more active today than during the "lull" period in the 1970s and 1980s, and that the testing of its scope and general capacities is far from over.

[2] Prominent representatives of this school who have, over the last twenty years, offered lexicostatistical classifications for various families, include A. Dybo (Altaic), A. Militarev (Semitic, Afro-Asiatic in general), O. Mudrak (Altaic, Chukchee-Kamchatkan, Eskimo), I. Peiros (Austro-Asiatic, Kra-Dai, Sino-Tibetan), E. Helimski (Uralic) and others; unfortunately, only parts of this data have been published officially.

It must be stressed, however, that, as of now, the word 'lexicostatistics' itself is frequently applied to two significantly different procedures, causing deep confusion among proponents as well as opponents of the method. This confusion is perhaps best exemplified by the following quotation:

"…glottochronology cannot find or demonstrate remote relationships; rather, in the application of the method, forms which are phonetically similar in the languages being compared are checked/ticked as possible cognates and then, based on the number counted, a date is calculated for when the languages split up. That is, the method does not find or test distant genetic relationships, but rather just assumes relationship and proceeds to attach a date. This is illegitimate for research on possible remote linguistic relationships" [Campbell 1998: 185–186].

Lyle Campbell's unwillingness to distinguish (at least, on a practical level) between "lexicostatistics" and "glottochronology" is of no great concern in this context, but his use of the expression "phonetically similar" may be so. The original application of lexicostatistics, as demonstrated in the earliest works of Morris Swadesh on the subject [Swadesh 1952, 1955], was essentially limited to languages whose relationship *had already been demonstrated* through more "conventional" means — such as systematic morphological evidence or the use of the basic comparative method, either thorough (in the case of Indo-European test languages) or partial, but effective (in the case of Eskimo-Aleut). This means that, for Swadesh and everybody else, it is not the forms that are "phonetically similar" which hold the most relevance, but the forms that *correspond to each other historically*, regardless of whether they remain "similar" or not. Were it otherwise, we would hardly expect words like English *eye* and German *Auge*, quite dissimilar phonetically, to be checked as cognates on the list given in [Swadesh 1955].

This original application of the method should, perhaps, be called **classic lexicostatistics** (CL), and it is strange that, in his rejection of the lexicostatistical procedure as such, Campbell does not even refer the reader to its existence. In the general framework of comparative-historical research, CL constitutes merely *the final phase* of the lengthy process of suggesting and testing language relationship through other means such as the ones listed above. Once the process is finished, or, at least, has reached a "respectable" stage at which the relationship is no longer doubted, CL is applied to certify the internal classification of the taxon. CL is, therefore, applicable to language families like Indo-European, Uralic, Eskimo-Aleut, or Mayan, for which we know (or mostly know) the phonetic correspondences, but — at this stage — unapplicable to (in comparison) poorly studied families like Pama-Nyungan, Kwa, or Jê, for which we do not have reliable proto-language reconstructions, even if there is little general doubt of their existence. Even less possible is the application of CL to hypothetical macro-families like Austric or Nilo-Saharan, whose very reality is questioned by numerous specialists in the field(s).

The other way of using lexicostatistics — namely, applying it to assembled wordlists *before* the proper historical research has been performed on them — may be called **preliminary lexicostatistics** (PL). It is true that Swadesh rarely, if ever, explicitly stated the difference between CL and PL, and if his earliest works, meant to present and explicate the method, did not stray away from well-studied language families, some of his later theories, such as the "Dene-Finnish" relationship [Swadesh 1965], were based on a very crude and superficial application of PL, lacking any conclusiveness whatsoever. This, unfortunately, is one possible reason for the fact that the two procedures have also been mixed in works like [Campbell 1998] and others. Below I summarize the crucial differences between both methods:

Parameter	Classic lexicostatistics	Preliminary lexicostatistics
Object of analysis	Basic lexicon wordlists for 2 or more languages known to be related	Basic lexicon wordlists for 2 or more languages suspected of being related
Previous research on object	Relationship demonstrated; phonetic correspondences worked out; protolanguage reconstruction performed	None necessary
Main point of analysis	Cognates scored based on the established system of correspondences	Cognates scored based on phonetic similarity (along with some knowledge of the general typology of phonetic change, if and where possible)
Main result of analysis	Establishing the internal classification of the family	Confirming relationship (and only then establishing internal classification), or rejecting relationship
Typical examples	Isidore Dyen's Indo-European and (less rigorous) Austronesian classifications [Dyen 1965, 1992]; Bastin, Coupez, & Mann's classification of Bantu [1999]; Militarev's classification of Semitic and Afro-Asiatic [2000]	Swadesh's "Dene-Finnish" [1965]

Contrary to Campbell's generalization of PL as the most common understanding of lexicostatistics in general, examples of its application in scholarly literature are quite scarce compared to examples of CL. PL does serve as a major source of classificatory explorations in surveys carried out by members of the Summer Institute of Linguistics (for understandable reasons, given that, for the most part, SIL members work with very poorly studied languages), but very little of their data is actually published in any printed or Internet sources, and, besides, even in their work PL is mostly applied to closely related languages rather than any complicated cases.

I do not, therefore, feel any need to justify the existence and usefulness of lexicostatistics as such; in its CL form the method, applied many times over to relatively well-studied families all over Eurasia and the other continents, has yielded results that are perfectly well compatible with uncontroversial results obtained by other methodologies of classification (such as the "shared innovations" approach), especially with the addition of Sergei Starostin's correction that loanwords detected on the 100-wordlist must be excluded from calculation ([Starostin 1989]; unfortunately, this correction still remains largely unnoticed by critics of the idea of a constant rate of retention, even though it by and large eliminates the issues raised in [Bergsland & Vogt 1962] that once threatened to bury the idea, but, eventually, only helped to reinforce it). Situations in which CL results enter into direct and sharp contradictions with classifications obtained by different means are, by comparison, rare and indecisive, such as the Austronesian case (see, e. g., [Blust 2000], and the counter-argumentation in [Peiros 2000]); their existence no more discredits lexicostatistics than the existence of alternate Indo-European classifications, all of them supposedly based on the same foundation of "shared innovations", discredits the very concept of "shared innovations".

It is also not easy to understand Campbell's argument that, since lexicostatistics/glottochronology simply "assumes relationship and proceeds to attach a date", "this is illegitimate for research on possible remote linguistic relationships". The argument is obviously wrong in the case of CL, but even in the case of PL, where its observation on "assuming relationship" is correct, the conclusion remains obscure. Surely *every* demonstration of relationship, regardless

of the kind of evidence it is based upon, "assumes relationship" and then proceeds to prove it with this evidence. Knowledge or suspicion of language relationship does not fall on us from the sky; we arrive at it through various ways of analyzing data, and one such way can be PL, just as another way could be, for instance, analysis of morphological connections between languages. Perhaps "assumes relationship" is supposed to mean "assumes the relationship as having already been demonstrated beyond doubt by other methods, even though it has not"? But this would be untrue for any application of PL.

The crucial difference between CL and PL — the one that is responsible for widespread application of the former and only marginal and highly controversial application of the latter — is that the former rests on far more rigid standards: reliance on *phonetic correspondences* rather than *phonetic compatibility*[3], working as a solid and, in many ways, objective anchor for the cognate scoring procedure.

Of course, practical application of both procedures shows that, in quite a few cases, the distinction between the two is somewhat blurred, because even for well-studied families like Indo-European, there is always a "fringe" area where uncontroversial etymological decisions are impossible — for instance, do we judge Latin *canis* 'dog' to be cognate with Old Indian *çvan-*, Greek κυών, etc., despite the blatant discrepancy in vocalism, or do we consider it to be a different root altogether (or, perhaps, a contamination of the old root with some other lexeme, leading to the vocalic irregularity?). Another troubling issue is that, according to the procedure as modified by S. Starostin, we are required to filter out borrowings, and it is not always easy to understand if a particular form that has replaced the old root represents an old "native" morpheme in the language or represents a borrowing.

Nevertheless, it goes without saying that, on the average, the better we understand the history of a given language family, the better we can rely on the CL procedure to provide us with a fairly secure genealogical model for it. Complex cases like the one described above can be dealt with on a semi-formal basis, and it is reasonably safe to assume that they will not distort the picture to the point of rendering it useless, especially when the comparison is conducted not on a binary, but on a multi-lateral basis.

Much more troubling is the realization that, for an absolute majority of the world's languages, we simply lack the means to conduct CL in *any* way, because no proper work has been done on establishing a well-defined system of correspondences between them. This does not merely include such "infamous" potential stocks ("pseudo-stocks" from the "mainstream" point of view, which is, technically, not a good term because it intentionally discourages further work on these promising hypotheses) as Indo-Pacific or Amerind, large chunks of which have not even begun to be subject to the appropriate comparative-historical treatment; similar problems crop up with families that are generally thought of as much better understood — e. g. Sino-Tibetan, where the understanding of comparative phonology seriously differs from linguist to linguist (cf., for instance, the many disagreements between models offered in [Peiros & Starostin 1996] and [Matisoff 2003]), or Afro-Asiatic, where some general agreement on the basic correspondences does exist, but the issue of proper matching of cognates still stands tall for each

---

[3] I will be using the term **phonetic compatibility** to refer to situations when two or more words can be judged as cognates either due to their phonetic similarity *or* because their phonetic shapes, although dissimilar, can nevertheless be reasonably connected due to either our general knowledge of the typology of phonetic change or supporting data from other languages. E. g., to quote an example from the Bongo-Bagirmi group of languages, Bagiro *fàɗù* 'fire' would be phonetically similar to Kenga *pòòdò* (the consonantal matches *f : p* and *ɗ : ɗ* are quite straightforward) and phonetically dissimilar, but compatible with Mbay *hòr* id. (phonetic developments *p > f > h* and *ɗ (d) > r* are well-known in the world's languages; also, cf. such related "intermediate" forms as Ngambay *pàr* and Deme *hàdè* id.).

second, if not first, etymology (cf. the numerous discrepancies between, e. g., [Orel & Stolbova 1995] and the more recent and advanced, but still constantly changing, "Database of Afro-Asiatic etymology" by A. Militarev and O. Stolbova, available online at http://starling.rinet.ru).

It may be argued that, since CL is impossible to apply to such families and PL rests on shaky methodology and overestimated intuition, lexicostatistics as such should be ruled out in trying to determine both their internal classification and external relations. But, if so, then what other criterion should *not* be ruled out? Morphological isoglosses between languages are not a universal means of classification, and, besides, they are only as good as the phonetic correspondences they are based upon — which brings us back full circle: no genealogical classification of any family will be resting upon a rock-solid foundation unless a proper amount of historical research has been previously done on it. On the other hand, researching the history of a language family can hardly be done without at least some idea of the internal structure of this family, leading to a vicious circle of sorts.

Still, there can hardly be anything *wrong* in submitting compiled lexical data to a PL investigation as long as we do not forget to state that the resulting classification is not "final" or "proven", but merely a working model — a phylogeny that has to be validated further through much more detailed comparative research. By its very nature, PL will inevitably share some of the flaws of J. Greenberg's "mass comparison" method — although, as will be shown below, many of them will be greatly reduced or completely eliminated — but an *a priori* admittance of its relative non-robustness should save us the trouble of engaging in the same kind of spirited debates that have always accompanied "mass comparison". The statement I want to make is not that "PL is sufficient to establish, beyond reasonable doubt, a general classification of the world's languages", but only that "PL is sufficient to establish a general working model of the classification of the world's languages, prone to refining or refuting, in part or even *in toto*, through ensuing research founded strictly on the comparative method in its Neogrammarian application".

Use of PL as a valid technique to form hypotheses on language relationship and classification is not at all new; it has been employed, in various shapes, by many members of EHL[4] as well as other linguists outside the project. The primary goal of this paper is, therefore, not to introduce and promote it as some radically different technique guaranteed to yield quick and ready solutions, but rather to define, as precisely as possible, the exact conditions under which PL, the way I see it, can and should be used to arrive at a preliminary picture of the world's linguistic situation. First and foremost, this involves answering the following set of questions:

a) What should be the **object** of PL? How much, and what kind of, data, should the compared wordlists include?
b) What should be the basic principle of **cognate scoring**? Should it be "phonetic similarity", "phonetic compatibility", or something else, and how should we avoid subjectivity in this matter?
c) What is the solution offered for the "common plague" of lexicostatistics — the **synonymity** issue? Should synonyms be allowed on the list?

---

[4] In particular, the author of the present paper has himself tested one variant of PL on the Elamite language, leading him to reject the dubious theory of Dravidian-Elamite relationship [G. Starostin 2002], and on the hypothetical Khoisan macrofamily, resulting in a preliminary classification of Khoisan as well as the elimination (for now) of Hadza from the phylum [G. Starostin 2003]. The EHL team also possesses numerous 100-wordlists on Papuan, Australian, Siberian, and Native American families that have been subjected to PL treatment (by O. Mudrak, S. Nikolaev, I. Peiros, and T. Usher), although the results are still being refined and not yet ready for publication. Finally, some PL on the "macro-macro-family" level has been performed by S. Starostin [Starostin 2003], although he usually preferred relying on lexicostatistics exclusively in its "classic" form.

d) In the particular situation when the PL procedure is testing potential long-range rela-
tionship, should there be any **"special" rules** for cognate scoring (distinct from the ba-
sic rules for testing relationship between chronologically more shallow units)?

e) Is there any particular safeguard about mistaking old **contacts** for cognates, and what
kinds of PL lists would decrease the risk of this happening?

Below I will try to answer, one by one, all of these questions, based on both theoretical
considerations and practical results already obtained by myself and my colleagues in the proc-
ess of applying PL to a wide range of families across the world.

3. Selection and compilation of wordlists for preliminary lexicostatistics (PL)

The first issue to be settled within the general task of applying the common PL procedure
to all of the world's major and minor linguistic families is the *degree of shortcutting* that will be
permissible and reasonable in this procedure. To compile Swadesh 100-wordlists — better still,
200-wordlists; better still, 500-wordlists, etc. — for all languages all over the world is a grand
endeavor indeed, but, unfortunately, one that is completely out of the question for now due to
serious lack of manpower, working hours, and, above all, reliable linguistic data, or, in fact,
any kind of data on at least half of these languages.

Fortunately, such an endeavor is also quite obviously excessive if our main goal lies not in
the establishment of a fine-grained internal classification of small, chronologically shallow
groups, but rather in the creation of a general framework, within which it will later be possible
to ascertain individual relations with increased precision. To be more exact, we need not be
significantly concerned with the inner structure of compact groupings that descend from
proto-languages whose age is commonly estimated not to exceed 2,000 — 2,500 years, such as,
e. g., Germanic, South Dravidian, Mongolic, Athapaskan, Daju, North Khoisan, etc. The very
existence of such groupings is generally undisputed (and, more often than not, intuitively evi-
dent even to native speakers), and, for our purposes, it would be more productive to have each
such "primary grouping" represent *one* node on our future "global" tree than to insist upon
"maximum splitting".

One way of achieving that would be to have each such grouping be represented on our
tree by just one "diagnostic" member — e. g., have German (or Dutch, or English, or Swedish)
represent Germanic, Tamil (or Kannaḍa, or Kota) represent South Dravidian, Khalkha Mongo-
lian represent Mongolic, etc. However, such an approach would be painfully anti-historical to
the point of irrationality. Thus, for language groups whose history is relatively well under-
stood, we would frequently find ourselves forced to throw away important data. Limiting our-
selves to German as our "Germanic representative", we would have to note that the word for
'bone' is *Knochen*, and intentionally ignore that it has nothing to do with the common Ger-
manic word **bain-an* for this item [Orel 2003: 32]. Limiting ourselves to Tamil, we would have
to acknowledge (and, in accordance with the procedure, discard) the obvious Sanskrit borro-
wing *nakam* for 'fingernail', instead of the perfectly legitimate Common South Dravidian
*ugur(u) [DEDR: 55], etc.

Things would work even worse in the case of poorly studied or described language fami-
lies, where individual languages almost always are less reliable than comparative data. Thus,
were we to take Mursi as our representative for the Surmic subgroup of Eastern Sudanic, we
would be stuck with the word *hoho* for 'heart', even though the other languages mostly agree
in having an entirely different root: Tennet *zinzet*, Baale *sɪɪnɪ*, Chai *hini*, Koegu *šen*, Meʔen *šini*,

Didinga *dhinit*, etc. Here, not only would we have to discard more important evidence, but we would also have problems with certifying the status of Mursi *hoho* — is this a native Surmic word or a borrowing from some extraneous source?

For these and other reasons, it seems preferable to have the **primary nodes** represented not by any "diagnostic" forms from particular languages, but rather by the likeliest common invariant — in historical terms, the *protoform* for each of the primary groupings.

Usage of reconstructed rather than attested forms in lexicostatistical lists is a slightly controversial, but, perhaps, inevitable application of the method. Its most ardent supporter used to be the late S. Starostin, who was particularly adamant about using reconstructed forms to test hypotheses of long-range relationship [Starostin 2003], an issue which we shall consider in more details below. Most Western linguists have generally refrained from following his example, but this mostly has to do with the fact that, for their particular purposes — usually having to do with building an internal classification for just one family — this was simply unnecessary. Even if we want to build a grand lexicostatistical tree for such a huge family as, say, Austronesian ([Dyen 1965], [Blust 2000], [Greenhill et al. 2008]), we do not require the use of reconstructions: most of the attested Austronesian languages have preserved sufficient quantities of "Proto-Austronesian lexical stock" for us to be able to measure and grade these quantities. But if our aim is to cover the entire globe, this is a different matter; it requires "short-cuts", and reconstructions are both the most logical and the most honest ones.

There are, however, two obvious questions that arise from using low-level reconstructions. These are: (a) how can we be certain of the validity of the reconstructions, especially for families that have not been well studied in the historical perspective?; and (b) in the case of several alternatives, how do we select the one root to represent the entire family?

The first question requires a special answer, and we will tackle it in the corresponding section; for the moment, let us assume that *in general*, low-level reconstructions for our list can be obtained relatively easily and with plenty of confidence. As for the second question, it is tightly connected to the issue of dealing with synonymity on the wordlist, and will also be discussed specially. For now, I will simply say that both issues are problematic, but that there also are ways to minimize these problems or, at least, to deal with them on a formal basis.

Now that we have chosen low-level reconstructions[5] as our main object of study, the next obvious issue is quantitative: how many items do we need for our list? The initial consideration would, quite naturally, be to simply use the "classic" 100-item list as originally selected by Morris Swadesh, especially since for many languages, ready-made 100-wordlists are already available.

However, given our stated purpose, it can be argued that use of the *entire* list will be excessive, both for technical and substantial reasons. From a practical viewpoint, requiring that all the positions on the list be filled in would inevitably hinder the inclusion of quite a few low-level language groups in Africa, America and the Pacific region, where for many languages we only have very short — but, nevertheless, still informative — wordlists collected under specific "rapid survey" conditions. While these wordlists may, and should, be used as valuable data for genetic classification, demands for more data would force us to reject them as evidence, which would hardly be reasonable.

---

[5] For our purposes, here and below "low-level reconstructions" will be understood as "most probable lexemes with a particular meaning that can be reconstructed for the immediate ancestor of a group of languages that is uncontroversially understood to be related and whose members share, on the average, no less than 50% of cognates on the regular Swadesh 100-wordlist." It should be noted, of course, that language isolates, having no close relatives, will, in any case, have to be represented by modern attested forms on our list.

Another, more serious, consideration is that for our purposes 100 items may simply be excessive. It has always been clear, both to opponents and proponents of lexicostatistics alike, that some words on the Swadesh wordlist are generally more stable than others (e. g. the words for 'eye' or 'two' are empirically known to be replaced far less frequently than the words for 'round' or 'yellow'), and this, in turn, led to suggestions about replacing the original Swadesh "stability quotient" of 0.14 (or the "improved" Starostin quotient of 0.05) with individual stability quotients for each item on the list[6].

An attempt at empirically calculating the individual "stability level" for all 100 items was actually carried out by S. A. Starostin [Starostin 2007a], based on a simple procedure of calculating a "stability index" for the items within a particular family (the general criterion here is the number of different roots that are used within the family to denote the item) and then averaging the indexes across the world (calculations were performed for wordlists of the following families: Afro-Asiatic, Altaic, Australian, Austro-Asiatic, Austronesian, Daic, Dravidian, Indo-European, Kartvelian, Khoisan, North Caucasian, Sino-Tibetan, Uralic, Yeniseian). Since the results have not been published in English, it makes sense to reproduce the resulting list here, ranged from the most stable items to the least stable ones (I omit the 10 "additional" elements to the 100-wordlist that are sometimes used in calculations):

1. we	21. one	41. stand	61. meat	81. night
2. two	22. tooth	42. tree	62. road	82. see
3. I	23. new	43. ashes	63. know	83. walk (go)
4. eye	24. dry	44. give	64. say	84. warm
5. thou	25. liver	45. rain	65. egg	85. red
6. who	26. eat	46. star	66. seed	86. cold
7. fire	27. tail	47. fish	67. knee	87. woman
8. tongue	28. this	48. neck	68. black	88. round
9. stone	29. hair	49. breast	69. head	89. yellow
10. name	30. water	50. leaf	70. sleep	90. lie
11. hand	31. nose	51. come	71. burn	91. green
12. what	32. not	52. kill	72. earth	92. cloud
13. die	33. mouth	53. foot	73. feather	93. big
14. heart	34. full	54. sit	74. swim	94. bark (*of tree*)
15. drink	35. ear	55. root	75. white	95. sand
16. dog	36. that	56. horn	76. bite	96. good
17. louse	37. bird	57. fly	77. fat	97. many
18. moon	38. bone	58. hear	78. man	98. mountain
19. fingernail	39. sun	59. skin	79. person	99. belly
20. blood	40. smoke	60. long	80. all	100. small

Prior to the compilation of this index, Starostin and other EHL/Moscow school members would occasionally rely, instead of or in addition to the standard Swadesh wordlist, on a

---

[6] See, e. g., [Merwe 1966]. In [Starostin 1989], the idea was reflected indirectly by introducing a special parameter — deceleration of the rate of change of the original wordlist depending on the amount of unreplaced items remaining on the list at any given time — but later on, the method of using individual quotients instead of a fixed one was successfully incorporated by him into STARLING computer software, and is now tested by EHL members and their colleagues (as the "experimental method") along with calculations based on a fixed quotient (called the "standard method"). In most cases, "experimental" and "standard" calculations yield surprisingly similar results, although the "experimental" method tends to yield slightly earlier glottochronological dates.

shortened 35-item version of it, compiled by S. Jaxontov (the list originally appeared not in any of Jaxontov's own publications, but in [Starostin 1991: 59–60]). This 35-item list, in Jaxontov's opinion, constituted the generally more stable part of the Swadesh list, and the theoretical idea behind it was that any two or more related languages always had to show a larger percent of matches within this section than within the remaining 65-item section, the reverse situation indicating language contact rather than language relationship. This idea was heartily embraced by Starostin in much of his work (in particular, to validate the Altaic theory); more importantly, the 35-wordlist has been used by him as a possible "shortcut" to arrive at a preliminary classification of the language families of Eurasia (unpublished).

The major problem with Jaxontov's list, however, has always been that the exact considerations underlying the selection of 35 items out of a total of 100 have never been stated expressly; it seems that, for the most part, the words had been chosen simply based on his own linguistic experience, gained from working on the history of language families in one particular area — Southeast Asia. However, the list from [Starostin 2007a], compiled on the basis of a somewhat more formal and objective principle, shows that Jaxontov's intuition has misled him into "overrating" the overall stability of some items (namely, 'sun', 'bone', 'give', 'fish', 'salt', 'horn', 'egg', 'know') while "underrating" others ('we', 'fingernail', 'heart', 'not', 'liver', 'eat', 'mouth', 'dry', 'hair', 'drink')[7].

Now that we stand on somewhat firmer ground in determining which items are more stable and which ones are not[8], it is only natural that, for the purpose of establishing a general classification scheme even for one macrofamily, we do not really need all one hundred items. To take but one example: S. Starostin quotes 26 cognate matchings between Indo-European and Uralic on the list [Starostin 2003: 482], but if we split the list into two equal parts — the generally more stable items 1–50 and generally less stable items 51–100 — the first part, predictably and in accordance with "Jaxontov's law", will yield more matches (17) than the second part (9); in addition, these 17 matches are generally less questionable from a phonetic, semantic, and distributional point of view than the other 9. The situation does not change much if we look at more shallow time depths: out of the 42 direct matches between Finnish and Saami, 28 belong to the "stable" half of the list, and only 14 — to the "non-stable" part of it.

---

[7] Jaxontov's full list looks as follows: 'blood', 'bone', 'die', 'dog', 'ear', 'egg', 'eye', 'fire', 'fish', 'full', 'give', 'hand', 'horn', 'I', 'know', 'louse', 'moon', 'name', 'new', 'nose', 'one', 'salt', 'stone', 'sun', 'tail', 'this', 'thou', 'tongue', 'tooth', 'two', 'water', 'what', 'who', 'wind', 'year'. Note that three of these words — 'salt', 'wind', 'year' — do not constitute part of Swadesh's original 100-wordlist (taken from the second half of the 200-wordlist instead).

[8] A radically different approach has recently been advocated by Mark Pagel and others [Pagel et al. 2007], who propose to predict "stability" of particular items based on their relative frequency in the language (more frequently used items tend to be more stable), illustrating this on the example of large lexical corpora drawn from four Indo-European languages. While it would be rash to claim that frequency of usage has nothing whatsoever to do with "stability", it is also safe to assume that it is but one of the supposedly many factors influencing "stability". Pagel and his co-authors do not give individual statistics for each word, but it is very hard to believe that, for instance, the word for 'fingernail' in Indo-European (very high stability rate of 0.92, according to Starostin) is used more frequently by active language speakers than the word for 'blood' (very low stability rate of 0.18), or that the word for 'new' (0.90) is used more frequently than the word for 'many' (0.19). In addition, what works for Indo-European will not necessarily work for other language families. Thus, numerals 'one' and 'two' are almost never replaced in Indo-European, which may be accounted for by the extremely high frequency of both words; outside Indo-European, however, we constantly find that the word for 'two' has a much slower rate of replacement than the word for 'one' (cf. in Uralic: 1.0 vs. 0.65, in Daic: 0.79 vs. 0.55, in Kartvelian: 0.86 vs. 0.57, in Sino-Tibetan: 0.92 vs. 0.37), even though there is little reason to think that speakers of these languages resort to saying 'one' far less often than they say 'two'. As attractive as Pagel's model is on the surface, at this point it cannot be used for any practical purpose.

To cut a long story short, it is not very likely, given their observedly poor "performance" on shallow chronological levels, that words like 'road', 'swim', 'cloud', or 'yellow', to name but a few, will persevere over several millennia[9], yielding precious lexicostatistical information about long-distance relationship. Since there is no general, exceptionless "law of retention" for each individual word, occasional exceptions must and will occur, but their efficiency will be quite low compared to the troubles of compiling full-fledged 100-item wordlists and, more importantly, the troubles of cognate matching between poorly studied families, which will increase significantly for unstable words (any historical linguist who has seriously studied existing reconstructions, or contributed to any of them him/herself, knows how much more difficult it generally is to reconstruct the protoform for 'big' or 'warm' or 'root' than it is for 'ear' or 'eye' or 'die').

It may be argued, in fact, that testing relationship hypotheses on different chronological levels requires wordlists of different sizes. Obviously, if we want to measure the lexicostatistical distance between closely related languages or dialects, such as East Slavic, Scandinavian, Oghuz, or North Khoisan, limiting ourselves to the "stable" half of the Swadesh wordlist will almost certainly result in an incorrect classification: most, if not all, of the words will simply match, and we will get, at worst, a zero degree of separation, at best, minimal degrees that will all lie within the margin of error and tell us virtually nothing. For such purposes, we would definitely need all 100 words, or perhaps, better still, the full original 200-word list. But already for Indo-European, utilizing only the "stable half" seems to yield results that are not too far removed from results of the regular classification based on all 100 items — at the very least, all the subgroupings are "recognized" properly.

The choice of 50 as the "magic number" is somewhat arbitrary, but not entirely so: a 50% discrepancy between the wordlists of two different languages (corresponding, according to the glottochronological formula of S. Starostin, to approximately 3,000 years of divergence time) is generally the threshold beyond which relationship ceases to be "intuitively obvious" and requires resorting to more sophisticated methodology in order to become transparent, and we may reasonably expect that the non-stable elements will, overall, be the first to go, or, at least, will fade away about twice as fast as the stable ones. On the other hand, the 35-item list, previously employed in some long-range calculations, will not be convenient for us if we want to utilize the material of units like Proto-Germanic and Proto-Slavic (on Indo-European territory), or Proto-Ethiosemitic and Common Arabic (on Semitic territory) — there will be way too few differences to be of any statistic relevance. At the moment, 50 items looks like the most promising alternative, by way of compromise between the different extremes.

On the other hand, mechanistically selecting the first half of the list (stopping at the word 'leaf') will inevitably lead to certain practical difficulties and imbalances. Up until the number 24, I have no problems with it, but beyond that number I propose nine replacements of "more stable" items by "less stable" ones in order to facilitate the work on both the compilation of the wordlists and the scoring. The following items are to be discarded:

a) 'this', 'that': first, the wordlist is already heavily biased towards pronouns ('I', 'thou', 'we', 'what', and 'who' are all included), second, stems like *a* 'that', *i* 'this', etc., are nearly universal, rendering them of little use for global classification purposes, and third and most important, many languages around the world show far more than these two basic degrees of

---

[9] It should perhaps be strongly emphasized that, in the strict lexicostatistical spirit, I am talking about *words*, i. e. "form : meaning" pairs, not etymological *roots*, prone to meaning shifts. A *root* with an original meaning like 'swim', 'yellow', etc., obviously has a better chance of being preserved over lengthy time periods than the original bundling of its meaning with its form.

deixis (e. g. triple systems like 'this near' — 'this/that neither near nor far' — 'that over there', etc.), complicating the already pressing synonymity issue;

b) 'liver': this word, despite its relative stability, is very frequently not included in short wordlists collected on data survey trips, and would have to go missing in quite a few cases anyway;

c) 'fish': this item is frequently lacking in desert communities (e. g. it is not attested at all, or represents an obvious recent borrowing, in quite a few Khoisan languages), for the languages of which it will be of no use whatsoever;

d) 'neck', 'breast': these words are not only at the very bottom of the "stable" list, but they also frequently tend to be sound-symbolic ('neck' frequently is the same as or stems from 'throat', where onomatopoeic forms like *kurʌ, qurʌ* are of little diagnostic value, and 'breast' is frequently the same as 'mother', representing nursery forms; also, confusion frequently arises as to whether the intended meaning is 'male chest' or 'female breast(s)');

e) 'full', 'stand', 'give': the semi-abstract semantics of these verbal/adjectival roots has been frequently found a big "nuisance" (they tend to have multiple synonyms where it is frequently impossible to tell the difference), and, in general, it is advisable to have as few verbal roots on the list as possible[10].

For these reasons, it looks justified to remove these nine items and replace them, respectively, with nine other ones that may not be as stable, yet, on the average, turn out to be less of a bother on practice: 'kill', 'foot', 'horn', 'hear', 'meat', 'egg', 'black', 'head', 'night'. I shall not give out detailed reasons for these particular choices; let us simply assume that the swap will hardly make any profound substantial difference, but will inevitably facilitate the overall work process.

We will designate this array of 50 lexical "genetic markers" as the *main wordlist* (MW), opposed to the *original wordlist* (OW) that contains all 100 items. The presumption is that the slots on the MW are occupied by low-level reconstructions; these low-level reconstructions, in turn, are generally based on OWs (and, where possible, on even more detailed etymological databases) for the respective low-level families — data that actually allows us to produce a low-level reconstruction, as well as establish the internal classification of the low-level family.

E. g., the MW for "Slavic" looks like [1] *pepel-ъ* 'ashes', [2] *pъt-a* 'bird', [3] *čьrn-ъ* 'black', etc.; the reconstructions are validated by OWs for several Slavic languages, which not only confirm these reconstructions, but also contain etymological information on other words like 'all', 'bark', 'belly', 'big', etc., to ensure more accurate internal classification of Slavic languages.

4. Cognate scoring: a compromise between the comparative method and "phonetic compatibility"

Now that we have established the basic constituency of the MW and the type of information in it (low-level reconstructions), the most important question is setting up the rules for scoring potential cognates. This is tricky, since any such procedure, unless operating on a fully automatic, machine-conducted basis, could easily lead one into the trap of subjectivity. Even well-established families frequently show irregularities that allow for different interpretations

---

[10] M. Robbeets [2005: 50], on the contrary, advocates for an increased use of verbal roots to demonstrate relationship, claiming that verbs tend to be borrowed far less frequently than nouns. Her observation is quite correct, but this advantage is completely annulled by the tendency of verbal roots to be generally less stable within the basic lexicon than nominal ones — a tendency that is fully confirmed by the adduced stability index, where we find only 5 verbs ('die', 'drink', 'eat', 'stand', 'give') in the upper half and 14 in the lower half, and the ratio is even worse for adjectives (which are frequently undistinguishable from verbs in languages around the world) — 3 vs. 13!

(a typical example would be Latin *canis* 'dog', whose correspondence to Proto-Indo-European *$\hat{k}won$- ~ *$\hat{k}un$- is obviously irregular, but no consensus has been reached on whether the form itself represents an entirely different root or a regional 'permutation' of the original entity), and the situation becomes much worse when we start dealing with medium-level (or even low-level) families that have not been subject to a great deal of historical research, not to mention any possible long-range connections.

On the other hand, use of a fully automated procedure, completely wiping out subjective approaches to etymology, would deprive us of the same factor of *historicity* that we tried to bring in by choosing low-level reconstructions as the main point of entry. Such procedures chiefly operate on the principle of "phonetic similarity" — matching phonemes (usually consonants) across compared languages according to their belonging to one of several distinct phonetic classes — and, in the end, this is exactly what is actually being measured: the degree of phonetic similarity, meaning that, for instance, languages that are in reality more distantly related to each other but more archaic in their phonetic systems may end up as more closely related than languages with innovative phonetic structures.

The major weaknesses of getting history out of the picture are, perhaps, most clearly illustrated by the recent results of the international ASJP (Automated Similarity Judgment Program) project hosted by the Max Planck Institute, whose major aim is presented as "achieving a computerized lexicostatistical analysis of ideally all the world's languages" (http://email. eva.mpg.de/~wichmann/ASJPHomePage.htm). The selected method — a moderately sophisticated procedure of estimating "degrees of phonetic similarity" between pairs of words — results in the construction of a phylogenetic tree [ASJP 2009] where historically correct nodes are hopelessly mixed with nodes that reflect either areal convergence (e. g. the closest branch to Sinitic turns out to be Hmong-Mien instead of Tibeto-Burmese), differences in the rate of phonetic evolution as mentioned above (e. g. Kota is not recognized as a South Dravidian language, although it most certainly is), or straightforward absurdities (e. g. the closest neighbour of Khoisan languages turns out to be… Kartvelian!)

Participants of ASJP obviously understand these limitations of the method and are able to correctly identify most of the underlying causes [Wichmann et al. 2009]. This understanding, however, does not really answer the inevitable question — of what particular use is the produced tree? The importance of assessing an average degree of "lexical similarity" between the world's languages without distinguishing between various factors that cause this similarity is quite dubious, since such information cannot be reliably used for any further scientific purposes. And if our specific purpose is to arrive at the likeliest — in the light of available data — genealogical tree for the world's languages, then the importance of the ASJP assessment drops to zero, as it is quite liable to rewarding us with large quantities of false positives and equally false negatives.

Less "global" applications of various statistical procedures measuring and analyzing degrees of phonetic similarity have yielded interesting, but inconclusive results. Thus, Baxter & Manaster-Ramer [2000] have, based on the comparison of only one phonetic segment (the initial consonant), shown that the number of phonetic resemblances on Jaxontov's 35-wordlist between English and Hindi exceeds chance expectations and serves, therefore, as proof of relationship (presumably, contact is all but excluded in this particular situation), disproving the popular myth that it is impossible to demonstrate the existence of Proto-Indoeuropean without having access to ancient language data. However, there is no guarantee that the same procedure would work equally well on *any* pair of languages known to be related[11].

---

[11] Baxter & Manaster-Ramer's method established nine potential cognates between English and Hindi, only five of which were true from a historical point of view. The method determined that number to be sufficient; how-

Recently Turchin, Peiros & Gell-Mann [2010] have tested a similar, but slightly more sophisticated method, with extra safeguarding against the effects of language contact, that seems to yield true positives for the case of Altaic relation. Their case, however, is not one of simply measuring "pure" phonetic similarity between attested languages: the procedure is tested on reconstructions of Proto-Turkic, Proto-Mongolic, etc., meaning that they are not unwilling to take historical information into consideration. Tests that have tried to verify hypotheses of long-range relationship based exclusively on data from modern or historically attested languages — e. g., [Ringe 1992], [Kessler 2001] — have almost invariably failed to come up with any positives (but it must be noted that Ringe does report "weak positive" results for Indo-European and Uralic; somehow, though, even this has not brought mainstream linguistics any closer to a common acceptance of "Indo-Uralic" as a historically valid taxon).

Of course, automatic procedures need not necessarily be as simple as that. In addition to estimating degrees of phonetic similarity between compared words (either absolute or relative), such a procedure can attempt to establish patterns of potential correspondences — essentially, doing much the same things that a real comparative linguist, equipped with knowledge of Neogrammarian methodology, would try to do with a bunch of unfamiliar material. This implies that the algorithm will try to match not merely similar, but, in fact, *any* consonantal classes, and try to determine those matches that are statistically significant. One such procedure, designed by the author of this paper with the help of programmer Phil Krylov (see [G. Starostin 2008]), does show far more promising results for relatively closely related languages; results report, among other things, a total of 64 out of 77 cognate forms between modern English and modern High German recognized — a number which is further increased to 72 out of 77 when the comparison procedure is extended from binary to multilateral (including lexicostatistical data from other Germanic languages). The algorithm even seems robust enough to recognize some of the "controversial" intermediate level groupings, such as Altaic or North Caucasian (relationship between Nakh-Daghestanian and Abkhaz-Adyghe languages).

On the other hand, the capacity of this procedure is, even at this point, insufficient to match quite a few of the obviously correct etymological decisions that comparative linguists have "manually" established over the years. The main reason is clearly the insufficience of data present on the 100-wordlist. For instance, the algorithm was incapable of understanding the cognacy of English *mouth* and German *Mund*, because the regular correspondence "English *zero* : German *n*" (more precisely, of course, "English *th* : German *nd*") could not have been substantiated by any other examples[12]. Stepping outside the wordlist, it is easy to ascertain that the correspondence is indeed regular even without resorting to the more archaic stages of both languages (cf. such examples as *other* : *ander*, *youth* : *Jugend*, *lithe* : *linde*, *un-couth* : *kunde*, etc.), but this would require having the algorithm run through the entire compared vocabularies and, in addition to valuable information, picking up a huge lot of "noise" (false cognates, shared borrowings from third sources, etc.) that could seriously distort the desired results.

The conclusion is that "rough" automatic data handling is, at present, unable to arrive at the same level of precision in its results that can be provided through manual handling of the same data; the obvious benefit of "weeding out subjectivity" does not fully compensate for the

---

ever, e. g., a similar search that I have attempted between Modern Chinese and Lhasa Tibetan finds only six potential cognates, with only four of them historically true — although I have not performed the second part of their test (the "shuffling" trials), I believe its results are quite predictable.

[12] Of course, the actual "recognizal" of this cognacy depends on the specific rules of segmental alignment that are set up; e. g., if free deletion of the middle segment in a triconsonantal sequence (MNT) is allowed so that MNT = MT, the two words are judged as cognate. It is, however, always questionable whether such "free deletions" are admissible in these automatic procedures and do not undermine their robustness.

lack of fine-graining analysis techniques — techniques which, more often than not, are a very serious influence on classification schemes. This does not mean that automatic procedures should be abandoned; on the contrary, one of our major goals should be to refine and readjust them in accordance with the basic principles of historical linguistics[13]. In the meantime, however, we can only place more trust in manual procedures, all the while attempting to enforce maximally formal criteria. In other words, it may be too early to teach the machine to behave like a human, but it is, in some respects, easier to make the human behave like a machine.

Therefore, for our classification based on 50-item wordlists we will ultimately be relying on manual rather than automatic cognate scoring. This gives us the important bonus of being able to use all kinds of historical information and reliable historical conclusions accumulated over two hundred years of incessant work by specialists in language comparison. The two basic principles of scoring will be defined in the following way:

1. For language groups already studied by the comparative method, judgements about the cognacy of particular items will be made on the grounds of recognized **regular phonetic correspondences** between said groups.
2. For language groups that lack serious comparative study, judgements on cognacy will be made on the grounds of (a) **phonetic similarity** of the items concerned, or (b) **phonetic compatibility** of the items, provided it is possible to base the judgement on **traces of regularity**.

Both points require more precise comments. First of all, it must be made clear that in a lot of situations it is hard to make a clear distinction between the two types of scoring. "Historically studied" is not an absolute definition: no two language groups in the world have received a completely equal amount of study, and our knowledge of the regularity of correspondences is always relative rather than absolute. Even Indo-European is prone to cases where it may be reasonable to sacrifice regularity and resort to scoring on the grounds of phonetic similarity instead.

Case in point: do we judge Old Indian *hṛd* 'heart' as cognate to Germanic **xirt-*, Slavic **sъrdь-ce*, Greek κῆρ, etc. ← IE **k̂ṛd-*, or do we score it differently, since it violates the regularity principle (the Old Indian form should reflect IE **ĝhṛd-*)? In Pokorny's dictionary, an authoritative but by no means dictatorial source, the Indo-Iranian root is judged to represent a separate "Reimwort" [Pokorny 1958: 580], not to be related to **k̂ṛd-*. Intuitively, however, it is extremely hard to think of the two variants as having nothing to do with each other — apart from complete regularity in every other respect, there is also the important issue of *representativity*: the two variants are in complementary distribution throughout Indo-European, and no non-conjectural evidence can be found as to their co-existence in at least one branch of the family. Hence, probably, the "compromise" solution of **ĝhṛd-* as a "rhyme word", adopted by Pokorny — a solution that achieves nothing, since nothing is explained about the mysterious

---

[13] In this respect it is necessary to mention a project (to the best of my knowledge, there is no official name for it as yet, but "Network models of sound change" has been offered as one way of description) recently undertaken by several linguists and specialists from other fields, also based at the Santa Fe Institute and supervised by some of its resident professors (Tanmoy Bhattacharya, Daniel Hruschka, Eric Smith, Jon Wilkins and others). The project's aim is to produce a major quantitative framework for recognizing and describing patterns of regular change, which could, if successful, be then used as the best possible automatic tool for generating classification schemes. On the other hand, the aim is so global that it is so far unclear how much time it will be needed for it to come to fruition. A little more information on it can be found in [Christiansen et al. 2009], as well as the official site of the Santa Fe Institute (http://www.santafe.edu).

origins of this "rhyme word" (did it exist in Proto-IE? was it an original concoction on Indo-Iranian grounds? how did it originate? are its origins related to the existence of *k̂r̥d- or is it just a fortunate coincidence? etc.), but at least spares the author from the painful Neogrammarian duty of declaring the phonetic similarity between the two variants as the result of pure coincidence.

The *representativity* criterion — which, in this case, merely represents a particular application of Occam's razor — would strongly speak in favor of judging the Old Indian form as cognate with the rest of Indo-European. The exact reason that underlies the irregularity remains unknown, with several *ad hoc* explanations possible (idiosyncratic development of some old non-trivial cluster, perhaps with a laryngeal; assimilatory influence of two ensuing voiced segments; analogy/contamination with some other word; taboo, etc.) but none of them supported by strong independent arguments. But the assumption of a *lexical replacement* in this case would reduce the Neogrammarian model to absurdity, and, more importantly, leave us with a far larger number of unanswered questions (see above) than the assumption of an unexplainable irregularity.

Therefore, in making cognation decisions even for families with a generally elaborated historical phonetics and a large etymological corpus, it is reasonable to allow for occasional irregularities in the forms, *especially* when the two irreconcilable variants appear to be in complementary distribution *and* there is no easy way to "explain away" one of the variants as having an entirely different origin. A demand for utmost mechanistic rigor will inevitably result in our throwing away true historical cognates and coming up with unnecessarily distorted classification schemes. We may formulate the main rule of exception as follows:

1a. Phonetic irregularities between potential cognates within groupings for which a system of phonetic correspondences has been established may be ignored if [a] they concern *not more than one* consonantal segment of the root (out of two or more), [b] the phonetic distance between the two segments does not make them *phonetically incompatible*, [c] the two variants — "regular" and "irregular" are in *complementary distribution* across languages and cannot be clearly shown to fall under two different etymologies[14].

Concerning the second type of situations — those for which comparative studies are in their initial phases, or non-existent — it is also easier to illustrate the exposed methodology with real examples, this time taken from the African area. Let us consider the following forms from various "branches" of the hypothetical Nilo-Saharan macrofamily, all of them with the meaning 'to drink'[15]:

a) East Nilotic: Teso *akı-mát-à*, Turkana *akı-mat*, Nyangatom *t̪ɛ-mɛt-*, Karimojong *aki-mát*, Maasai, Sampur *a-mát*, Ongamo *-mát-à*, Lotuko *a-máð-à*, Oxoriok *mat-a*, Lopit *mát-à*, Dongotono *a-mát*, Lokoya *a-mát-à*. East Nilotic is a relatively compact and well-recognized language family, with a preliminary reconstruction published by R. Vossen, who reasonably reconstructs this particular root as PEN *-mat- [Vossen 1982: 356], and there are no grounds to doubt that it functioned as the main root for 'drink' in that proto-language. (It is unclear if the Bari subgroup form *mō-ʒu is also related — probably not, but in any case it will not affect our selection of *-mat-, since it is overall better represented in the family).

---

[14] A counter-example to 'heart' would be the case of Slavic *kostъ 'bone' vs. IE *(H)ost- id. → Hittite *hastai-*, Old Indian *asthi-*, Latin *os*, etc. Not only is the correspondence "Slavic *k- : IE zero" completely irregular and phonetically incomprehensible, but, more importantly, IE *(H)ost- is easier relatable to Slavic *ostъ 'sharp edge, awn', while Slavic *kostъ is better etymologized together with Latin *costa* 'rib'. There may have been semantic contamination between the two words in Proto-Slavic, but there is little reason to doubt the presence of *two* roots on the IE level, and the Proto-Slavic item on the list should be scored differently from the rest.

[15] Since this is merely a methodological example, I do not quote all the data sources for particular forms so as not to inflate the list of references too much. Only the sources for protoform reconstructions are quoted.

b) West Nilotic: Nuer, Shilluk *math*, Anywa *màath*, Luo *màð-*, Päri *maath*, Lango *mato*, Mabaan *mɔča*, Jumjum *maaṉ-ṉʌ*. All the forms are clearly related (with nasal assimilation in Jumjum), and, although no special published reconstruction of West Nilotic is available, we may safely follow G. Dimmendaal in setting up the proto-form **maḓ* [Dimmendaal 1988: 38].

c) Surmic: Chai *mat*, Koegu *amátiyaa*, Me'en *mad-*. These three forms are phonetically similar and most likely related, even though we have so far had no attempts at a Proto-Surmic reconstruction. We may provisionally set up a reconstruction **maT-*, indicating lack of knowledge about the exact manner of articulation of the intervocalic coronal stop.

d) East Jebel: Aka *mɛɛtu*, Molo *mootu*, Kelo *məd-ea*, Beni Sheko *mɪdi*, Gaam *məϑ-*. This is also a well recognized language group, and we feel justified following M. Lionel Bender's preliminary reconstruction **mVt-* [Bender 1998: 56].

e) Berta: *meera*. Berta is an isolated cluster of several extremely similar dialects, with no uncontroversial "relatives" to speak of (C. Ehret thinks of it as the closest relative of East Jebel, but this classification is highly disputed).

f) Central Sudanic: Moru *u-mvú*, Avokaya, Ma'di *mvu*, Logo, Keliko, Lugbara *mvú*, Lulubo *mbú*, Lendu *mbu*, Ngiti *ɔmvʋ̀*, Mangbetu *ɔmbuo*, Kresh *ɔmɔ*, Aja *amú*. This is one of the primary roots for 'drink' in this large language family, and its proto-invariant should be approximately (for lack of an overall credible Central Sudanic reconstruction) reflected as **mvu* (Ehret [2001: 275] reconstructs East Central Sudanic **mbu*, but the root has a wider distribution, since Kresh and Aja are not ECS).

All of these six branches are included by J. Greenberg within his "Nilo-Saharan", and this decision is upheld by such prominent Africanists as M. Lionel Bender, C. Ehret, and others. However, at the moment, only the relationship between (a) and (b) happens to be completely uncontroversial. The grouping of Surmic and East Jebel languages together with the large Nilotic family as separate units of "Eastern Sudanic" is generally questionable; the grouping of Berta within the same family even more so; and the relations of the whole *ensemble*, on a seriously "macro"-level, with Central Sudanic, are a problem of about the same scope as Nostratic or Austric relationship, if not more so.

In the light of this, we approach all of these groups as potentially related, but consider this relationship, for the moment, insufficiently substantiated through the comparative method, meaning that the situation here clearly falls under type (2). The scoring will, therefore, be conducted as follows:

— West Nilotic **maḓ* and East Nilotic **-mat-* are scored as cognates, based on phonetic similarity as well as preliminary correspondences, established in [Dimmendaal 1988] and elsewhere;

— Surmic **maT-* and East Jebel **mVt-* are also scored as cognates both between themselves *and* with Nilotic, based on phonetic similarity;

— Berta *meera*, theoretically, could be scored as cognate to all four. However, there is a serious problem with the second consonantal segment: it belongs to a somewhat, if not crucially, different consonantal class[16], and, in order to be more secure about the cognacy, we need to support it by finding *traces of regularity*, i. e. at least one or two more exact or near-exact se-

---

[16] On the basic principles of classifying consonants into non-intersecting "classes" based on similarity of articulation, see [Baxter & Manaster Ramer 2000; Dolgopolsky 1986; G. Starostin 2008]; proposed models frequently differ as to the degree of detalization (e. g., do we place such front consonants as *t, s, r* in the same class or in three different ones?) — I would opt for a more detailed classification, so that such forms as [*pata*] and [*para*] be judged phonetically compatible rather than phonetically similar, and require the presence of additional "traces of regularity" to be scored as cognates.

mantic matches — not necessarily within the Swadesh wordlist — that would support the correlation. So far, I have been unable to do that, and this means that Berta *meera*, for now, has to be judged as a different root[17];

— the Central Sudanic forms certainly share the initial consonant with the rest, but there is no evidence (for now) that the protoform at one time suffered the loss of the root-final coronal consonant, or, vice versa, that the Eastern Sudanic form had, at one point, become expanded through the addition of some sort of coronal suffix. There is also no question here that these forms should be scored differently from Nilotic/Surmic/Jebel, on one hand, and Berta, on the other.

Let us now take a different example, one that illustrates how "traces of regularity" can influence the scoring. In Khoisan languages, the word for 'star' is represented in the Northern (Ju) and Southern (!Wi-Taa) families by two roots that are significantly different as to their segmental structure:

— North Khoisan: Ju|'hoan ǂũ, Ekoka !Xũ !!ũ, etc. ← Proto-NK *ǂũ (ǂ = alveolar click);

— South Khoisan: !Xóõ ǁona, N|u ǁqʔoe-si, etc. ← Proto-SK *ǁ[ʔ]o- with different suffixes (actually, not fully clear if !Xóõ and N|u forms themselves are related, but our main concern here is !Xóõ; ǁ = lateral click).

The biggest obstacle that prevents us from scoring NK and SK as cognate forms is the difference in click articulation, which cannot be overlooked, since clicks are as different from each other as "regular" consonants with different manners of articulation. Cf., however, the following additional comparisons, relatively easy to come by: Ju|'hoan ǂaʔu 'cold' : !Xóõ ǁaʔũ id., Ju|'hoan ǂoe 'young man' : !Xóõ ǁquV 'new, young', Ju|'hoan ǂaʰ 'old (of things)' : !Xóõ ǁahã 'old, mature'. These (and other) examples — impeccable from the semantic side and quite convincing phonetically as well — show that, despite the dissimilarity, there is reason to consider this set as displaying traces of regularity. The obstacle is, therefore, overcome, and we can safely score the forms for 'star' as cognate.

It is important to stress that the requirement of *traces of regularity* is more lax than that of a complete *system of regular correspondences*, but should not be underestimated. The principal difference is that finding traces of regularity does not require us to thoroughly explore *all* the lexical evidence of the compared idioms and present a detailed reconstruction. But it does require us to demonstrate that our comparison is not completely ad hoc. It is not enough to take Proto-Japanese *pa* 'tooth' and compare it with Proto-Dravidian *pal* id., saying "final -l probably got lost in Proto-Japanese"; at the very least, it is necessary to find and quote several other transparent examples in which Japanese loses its final or intervocalic *-l-* compared to the rest of Altaic, such as Japanese *á-* 'receive' = Tungus-Manchu *al-* id., *kə̀-* 'to come' = Turkic *gel-* id., *kái* 'hair' = Turkic *Kïl* etc. (examples quoted from [EDAL]).

Obviously, *scoring* two or more forms as 'cognate' based on PL-related considerations of similarity or compatibility is not the same as demonstrating " beyond reasonable doubt" that said forms are cognate. Nevertheless, if this procedure is relatively strictly adhered to, it is to be expected that mistakes in scoring will be reduced to a minimum, and, furthermore, their negative effect will decrease in direct proportion to the number of language families enlisted in the scoring, since a global perspective will tend to "even out" individual distortions.

---

[17] Ehret [2001: 282] finds the correspondence between Berta *meera* and the East Jebel forms (but not the Nilotic ones!) to be regular, reflecting Proto-Nilo-Saharan *ḷ (the entire root is reconstructed as *mɛ́:ḷ 'to lick'). However, I have been unable to find any other satisfactory examples for this correspondence, and have every reason to doubt its regularity (unfortunately, similar situations arise with a great many more examples of particular correspondences given in this work, which cannot be said to give a reliable account of Proto-Nilo-Saharan historical phonology).

5. The issue of synonymity on micro- and macro-levels

One major problem that has pursued lexicostatistics and glottochronology from the very beginning is that of choosing, for a particular language, the correct equivalent for each item on the Swadesh list — and sometimes realizing that a single choice is all but impossible to come by, since "for many items on the list, languages often have more than one neutral equivalent" [Campbell 1998: 181].

This problem is very frequently exposed in works that are critical of lexicostatistics, sometimes in a very grave tone, as if its very existence automatically rendered the whole method useless. In reality, there are multiple reasonable ways to overcome it. For instance, S. Starostin, in all of his writings and calculations, advocated to disregard the issue as such and simply include both (or even *more* than both) synonyms in the calculations; e. g., if, for a particular item, language 1 yields synonymous lexemes A and B, and language 2 yields B and C, the situation should be qualified as "lack of replacement", since at least one out of two different synonyms is the same in both languages.

This solution is highly practical, but may create an uncomfortable illusion of "lack of rigor". Alternatively, one can simply tighten the demands by more precisely specifying the semantics of the "Swadesh notions", whose principal flaw arguably lies in their having been originally rendered in standard English, thus reflecting all the ambiguities of that language. E. g., a word like 'hair' is quite problematic, since it can be understood in at least three different ways: (1) 'hair' as material, i. e. 'wool, body hair'; (2) 'hair' as collective 'head hair'; (3) 'hair' as a singulative noun, 'one hair'. Quite a few languages have a different root for each of the three meanings, and entering them all as synonyms would clearly be excessive. The "default" (i. e. most frequent) usage would probably be (2), and this is the more precise meaning that I would advocate for the word — but it would be hard to get linguists all over the world readily agree upon one universally approved semantic standard[18].

Nevertheless, for the purposes of our global PL enterprise, conducted in accordance with a single standard, all of these technicalities are easily overcome, so that the issue of making the right choice with historically attested languages will depend exclusively upon the quality of known lexical descriptions for these languages.

In our situation, however, there exists a much more serious and important problem that also has to do with synonymity: selection of the appropriate synonym for the protolanguage form, both on low levels that serve as the starting nodes in our tree and on higher ones. The seriousness of this problem, in fact, goes way beyond the needs of lexicostatistics, as it is directly tied in with the whole issue of *semantic reconstruction* in historical linguistics — a sphere that, even today, is still barely tapped, despite certain theoretical breakthroughs, achieved above all in the works of J. Trier [Trier 1981] and in A. Dybo's monograph on semantic networks [Dybo 1996].

Even limiting ourselves to low-level reconstructions and a total of 50 most stable items, we will frequently fall upon cases where it is difficult, or even impossible, to ascertain one particular choice over the other (or, perhaps, even more than the other *ones*). Only in two types of situations do we find ourselves in a relatively secure position; these types have been explicitly formulated in [Kogan 2006], an article specifically dedicated to the issue of reconstructing a reliable wordlist for Proto-Semitic, but whose methodology is equally applicable to any other language family:

---

[18] Several recent sessions of the Nostratic seminar were dedicated to this particular issue, and a paper suggesting a set of more precise specifications for meanings on the Swadesh list — based on setting these meanings within particular sentential contexts — is under preparation by A. Kassian.

"If a PS (Proto-Semitic — *G. S.*) root functions with the same basic meaning in all Semitic languages, there is hardly any reason to doubt that it did so also in the proto-language… the same conclusion can be safely achieved if the root in question lost its basic function in a limited number of languages or minor subdivisions… finally, if a term is lost in some languages of a minor subdivision but persists in others, its archaic status is strengthened" (p. 465);

"if a PS root functions as the main term for the respective basic notion in several geographically distant languages without special genealogical proximity, it is likely that this meaning goes back to the proto-level. In this case, too, it is usually preserved as peripheral in other languages and, importantly, no alternative basic term suggests itself" (p. 474).

Based on the first criterion, Kogan is able to reliably fill in 39 slots on the 100-wordlist; based on the second, he adds 12 more, bringing the total up to 52. Even without looking, I can reasonably predict that significantly more than half of these words will belong to the 50-item wordlist specified above, and, indeed, 38 of Kogan's semantically reliable Proto-Semitic items coincide with elements on that "ultra-stable" half of the Swadesh wordlist. Since, in general, I agree with both of Kogan's criteria, this means that, for our PL procedure, the problem of choosing the correct entry for (at least) low- and mid-level reconstructions will not be a critical one.

Nevertheless, we still have to find some way to deal with the remaining 12 items, i. e. cases where descendant languages display way too much variability in order to allow for an unambiguous reconstruction. First, it is quite possible to add a few more *internal* criteria that may raise the chances of a particular choice. These include:

(a) The criterion of *internal etymologization*: if we have a choice between two items, one of which shows a clearly derived (most likely, recently derived) semantics, while the other one does not, it is the second item that has a better chance of preserving the protolanguage state.

For instance, in trying to establish the proto-root for 'meat' in Samoyed languages, we find that the main South Samoyed form (Selkup *wẹči*, Kamassian *uịa* ← Proto-Samoyed **åjå* [Janhunen 1977: 17]) differs from the main North Samoyed form (Nganasan *ŋȯmsu*, Enets *ud'a*, Nenets *ŋamza* ← Proto-Samoyed **ǝmså* [Janhunen 1977: 15]). Without any additional information, selection of the more representative variant is impossible. However, we have every reason to think, following Janhunen, that **ǝmså* is, in fact, a nominal derivative from the verbal root **ǝm-* 'to eat' [ibid.]. There is still a chance, of course, that **ǝmså* had already been formed and acquired the meaning of 'meat' on the Proto-Samoyed level, after which a root **åjå*, of unknown origin, mysteriously replaced it in Proto-South Samoyed; but since we have no clue as to where **åjå* actually came from, yet have every clue for internally etymologizing **ǝmså*, it is more reasonable to think of the former as an archaism and of the latter as an innovation[19].

(b) The criterion of *polysemy*: if one of the roots has several different meanings across languages, while the other one only has the "Swadesh meaning", this may mean — although it also depends on the representativeness of both forms — that the latter is the more archaic. Case in point: Lettish *jaûns* means either 'new' (of a thing) or 'young' (of a person), whereas in Lithuanian *jáunas* is used exclusively to denote 'young' (people), and in the "Swadesh meaning" of 'new (thing)' we have the more archaic *naũjas*.

(c) The criterion of *borrowing*: if we can reliably show that one of the competing roots is a borrowing from a distantly related or non-related language, this obviously raises the chance of

---

[19] A more detailed analysis shows that both lexemes can actually be traced back to the Proto-Samoyed level, since we also find Selkup *apsị* (← **ǝmså*) in the meaning 'food; body', as well as Enets *aija* (← **åjå*) 'flesh' (not the default Swadesh notion of 'meat', for which *ud'a* is used, as specifically indicated in the Uralic wordlists compiled by E. Helimski). This only confirms the conclusion reached without considering this additional evidence.

the non-borrowed item. Examples are numerous; cf., e. g., the abovementioned case of Tamil *nakam* 'fingernail' = Malayalam *nakham* id., both forms replacing the older root *ukir* = Kannada *ugur*, Tulu *uguru* etc. Since the Tamil and Malayalam forms are transparent borrowings from Indo-Aryan, this leaves Proto-South Dravidian **ugir* as the likeliest candidate for 'fingernail' at that stage.

Nevertheless, all of these criteria have a significant drawback: the *reverse* situation, in all three of these cases, is not much less probable. It is not at all excluded that derivation, polysemy, or borrowing could have already been present at the proto-level of the families that we are dealing with, and that new roots were introduced into specific subgroups later, obscuring the situation. Such solutions are, overall, uneconomical, prompting us to set up extra "dark horses" that are, in fact, unnecessary (such as, e. g., an obscure "para-Samoyed" substratum that donated the root **åjå*), but they cannot be excluded.

This means that the most important criterion for settling ambiguous cases must be the *external* criterion, which we may formulate as follows:

Where two or more equal or near-equal choices are possible for the proto-item, strong priority is given to one that demonstrates the most reliable external genetic connections.

Let us illustrate this on an example from the Germanic group. Germanic languages have a wide variety of roots for the notion 'meat': Scandinavian **kiut-* (→ Icelandic *kjöt*, etc.), West Germanic **flaiska-* (→ Dutch *vlees*, German *Fleisch*, cf. also English *flesh*, etc.), English *meat* = Old Norse *mat-r* 'meal', etc. However, out of all this variety, unquestionably the best candidate for Proto-Germanic 'meat' would be the ancestor of the Gothic form *mimz* — even though, apart from Gothic, neither the form itself, nor even any different forms with the same root have been attested in any other Germanic language.

The reason, of course, lies in the external connections of *mimz*: it is a perfect phonetic and semantic match with such forms as Old Indian *māṁs(a)-*, Armenian *mis*, Albanian *mish*, and Proto-Slavic **męso*, all of them related and pointing to Proto-Indo-European **mems-* as the original form. Assuming that **mimz(a)-* continued to be used in that function in Proto-Germanic, we conclude that it was preserved in the Gothic branch of this family (apparently, until the very end, cf. Crimean Gothic *menus* id.), but replaced by different other roots in the other branches. Assuming the opposite — that it is Gothic *mimz* that represents a semantic innovation — we would have to conclude that Proto-Germanic lost the original semantics of the Indo-European root, and then *restored* it in the case of Gothic: a highly unlikely situation, very rarely (if ever) observed in or surmised for the world's languages.

There is one obvious and significant problem with this criterion: if it is our *aim* to use PL as a means of verifying hypotheses on language relationship and establishing a global classification of the world's languages, how can we allow ourselves to use external data as if we already knew everything about these relations? Let alone Indo-European, how is this criterion supposed to work in areas such as America or Papua, where external connections even on relatively low time depths have been studied so poorly? And is this not, overall, a typical example of poorly masked circular logic?

It goes without saying that the external criterion has to be applied very carefully. The best, and most certain, type of situation in which it may be employed is a sort of "bootstrapping" mode, in which "proto-list" reconstruction and cognate scoring is achieved in two stages. First, we only populate those slots on the list for which internal data suggest a non-ambiguous candidate, leaving the problematic slots empty. Then we run the first stage of preliminary scoring, establishing its likeliest external relatives. *After* this has been achieved, we can now use exter-

nal data to try to solve the internal problems of the low-level family, i. e. populate its "dubi-ous" slots with those roots that better fit in with the external data.

In the case of Germanic, for instance, we have little methodological reason to worry about the selection of *mimz(a)-* as opposed to, e. g., *flaiska-*, simply because the unambiguous en-tries on the Germanic list — of which there are plenty — clearly demonstrate the Indo-European character of Germanic. Other situations may not be as immediately transparent, but careful application of this "two-step" principle is possible practically in all cases.

Of course, it may — and will — frequently happen so that the external criterion is unable to help us as well, if *none* of the candidate items have any significant external matches. In the same Germanic subgroup, for instance, there are at least four or five different roots denoting 'tail', but not a single one has any serious 'tail'-type parallels in other branches of Indo-European (almost all of which have their own problems with this infamously unstable — in Indo-European — notion). This means that neither internal nor external data allow us to make a choice. In this case, for *internal* needs we should leave the slot open, but for *external* needs we may choose any of the forms — it does not make a difference whether it is *swanka-* (→ Ger-man *Schwanz*), or *tagla-* (→ English *tail*), or *xalēn* (→ Icelandic *hali*), because, regardless of our choice, we will have to count it as a non-match with the rest of the Indo-European subgroups.

We now come to a less obvious, but equally challenging issue that awaits us on levels of "middle" time depth (such as Indo-European or Semitic), and even more so with macro-family relationships. Since we are establishing our classification "rung by rung", it is important to establish the likeliest candidates for proto-items on every level, i. e. figure out such a candidate for Indo-European before starting to probe Nostratic, and for Semitic before starting to probe Afro-Asiatic.

In order to do this, we accept Kogan's criteria as quoted above, and expand them with several internal criteria (also quoted above). Note, however, that the second criterion has an important catch: "…importantly, no alternative basic term suggests itself". What if, however, an alternative basic term *does* suggest itself?

Let us suggest that we have a language family descended from proto-language L, con-sisting of four branches: A, B, C, D. Out of these four, for a certain Swadesh item N on our list branches A and B share one cognate (let us call it *X), whereas branches C and D share a dif-ferent one (let us call it *Y). Let us now suppose that we have already run through the first stage of scoring for the entire family. If the resulting tree structure looks as follows:

$$L$$

$$A_X \quad B_X \quad C_Y \quad D_Y$$
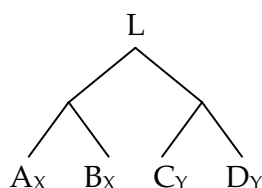
— this is in full agreement with our information on item N. In this case, internal data are con-sistent, although we will have problems understanding which of the two roots — *X or *Y — has to be posited at the top node; in order to do this, we will probably have to resort to exter-nal data. However, it is quite possible that our overall tree structure, based on an overall as-sessment of the lexicostatistical data, will look quite differently, e. g. the following way:

$$L$$

$$A_X \quad C_Y \quad B_X \quad D_Y$$
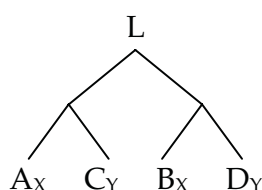
Such a tree would not be in very good agreement with the behaviour of *X and *Y, and would require one of four historical explanations:

(a1) *X and *Y were easily interchangeable synonyms in protolanguage L, as well as the intermediate protolanguages for AC and BD. The situation changed drastically only after the second split, with each of the four new branches "wiping out" one of the synonyms. The four "eliminations" could have been completely and utterly independent, or

(a2) the result of two areal lexical isoglosses that caused the loss of *X in geographical area AB and the loss of *Y in geographical area CD.

(b1) The regular word for notion N in protolanguage L, as well as the intermediate proto-languages for AC and BD, was *X, whereas *Y was semantically close, but not an exact synonym (or vice versa). After the second split, *Y replaced *X in branches C and D, but not in branches A and B. The two replacements could have been completely and utterly independent, or

(b2) the result of an areal semantic isogloss that affected the (supposedly contiguous) geographical area occupied by speakers of C and D, but not of A and B.

Needless to say, explanations (b1–b2) *by default* look more promising than explanation (a), since they require fewer assumptions (two independent or one common areal replacement vs. four independent or two common areal replacements). Moreover, explanation (a) requires us to set up freely interchangeable synonyms for Swadesh notions, a situation that is typologically rare and should better be avoided in reconstruction. Cases of such "semantic criss-crossing" are not frequent in non-controversial, low- or mid-level families, but they do exist, and it is strange that works on lexicostatistics have so far overlooked the existence of this problem.

A good actual illustration would be the word 'moon' in Indo-European languages. The most common and, undoubtedly, archaic root to express this notion is IE *mēns-, yielding Old Indian mā́s, Iranian *māh-, Baltic *men-, Slavic *měsęcь, Germanic *mēn- etc. [Pokorny 1958: 731–32]. However, Armenian lusin, Latin lū-nā, and certain Slavic forms going back to Common Slavic *lū-nā reflect a different root, usually — and with perfect reason — etymologized as IE *louk-s-nā, derived from the verbal root *leuk- 'to shine' and further compared with such forms as Avestan raox-š-na- 'shining', etc.

Trying to explain this as a common Armenian-Latin (or Armenian-Latin-Slavic?) isogloss is out of the question; "areal" explanation is excluded[20], and no other evidence exists to justify the postulation of a special "Armenian-Latin" node within Indo-European. This is, therefore, a typical example of "semantic criss-crossing", which we can attempt to solve in either of the two ways described above.

First, we can think of *mēns- and *louksnā as two freely interchangeable synonyms already on the Proto-IE level. This is, however, not realistic. Such a situation is not reflected in any of the attested descendant languages, which either only have one of two terms or feature a sharp semantic distinction between the two (as in Latin lūnā 'moon' vs. mensis 'month', or Russian луна 'full moon' vs. месяц 'crescent moon; month'). Even if we think of a possible stylistic dif-

---

[20] Unless, of course, we declare ourselves adherents of the strongest version of the "wave theory", according to which "Proto-Indo-European" as such never existed as even a minimally coherent linguistic entity, and that all of its twelve or so main branches have *always* been, in some ways, distinct from each other, co-existing peacefully on a small piece of territory before dispersing. Such a scenario, rendering useless the very idea of a genetic tree (and replacing it with the much more trendy concept of a "network"), would allow for just about any "areal" iso-glosses between just about any two or more branches of Indo-European, but I regard it as completely absurd and unsubstantiated by hard evidence, more of an artificial "easy way out" of the need to unravel the complex web of genetic and areal isoglosses between different branches of Indo-European than a solid model that makes real historical sense.

ference — e. g., *mēns-* as the "neutral" word and *louk-s-nā* as a "stylized", "poetic" moon — this already surmises incomplete synonymy, as it is always stipulated that each slot on the Swadesh wordlist be strictly filled in with the most "neutral" item, and that stylistically embellished quasi-synonyms should be left out.

On the other hand, if we do think of such a difference, or, indeed, consider it in terms of the possible existence of a special compound *mēns louksnos* (or, in the feminine, *mēnsā louksnā*) 'shiny moon', i. e. 'full moon' (cf., for instance, Avestan *raoxšnəm måŋhəm* acc.), it becomes very clear how easily the formerly adjectival form could have independently replaced the former noun *mēns* in at least several branches of Indo-European. To this should be added the additional "polysemy pressure" — since *mēns* was used both in the meaning of 'moon' and 'month', its replacement in at least one of these meanings could have been anticipated.

Work on semantic reconstruction for mid-level "non-controversial" families shows that such "criss-crossings" are relatively rare. Generally, if one item is replaced in several branches, it tends to be ushered out by different roots, because for each item on the list at least several different paths of semantic evolution are possible, and the more such paths we know, the less is the probability that the same path will be independently selected by two or more languages.

Nevertheless, semantic typology shows that some paths are more frequent than others, and in such cases, we must be prepared to expect independent developments. For instance, the term for such a body part as 'ear' is, every now and then, all over the world, re-formed as a nominal derivative from the verb 'to hear' (= 'hearing-thing'). In Indo-European, there is little doubt as to the original proto-root for 'ear' — IE *ous-* — but in Tocharian, we find that old root replaced by such a derivative: Tocharian A *klots*, B *klautso* ← Proto-Tocharian *kleutsā(i̯ä)n-* [Adams 1999: 230] ← IE *k̂leu̯-* 'to hear'. Not surprisingly, we also find a similar (although morphologically slightly different) development in Celtic: cf. Irish, Gaelic *cluas*, Welsh *clust* etc. Does this mean that Tocharian and Celtic share a common node on the tree, or, perhaps, this should be considered a special "areal" Tocharian-Celtic isogloss? Hardly likely.

But the one area where the issue of "semantic criss-crossing" hits the hardest is, of course, macro-comparison. Taking advantage of the fact that semantic reconstruction is one of historical linguistics' weakest spots, macro-comparative lexicostatistics may, in dealing with a particular Swadesh item, take *any* root which has the appropriate Swadesh meaning in *any* of mid-level family A's subbranches (or, in fact, even in any of its individual languages) — and score it as a positive cognate with *any* root with the appropriate Swadesh meaning in *any* of mid-level family B's subbranches (provided, of course, that the scoring is sanctified by phonetic correspondences or phonetic similarity). This approach is more or less explicitly stated by S. Starostin for his lexicostatistical calculations for language of Eurasia: "I have chosen the following principle: a word can be used as representing a particular meaning in the protolanguage if it has exactly this meaning in at least one subbranch of the family" [Starostin 2007b: 807].

Frankly, I have the gravest doubts about the statistical validity of this approach. Suppose that, in a certain language, we have a pair of semantically close roots (e. g. 'fire' : 'light'; 'star' : 'shine'; 'bird' : 'fly'; 'head' : 'top', etc.), the second of which is easily liable to usurp the functions of the first at some future point in time. How high are the chances of at least two of its future descendants to effectuate that transition independently of each other? Obviously, the primary dependency is on the *number* of those descendants. In the case of ten — twelve branches of Indo-European, chances for independent unidirectional semantic change will be quite modest (and this is explicitly confirmed by the actual historical analysis of the Swadesh wordlist), but if we multiply that number by a factor of five or six (the number of large families that constitute Nostratic), these chances will increase quite rapidly. (This could relatively easily be illustrated with a probabilistic model).

Not coincidentally, even a brief survey of the comparative tables for lexical matches between nine mid-level families of the Old World (Indo-European, Uralic, Altaic, Dravidian, Kartvelian, representing the Nostratic macrofamily; Semitic, representing the Afro-Asiatic macrofamily; North Caucasian, Sino-Tibetan, Yeniseian, representing the Sino-Caucasian macrofamily), presented in [Starostin 2007b: 807–815], reveals a picture that can only be called "Synonymity on the Rampage": two, sometimes three roots for each Swadesh item within one mid-level family — and, consequently, three to five roots on average within one macrofamily — are the norm. The word 'sun' in Nostratic languages alone, for instance, is illustrated by (a) a match between Indo-European *seHw- and Altaic *si̯àgu; (b) a match between Uralic *pVjwV and Altaic *pʿi̯agV; (c) a match between Altaic *nèra and Dravidian *ńejir-. Should this be historically interpreted as reflecting three freely interchangeable synonyms for 'sun' in Proto-Nostratic (and, further down, in Proto-Altaic)? Apparently not. In order to admit such a possibility, we should either find some typological support for it on less remote time scales — in all likeliness, an impossible task — or suggest that language speakers in pre-Neolithic times had a far more liberal attitude towards synonymity than their descendants, being accustomed to freely sharing two or three words for each meaning. This, however, would simply plunge us into the world of fantasy[21].

Let us look at this situation with 'sun' more closely. The three matches, as can clearly be seen, are determined by the three roots in Altaic — itself a "near-macro-family", still controversial among mainstream linguists. I do not doubt the existence of Altaic — evidence for a special relationship between Turkic, Mongolic, Tungusic, Korean, and Japanese is too overwhelming to make room for skepticism — but I will be the first to admit that this evidence is in dire need of further filtering and refining, and that one of its major problems is the lack of a detailed semantic reconstruction.

The three mentioned Altaic roots for 'sun' are not, in fact, "Altaic": they are rather the main roots to denote this object in separate subdivisions of Altaic. Proto-Altaic *si̯àgu (newer reconstruction is actually *si̯ŏgu) is reflected as Tungus-Manchu *sigū-n 'sun' and Korean *hǎi id., with a possible further correlate in Japanese *suà-rá 'sky' [EDAL: 1274]. Proto-Altaic *pʿi̯agV is reflected as Japanese *pí 'sun', but also Korean *pài 'dawn', Tungus-Manchu *pigi 'to warm (smth.), warm oneself', and Mongolic *heɣe- 'to heat, be heated' [EDAL: 1147]. Finally, Proto-Altaic *nèra (*ŋèrá in EDAL) is reflected as Mongolic *nara-n 'sun', but also Turkic *i̯ạr-ï̇n 'morning; tomorrow', Tungus-Manchu *ŋēr(i)- 'light', Korean *nár 'day (24 hours); weather', and Japanese *àrí- 'dawn' [EDAL: 1028].

Out of these three roots, only *si̯àgu has the meaning 'sun' in at least two branches of the family, and it is interesting to see that the Japanese parallel shows a suffixal extension, indi-

---

[21] The existence of this problem was well realized by S. Starostin himself, who wrote: "the "protolanguage synonymy" may produce a higher number of coincidences and make the dates of separation somewhat younger" [Starostin 2003: 465]. He, however, believed that the negative effects of this kind of scoring may be counterbalanced and cancelled by a reverse factor: "the impossibility of identifying loanwords may result in an earlier date of divergence (according to the standard procedure adopted by us, a mismatch caused by the borrowing is not taken into consideration; consequently, if loanwords cannot be detected, the percentage of coincidences between the proto-languages becomes lower)" [ibid.].

Perhaps for the full 100-wordlist this may, to a certain degree, be true. But when we pare it down to 50 most stable items, the loanwords issue loses much of its significance, since these items, by default, are expected to contain an absolute minimum of loans (see below). The synonymity issue, on the other hand, is equally disturbing for any version of the list, and I am afraid that, in "macro-calculations", adoption of a liberal stance on synonymity will inevitably result in an exaggerated number of matches between families and, consequently, younger dates of separation for macro-units like Nostratic or Sino-Caucasian.

cating that the original meaning of *suà-rá may have been something like 'sunny skies'. In very sharp contrast, the two other roots have only gained the meaning 'sun' in one branch each, and show a very different type of semantics elsewhere. In fact, a comparison between *pigi 'to warm' and *pí 'sun' is hardly imaginable *unless* the original semantics was that of 'heat', because the semantic development 'sun' → 'warm' is typologically unprecedented (at the very least, I have been unable to encounter any reliable examples in EHL's huge collection of data). Likewise, *ŋèrá is easily understood as an original 'day, light time period', but hardly as an actual designation of the celestial body.

The likeliest candidate for an original Proto-Altaic 'sun' is, therefore, only *si̯àgu — for the other two roots, none of the possible scenarios are credible from the point of view of semantic typology. How does this reflect upon the Nostratic comparison? Fairly well: as suggested originally, *si̯àgu is a solid match for Indo-European *seHw-, or, more traditionally, *sāw-el- ~ *sw-en- with fluctuating suffixal extensions [Pokorny 1959: 881–882].

But what of the other two matches, with Uralic and Dravidian respectively? The interesting thing here is that, while Indo-European *sāw-el- ~ *sw-en- is, indeed, unquestionably the primary Indo-European root for 'sun', the same cannot be said neither of Uralic *pVjwV nor of Dravidian *ńejir-. The former, as a polysemous 'sun; day', is the main root in Balto-Finnic and Lappic (Finnish päivä, Estonian päev, Saami bæi've, etc., see [Rédei 1988: 360]), but not anywhere else. The latter, reconstructable as *ńējir̯ or *ńēsir̯, is seen only in the South Dravidian subgroup (Tamil ñāyir̯u, nāyir̯u; Kannada nēsar̯; Tulu nesur̯u 'morning'; Toda nȫr̯ 'sun (only in songs)') and, perhaps — although the phonetic correspondences are dubious — in North Dravidian, with different semantics (Malto nīr̯u 'sunshine, heat'); see [DEDR: 252]. It is certainly a far less likely candidate for Proto-Dravidian 'sun' than the far better represented *por̯ud- [DEDR: 403][22].

By applying nothing but the basic, simplest principles of semantic reconstruction, we have managed to show that, out of these three instances of 'sun' in Nostratic, there is really *one* strong case — strong on all sides — and *two* weak ones — weak on all sides. Note that the *etymologies* as such have not been killed off (at least the Uralic-Altaic connection is still relevant), only their lexicostatistical significance. The evidence in favor of Nostratic has not been weakened; on the contrary, it has only become tighter, as the "evolutionary scenario" for Nostratic 'sun' is now more comprehensible and realistic.

There does, however, remain the issue of scoring. We have more or less certified that Proto-Uralic *pVjwV did not necessarily have the meaning 'sun', and that Proto-Altaic *pʼi̯agV almost certainly did not have this meaning. However, our list of proto-languages does not include Altaic and Uralic; the starting nodes are the smaller subgroups that constitute these two large families, and these happen to include Balto-Finnic, where the root for 'sun' is *päivä, and Proto-Japanese, where it is *pí. They generally satisfy the requirements for phonetic correspondences in Nostratic languages, and are quite compatible phonetically even without knowing these correspondences — yet they, most likely, do not go back to the respective Proto-Altaic and Proto-Uralic roots for 'sun'. Should they be scored as cognates or not?

From an etymological point of view, they *are* cognates — reflecting independent similar semantic development out of an older meaning — and should be scored as matches. However, the epistemological definition of a "match" on the Swadesh list would necessarily surmise the

---

[22] Actually, if the Altaic root *ŋèrá is really to be reconstructed with a temporal meaning ('bright period of day'), a much better parallel in Dravidian is Tamil nēram 'time, season, opportunity', Koḍagu nēra 'time, sun (!)', Tulu nēr-ḍè id., possibly (although loss of final -r is irregular) also Brahui dē 'sun, sunshine, day, time' [DEDR: 337] — still not the main Proto-Dravidian root for 'sun', but a very interesting semantic match all the same.

idea of either *common retention* (the word continues, substantially unchanged, to perform the original function as such in descendant languages) or *common innovation* (the word shifts from its original function in the intermediate language that serves as the specific common ancestor to languages displaying the innovation). In this particular case, as well as plenty of others, there is neither a common retention — chances of this word meaning 'sun' in Proto-Nostratic are minimal compared to other candidates — nor a common innovation (Baltic-Finnic and Japanese do not have an immediate common ancestor). Scoring *päivä* and *pí* as a match will, therefore, distort the overall calculation scheme, and, in combination with multiple other distortions of such sort, make the classification results less reliable.

On the other hand, it should not be forgotten that notions such as "Altaic", "Uralic", "Nostratic", etc., already surmise a pre-established idea of branching, and that we run the risk of succumbing to circularity if we modify our scoring results based on preconceived ideas of classification. Moreover, for linguistic areas in which there are no preconceived ideas of classification, or these ideas are at an embryonic stage (= much, if not most of the linguistic world outside Eurasia) such modifications will be impossible in principle. How should we proceed?

I suggest, once again, a return to "bootstrapping" mode. During the *first* stage of calculations our main goal is to establish the primary "linguistic building blocks" — perform a rough attempt of grouping a large number of families into a smaller number of higher-level units. In the case of Eurasia, this attempt will, without a doubt, let us see all of its principal families — Indo-European, Uralic, Altaic, Dravidian, Sino-Tibetan, Semitic, Austro-Asiatic, etc. — as well as indicate possible higher level connections between them. At this stage, it will be permissible to count *päjvä* and *pí* as (potential) cognates, because we have not yet certified the existence of such "blocks" as Uralic and Altaic.

Once the first stage is completed, we proceed to the second stage: fine-graining the results, using the "block" information we have accumulated as our basis. At this stage, our main task is to wipe out the "false leads", and this is accomplished through establishing, as precisely as possible, the *most likely* candidate for the given Swadesh notion at the top of each "block", i. e. for Proto-Indo-European, Proto-Uralic, Proto-Semitic, etc. By default, *only that particular item will be allowed to score as a positive match on the higher level of taxonomy*. All other matches will be eliminated, judged as either (a) chance similarities or (b) independent semantic innovations, even if the roots are related etymologically.

Let us demonstrate this on one more example, this time taken from the Sino-Caucasian sphere. In [Starostin 2003: 473], one of the proposed matches is North Caucasian *wĕnX̣V* 'head' vs. Sino-Tibetan *lŭH* id. This comparison satisfies S. Starostin's own system of phonetic correspondences between the two families (with regular reduction of the initial syllable in Sino-Tibetan) and, at the first stage of comparison, is acceptable. However, since both the "North Caucasian" and "Sino-Tibetan" labels are not quite allowed at this stage, it should rather be noted that the comparison is between (a) Proto-Lezghian *woX̣ul*, (b) Proto-Dargwa *beḳ*, (c) Lak (an isolated language) *baḳ*, (d) Khinalug (another isolate) *miḳir* (in other branches of North Caucasian the root is either missing or has such different meanings as 'beak; mouth; nose'; see [Nikolayev, Starostin 1994: 1041] for details), (e) Old Chinese 首 *s-luʔ*, (f) Kuki-Chin *lu* (Kuki-Chin is a large, but only one subgroup of Tibeto-Burmese; see [Schuessler 2007: 470] for the etymology). All these forms can be marked as cognates (even such superficially dissimilar forms as Lak *baḳ* and Kuki-Chin *lu*, since we have permission to use our knowledge about the internal and external historical phonology of these languages).

Once the primary stage has been completed, and the North Caucasian and Sino-Tibetan "blocks" established as firm taxonomic units, we run the second stage, checking the validity of *wĕnX̣V* and *lŭH* as the best respective candidates for Proto-NC and Proto-ST 'head'. First of

all, it should be noted that even the primary stage will clearly indicate a strong binary split in both cases: North Caucasian will be a combination of Northeast (Nakh-Daghestanian) Caucasian and Northwest (Abkhaz-Adyghe) Caucasian, and Sino-Tibetan — a combination of Sinitic (Chinese) and Tibeto-Burmese. Our ideal would be to see *wĕnX̌V represented in both the Northeast and the Northwest branches, and to see *lŭH in both Chinese and Tibeto-Burmese. The situation is, however, much more complicated.

NC *wĕnX̌V is not properly NC; it is only encountered as 'head' in several Daghestanian branches and is not necessarily even the best candidate for 'head' on that level. (In Andian and Tsezian languages the main root for 'head' reflects NC *hq̇wĕmV̄, and the default West Caucasian root is reconstructed as *SqIa). This is not a death blow, since it merely presumes that we are unable to reach a satisfactory conclusion based on internal evidence alone (see above).

But the situation is worse in the case of Sino-Tibetan. Here, semantic reconstruction strongly indicates that *lŭH may be an independent innovation in Old Chinese *and* Kuki-Chin — provided the roots are even related in the first place, and do not represent accidental look-alikes. The reason is that the primary root for 'head' in Tibeto-Burmese is not *lŭH, but *qhə̄wH (reconstruction following [Peiros, Starostin 1996]), reflected in a large number of subgroups: cf. Tibetan *m-go*, Burmese *u-h*, Sgaw Karen *khoʔ*, Garo *s-ko*, Pumi *khu*, Jiarung *ko* etc. (each language here represents a separate group). The idea that it is *qhə̄wH that represents an archaism and not *lŭH is further supported by its very likely cognate in Old Chinese: 后 *gō̆ʔ 'ruler, sovereign', suggesting a very usual semantic development from 'head'. The opposite transition 'ruler' → 'head' (as body part!) is not at all realistic.

Obviously, we should keep in mind that the general field of Sino-Tibetan etymology at its present state leaves a lot to be desired, and future research may yet show that *lŭH is, in fact, a more firmly grounded reconstruction than *qhə̄wH. But the current disposition is hardly in favor of that conclusion, and so, at the second stage of our cognate scoring, we should dispose of this match, since it fails to pass our criteria for choosing the most appropriate synonym.

It is very important to note that there *are* clear-cut cases when no single item can be unambiguously postulated for the "top of the block" position. The most typical situation here is that of a primary binary split, such as, e. g., Indo-European into Anatolian and "Narrow Indo-European" (or, in other terms, "Indo-Hittite" into Anatolian and Indo-European), Uralic into Fenno-Ugric and Samoyed, or North Caucasian into Northeast and Northwest Caucasian. In all such cases, whenever one has to reconstruct different roots for the same notion in each branch, both reconstructions carry the same "weight", regardless of their size and spread. E. g., "Narrow Indo-European" *onogh- 'fingernail' and Hittite *sankuwai-* id. have an equal chance of reflecting the original root for this notion, despite the fact that *onogh- is seen in at least seven different subgroups of Indo-European.

I predict a certain amount of criticism addressed at this methodology, and understand the main objection: the general inexperience of historical linguistics when it comes to strict semantic reconstruction, the usual uncertainties that we all feel about assigning one particular meaning to a proto-root whenever its descendants show even a slight amount of semantic variety. However, it is exactly this particular objection that makes me insist that the "no synonyms!" principle be applied and tested as rigorously as possible, if only for the reason that we all have to learn to perform strict semantic reconstruction, sooner or later, and that if there is one good place to start with it, it is the Swadesh wordlist. A global lexicostatistical database with an emphasis on semantic change would, in addition to its general goals, serve as an excellent foundation for all sorts of systematic studies on historical semantics.

Finally, a consistent application of the "semantic filter" would, hopefully, help dissipate the major accusation against global-scale lexicostatistics — namely, that the more languages

are added into the pot, the more chances we have of getting accidental look-alikes. Obviously, this accusation is true if we place no limits on "criss-crossing" — score one "Proto-Indo-European" synonym for a given item as a match with Uralic, another one as a match with Dravidian, a third one as a match with Old Chinese, and a fourth one as a match with North Halmaheran. But if it can be shown, for instance, that the best matches between Indo-European and Uralic are *truly* Proto-Indo-European *and* Proto-Uralic — most likely candidates for the proto-roots in both families — this leaves no space for such accidence.

6. Contacts, Contradictions, and Conclusions

In the three previous sections, we have attempted to describe the main methodological principles that should, in our opinion, guide the process of constructing a global lexicostatistical database for the world's languages. Their chief differences from previously employed techniques may be briefly summarized as follows: (a) use of a compact, ultra-stable 50-item wordlist with low-level reconstructions serving as the main entries; (b) use of a "mixed" scoring procedure, based on phonetic correspondences where they have been established and "phonetic compatibility with traces of regularity" where they have been not; (c) very strict limits on synonymity both on low, mid and deep chronological levels; (d) a "recursive" approach to scoring, where the first round of calculations is followed by a "fine-graining" round, weeding out false matches with no historical reality behind them.

A careful application of all these conditions, particularly (b) and (c), will minimize the number of accidental similarities in our calculations. But will it be able to neutralize the problem that we described at the very beginning of the paper — the risk of mistaking contact lexicon for genetic cognates? Obviously, words could be borrowed into proto-languages as easily as they can be borrowed into historically attested languages (so strict limitations on synonymity are not necessarily a safeguard), and if the borrowed strata are large enough, they always display "traces of regularity".

It would be an exaggeration to say that the proposed method is sufficiently robust to let us, in each and every type of imaginable situations, avoid the "contact trap". Nevertheless, there are two main considerations that make it significantly more waterproof than other methods of classification.

The first one is the choice of the wordlist. None of the 50 items — not even personal pronouns — are 100% immune to borrowing, but, in general, this "core" is much more resilient to being replaced by words of foreign origin than even the remaining half of the Swadesh wordlist. Having analyzed (preliminarily) the 50-item lists for approximately 200 low-level families of Eurasia and Africa, I have been able to detect only three explicit cases in which borrowings amounted to about 1/5 (10–11 items) of the entire list: these were Brahui (one-language group within Dravidian), Albanian (one-language group within Indo-European), and Northern Songhay (a small cluster of closely related dialects with a very heavy Berber influence; Southern Songhay is much more conservative).

Furthermore, Brahui displays a hodge-podge of borrowings from different sources (Indian, Persian, Arabic) that outcancel each other, and some of the alleged "borrowings" from Latin on the Albanian list are etymologically questionable and may actually represent inherited retentions of original Indo-European roots. This leaves the Songhay dialects as just about the only transparent example where one could really make a mistake (provided one had no access to supporting data from Southern Songhay) — and there is no reason whatsoever to think that this ratio of 1 to 200 must have been seriously different ten or more thousand years ago.

The second consideration is one of *context*. Let us suppose that we are running the first stage of calculations and have no idea of the genetic status of the Brahui language. In this case, we may want to score Brahui *haḍ* 'bone' as cognate with Old Indian *asthi*, Brahui *dandān* 'tooth' as cognate with *dant-,* and, perhaps, Brahui *draxt* 'tree' (although this is a Persian, not an Indian word) as cognate with *daru*. This will give us three false matches that will, nevertheless, be overridden during the tree construction process by the overwhelming number of true matches that Brahui has with the other Dravidian languages. Noticing the sharp increase of Brahui matches with Indo-European, even though the suggested classification clearly puts it with the rest of Dravidian, we will then — at the second, "fine-graining" stage — count the Brahui forms as borrowings (excluding them from calculations), since a true close relationship with Indo-European would require an equally sharp increase in cognation rate between every branch of Dravidian and every branch of Indo-European.

Similar analyses will easily help us weed out false matches between North Songhay and Berber, Fenno-Ugric and Indo-Iranian, Kartvelian and North Caucasian, etc. Counting these pairs of language groups as sharing a close genetic relationship will be out of the question because each of their elements will have a much stronger "attraction" on the part of its true closest relative.

If, on the other hand, potential cognates are found between the respective protolanguages A and B in their "blocks", and no "stronger" genetic affiliation is suggested between protolanguage A and, for instance, protolanguage C, this should be — by default — considered as indicative of deep-level relationship. "By default" here means that, if we want to interpret such a situation as reflecting contacts, the burden of additional proof here lies on the "arealist", not on the "heritagist".

Example: for Indo-European and Uralic, we find such serious matches on the 50-item list as IE **me-* : Uralic **mE* 'I', IE **tu* : Uralic **tE* 'thou', IE **k̂leu̯-* : Uralic **kule* 'to hear', IE **(H)nom-* : Uralic **nime* 'name', IE **wed-or* : Uralic **wete* 'water', IE **kʷi-s* : Uralic **kU* 'who' (several other, less obvious, cognates will be discussed in further publications on the subject). Similarly strong cognation suggestions also exist between IE, Uralic and some other language families that constitute the traditional "Nostratic", but none of them *override* this evidence quantitatively.

Interpretation of these matches in terms of prehistorical contacts is not entirely ruled out, yet, based on our empirical knowledge about contact situations around the world as well as common sense, is significantly less likely than its interpretation in terms of prehistorical genetic relationship. If the "arealist" thinks otherwise, it is up to him/her to provide additional evidence, preferably in the form of at least *dozens* (if not *hundreds*) of terms in the cultural lexicon, borrowed from Proto-IE into Proto-Uralic or vice versa — a condition that is, for instance, very easy to satisfy in the cases of Brahui, Albanian, and North Songhay. Until this is done, the default working model will be that of genetic relationship between Indo-European and Uralic[23].

Before concluding this discussion, three more small, but important technical points should be made on certain procedural aspects of PL:

1. As mentioned above, *glottochronological* interpretation of the results — with absolute dates of splitting accompanying the classification — is not obligatory, but is nevertheless use-

---

[23] Of course, there always remains the problem of the so-called "mixed languages" (pigins, creoles, etc.), whose existence in prehistoric times can be questioned, but not ruled out. Nevertheless, there are reasons to think that "contextual" considerations such as described above will help us single out and correctly identify such situations as well. For a detailed discussion on the identification of possible "creoles" in lexicostatistical databases, see [Burlak 2006].

ful for those who accept glottochronology as a valid method. However, basing the glottochronological calculations on the old Swadesh quotient of 0.14 or Starostin's "improved" quotient of 0.05 will be inadmissible, since these rates have been calibrated based on the average stability of the entire 100-wordlist, not its more stable half. We, therefore, either have to recalibrate the quotient — obviously, its value will be somewhat less than 0.05 — or, better still, rely on Starostin's "experimental" method with individual rates for each item on the list (see fn. 6).

2. It is evident that, no matter how tight we make the rules on scoring, in quite a few cases we will be presented with several alternatives of equal or near-equal probability, sometimes affecting classification results in a serious manner. (Within Indo-European, for instance, Albanian is a particularly difficult case; its position on the tree may depend on as little as one or two questionable etymological decisions). For such cases, it makes sense to consider all the alternate paths of scoring and present all alternate models; additional data will then be necessary to make a more precise choice.

3. Although the principal work should be conducted manually, this does not mean that fully automatic procedures — such as have been described in section 4 — are out of the question; on the contrary, it would make perfect sense to combine manual and automatic handling of the data. Similar results will strengthen the conclusions, while discrepancies may clearly indicate problematic areas in the manual handling as well as help refine the automatic algorithms.

The detailed description of the PL procedure in this paper would, of course, not be possible if the procedure itself still existed only in theory. As it is, 50-item lists have already been compiled by the author of this paper — and are, at the moment, collectively verified and modified at regular sessions of the Nostratic seminar at RSUH's Center for Comparative Linguistics — for most of the families and sub-families that constitute the traditional "Nostratic", and are now being compiled for subdivisions of "Afro-Asiatic" and "Sino-Caucasian".

Sergei Jaxontov, in an overview article on glottochronology, once wrote: "It would be desirable to apply glottochronology among all established and tentative language families. As a result, language groups could be revealed with a maximum divergence of 60–80 (or, probably, 80–100) centuries, as well as language isolates beyond such groups. Also, realistic and comparable classifications could be proposed for each group" [Jaxontov 1999: 59]. With the massive amount of comparative data that members of the EHL project have managed to put together over the past eight years, we now have every possibility of carrying out this work on a more detailed and professional basis than was possible even a decade ago. It is, at present, unclear what the "time ceiling" will be for this kind of approach — whether it will be Jaxontov's "80–100" centuries or significantly deeper than that — but this really depends on "data behaviour" and can hardly be predicted.

The present paper lays down the basic methodological aspects of PL, yet its real value will only be evident on practice — with the actual discussions of the data for each individual "block" (family) and its comparisons with data from other "blocks". The paper is, thus, but an introduction to a series of publications (or, perhaps, a collective monograph) that I and other EHL members plan to dedicate to the presentation and analysis of the lexical data relevant for a PL-based global linguistic classification.

Appendix

The proposed 50-item wordlist for the global lexicostatistical database. Items are ranged according to their relative degree of stability. For some of the most ambiguous English lexemes, additional meaning specifications are given in parentheses.

1. we [24]	11. hand	21. one	31. mouth	41. leaf
2. two	12. what	22. tooth	32. ear	42. kill
3. I [24]	13. die	23. new	33. bird	43. foot
4. eye	14. heart	24. dry (*e.g. of clothes*)	34. bone	44. horn
5. thou [24]	15. drink	25. eat	35. sun	45. hear
6. who	16. dog	26. tail	36. smoke	46. meat (*as food*)
7. fire	17. louse (*head*)	27. hair (*of head*)	37. tree	47. egg
8. tongue	18. moon	28. water	38. ashes	48. black
9. stone	19. fingernail	29. nose	39. rain	49. head
10. name	20. blood	30. not [25]	40. star	50. night

Literature

Adams 1999 — Douglas Q. ADAMS. *A Dictionary of Tocharian B*. Amsterdam: Rodopi.

Aikhenvald & Dixon 2001 — *Areal Diffusion and Genetic Inheritance*. Ed. by Alexandra Y. AIKHENVALD and R. M. W. DIXON. Oxford University Press.

ASJP 2009 — André MÜLLER, Viveka VELUPILLAI, Søren WICHMANN, Cecil H. BROWN, Pamela BROWN, Eric W. HOLMAN, Dik BAKKER, Oleg BELYAEV, Dmitri EGOROV, Robert MAILHAMMER, Anthony GRANT, and Kofi YAKPO. *ASJP World Language Tree of Lexical Similarity*: Version 2. Available online at: http://email.eva.mpg.de/~wichmann/language_tree.htm (April 2009).

Bastin, Coupez, & Mann 1999 — Yvonne BASTIN, André COUPEZ, Michael MANN. *Continuity and Divergence in the Bantu Languages: Perspectives from a Lexicostatistic Study.* Human Sciences Annals of the Royal Museum for Central Africa, Series in –8°, 162. Tervuren: RMCA.

Baxter & Manaster-Ramer 2000 — W. BAXTER, A. MANASTER-RAMER. Beyond lumping and splitting: probabilistic issues in historical linguistics // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 167–188.

Bender 1998 — M. Lionel BENDER. The Eastern Jebel Languages of Sudan II. Comparative Lexicon // *Afrika und Übersee*, Band 81.

Bergsland & Vogt 1962 — K. BERGSLAND, H. VOGT. On the Validity of Glottochronology // *Current Anthropology*, 3; pp. 115 — 153.

Blažek 2006 — Vaclav BLAŽEK, Petra NOVOTNÁ. On Application of Glottochronology for Celtic Languages // *Linguistica Brunensia. Sborník prací filozofické fakulty brněnské univerzity* A 54, pp. 71–100.

Blažek 2008a — Vaclav BLAŽEK. On application of Glottochronology for Saharan Languages // *Viva Africa* 2007. *Proceedings of the IInd International Conference on African Studies* (April 2007), ed. by Tomáš MACHALÍK & Jan ZÁHOŘÍK. Plzeň: Dryáda, pp. 19–38.

Blažek 2008b: Vaclav BLAŽEK, Šárka KRPCOVÁ. On the Application of Glottochronology to Kartvelian languages // *Mother Tongue* 12, pp. 125–133.

Blust 2000 — Robert BLUST. Why lexicostatistics doesn't work: the 'universal constant' hypothesis and the Austronesian languages // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 311–332.

Burlak 2006 — Svetlana BURLAK. Kreol'skije jazyki i glottoxronologija [Creole Languages and Glottochronology] // *Aspekty Komparativistiki* 3 [*Aspects of Comparative Linguistics* 3]. Moscow, RSUH, pp. 499–508.

Campbell 1998 — Lyle CAMPBELL. *Historical Linguistics. An introduction*. Edinburgh University Press.

Christiansen et al. 2009 — Morten H. CHRISTIANSEN, Chris COLLINS, Shimon EDELMAN. Language Universals: A Collaborative Project for the Language Sciences // *Language Universals*. Ed. by Morten H. CHRISTIANSEN, Chris COLLINS, Shimon EDELMAN. Oxford University Press, pp. 3–16.

DEDR — T. BURROW, M. B. EMENEAU. *A Dravidian Etymological Dictionary*. Second edition. Oxford, Clarendon Press.

---

[24] For personal pronouns, as an official exception, synonymity is allowed on the list by taking both the direct and indirect stem of the pronoun into account if they are suppletive (e. g. *I — me*).

[25] Basic negation, particle or negative verbal stem/suffix.

Dimmendaal 1988 — Gerrit DIMMENDAAL. The Lexical Reconstruction of Proto-Nilotic // *Afrikanistische Arbeitspapiere*, 16, pp. 5–67.

Dolgopolsky 1986 — Aharon DOLGOPOLSKY. A probabilistic hypothesis concerning the oldest relationships among the language families of northern Eurasia // *Typology, Relationship, and Time: a Collection of Papers on Language Change and Relationship by Soviet Linguists*. Ed. by V. V. SHEVOROSHKIN & T. L. MARKEY. Ann Arbor, Karoma, pp. 27–50.

Dybo 1996 — Anna DYBO. *Semantičeskaja rekonstrukcija v altajskoj etimologii* [*Semantic reconstruction in Altaic etymology*]. Moscow, Institute of Linguistics.

Dyen 1965 — Isidore DYEN. *A lexicostatistical classification of the Austronesian languages*. Memoir 19, Supplement to the International Journal of American Linguistics 31.1. Baltimore: Waverly Press.

Dyen 1992 — Isidore DYEN, Joseph B. KRUSKAL, P. BLACK. *An IE Classification: A Lexico-Statistical Experiment*. Philadelphia.

EDAL: A. V. DYBO, O. A. MUDRAK, S. A. STAROSTIN. *An Etymological Dictionary of Altaic Languages*. Brill, Leiden, 2003.

Ehret 2001 — Christopher EHRET. *A Historical-Comparative Reconstruction of Nilo-Saharan*. Köln, Rüdiger Köppe Verlag.

Gell-Mann, Peiros, Starostin 2009 — M. GELL-MANN, I. PEIROS, G. STAROSTIN. Distant Language Relationship: The Current Perspective // *Journal of Language Relationship*, 1, pp. 13–30.

Georg 2003 — Stefan GEORG. Haupt und Glieder der Altaischen Hypothese: die Körperteilbezeichnungen im Türkischen, Mongolischen und Tungusischen // *Ural-Altaische Jahrbücher*, Neue Folge B, 16, pp. 143–182.

Gray & Atkinson 2003 — Russell D. GRAY, Quentin D. ATKINSON. Language-tree divergence times support the Anatolian theory of Indo-European origin // *Nature* 426, pp. 435–439,

Greenhill et al. 2008 — S. J. GREENHILL, R. BLUST, R. GRAY. The Austronesian Basic Vocabulary Database: From Bioinformatics to Lexomics // *Evolutionary Bioinformatics*, 4, pp. 271–283.

Güldemann 2006 — Tom GÜLDEMANN. Structural isoglosses between Khoekhoe and Tuu: the Cape as a linguistic area // Yaron MATRAS, April McMAHON and Nigel VINCENT (eds.), *Linguistic areas: convergence in historical and typological perspective*. Hampshire: Palgrave Macmillan, pp. 99–134.

Hübschmann 1875 — Heinrich HÜBSCHMANN. Über die Stellung des armenischen im Kreise der indogermanischen Sprachen // *Zeitschrift für Vergleichende Sprachforschung* 23, pp. 5–42.

Janhunen 1977 — Juha JANHUNEN. *Samojedischer Wortschatz: Gemeinsamojedischen Etymologien*. Helsinki.

Jaxontov 1999 — Sergei JAXONTOV. Glottochronology: Difficulties and Perspectives // *Historical Linguistics and Lexicostatistics*. Ed. by Vitaly SHEVOROSHKIN and Paul J. SIDWELL. Melbourne, pp. 51–59.

Kessler 2001 — Brett KESSLER. *The Significance of Word Lists*. Stanford, California: CSLI Publications.

Kogan 2006 — Leonid KOGAN. Lexical Evidence and the Genealogical Position of Ugaritic (I) // *Babel und Bibel 3*. Orientalia et Classica vol. XIV. Eisenbrauns, Winona Lake, Indiana, pp. 429–488.

Matisoff 2003 — James A. MATISOFF. *Handbook of Proto-Tibeto-Burman: System and Philosophy of Sino-Tibetan Reconstruction*. University of California Press.

Merwe 1966 — N. van der MERWE. New mathematics for glottochronology // *Current Anthropology* 7, pp. 485–500.

Militarev 2000 — Alexander MILITAREV. Towards the chronology of Afrasian (Afroasiatic) and its daughter families // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 267–307.

Nichols 1992 — Johanna NICHOLS. *Linguistic Diversity in Space and Time*. University of Chicago Press.

Nikolayev, Starostin 1994 — S. L. NIKOLAYEV, S. A. STAROSTIN. *A North Caucasian Etymological Dictionary*. Moscow, Asterisk Publishers.

Orel 2003 — Vladimir OREL. *A Handbook of Germanic Etymology*. Brill, Leiden-Boston.

Orel & Stolbova 1995 — V. OREL, O. STOLBOVA. *Hamito-Semitic Etymological Dictionary: Materials for a Reconstruction*. Leiden & New York.

Pagel et al. 2007 — M. PAGEL, Q. ATKINSON, A. MEADE. Frequency of word-use predicts rates of lexical evolution throughout Indo-European history // *Nature* 449, pp. 717–720.

Pagel et al. 2008 — Q. ATKINSON, A. MEADE, C. VENDITTI, S. GREENHILL, M. PAGEL. Languages Evolve in Punctuational Bursts // *Science*, vol. 319, no. 5863, p. 588.

Peiros 1999 — I. PEIROS. Family Evolution, Language History and Genetic Classification // *Historical Linguistics & Lexicostatistics*. Ed. by Vitaly SHEVOROSHKIN & Paul J. SIDWELL. Melbourne, pp. 257–305.

Peiros 2000 — I. Peiros. Family Diversity and Time Depth // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 75–108.

Peiros, Starostin 1996 — I. Peiros, S. Starostin. *A Comparative Vocabulary of Five Sino-Tibetan Languages*. 6 vols. Melbourne.

Pokorny 1958 — Julius Pokorny. *Indogermanisches etymologisches Wörterbuch*. Francke Verlag, Bern & München.

Rédei 1988 — Károly Rédei. *Uralisches Etymologisches Wörterbuch*. Akadémiai Kiadó, Budapest.

Ringe 1992 — D. A. Ringe, Jr. *On Calculating the Factor of Chance in Language Comparison*. Philadelphia: The American Philosophical Society.

Robbeets 2005 — Martine Robbeets. *Is Japanese Related to Korean, Tungusic, Mongolic and Turkic?* Wiesbaden, Harrassowitz.

Sagart 2005 — Laurent Sagart. Sino-Tibetan-Austronesian: an updated and improved argument // Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas, eds. *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. London: Routledge Curzon, pp. 161–176.

Sands 2001 — Bonny Sands. Borrowing and diffusion as a source of lexical similarities in Khoesan // *Cornell working papers in linguistics*, 18, pp. 200–224.

Schuessler 2007 — Axel Schuessler. *ABC Etymological Dictionary of Old Chinese*. Honolulu, University of Hawai'i Press.

Starostin 1984 — Sergei Starostin. Gipoteza o genetičeskix sv'az'ax sino-tibetskix jazykov s jenisejskimi i severno-kavkazskimi jazykami [A Hypothesis about the Genetic Connections between Sino-Tibetan, Yeniseian, and North Caucasian Languages] // *Lingvističeskaja rekonstrukcija i drevnejšaja istorija Vostoka* [*Linguistic reconstruction and the Prehistory of the East*]. Moscow: Institute of Oriental Studies, pp. 19–38.

Starostin 1989 — Sergei Starostin. *Rekonstrukcija drevnekitajskoj fonologičeskoj sistemy* [*Reconstruction of the phonological system of Old Chinese*]. Moscow, Nauka Publishers.

Starostin 1991 — Sergei Starostin. *Altajskaja problema i proisxoždenije japonskogo jazyka* [*The Altaic Problem and the Origins of the Japanese Language*]. Moscow, Nauka Publishers.

Starostin 2000 — Sergei Starostin. Comparative-historical Linguistics and Lexicostatistics // *Time Depth in Historical Linguistics*. McDonald Institute for Archaeological Research, Oxford Publishing Press, pp. 223–259.

Starostin 2003 — Sergei Starostin. Statistical Evaluation of the Lexical Proximity between the Main Linguistic Families of the Old World // *Orientalia et Classica III: Studia Semitica*. Moscow, RSUH Publishers, pp. 464–484.

Starostin 2007a — Sergei Starostin. Opredelenije ustojčivosti bazisnoj leksiki [Defining the Stability of Basic Lexicon] // S. Starostin. *Trudy po jazykoznaniju* [*Works in Linguistics*]. Moscow, Jazyki slav'anskix kul'tur, pp. 825–839.

Starostin 2007b — Sergei Starostin. Indo-European among other language families: problems of dating, contacts and genetic relationships // S. Starostin. *Trudy po jazykoznaniju* [*Works in Linguistics*]. Moscow, Jazyki slav'anskix kul'tur, pp. 806–820.

G. Starostin 2002 — George Starostin. On the genetic affiliation of the Elamite language // *Mother Tongue*, vol. VII, pp. 147–170.

G. Starostin 2003 — George Starostin. A Lexicostatistical Approach Towards Reconstructing Proto-Khoisan // *Mother Tongue*, vol. VIII, pp. 81–126.

G. Starostin 2008 — George Starostin. *Making a Comparative Linguist out of your Computer: Problems and Achievements*. Presentation at the Santa Fe Institute, August 12, 2008. Available online at: http://starling.rinet.ru/Texts/computer.pdf.

Swadesh 1952 — Morris Swadesh. Lexicostatistic dating of prehistoric ethnic contacts // *Proceedings of the American Philosophical Society* 96, pp. 452–463.

Swadesh 1955 — Morris Swadesh. Towards greater accuracy in lexicostatistic dating // *International Journal of American Linguistics* 21, pp. 121–137.

Swadesh 1965 — Morris Swadesh. Lingvističeskie sv'azi Ameriki i Evrazii [Linguistic Ties Between America and Eurasia] // *Etimologija 1964*. M., Nauka Publishers, pp. 271–322.

Trier 1981 — Jost Trier. *Wege der Etymologie*. Berlin, Schmidt.

Turchin, Peiros, Gell-Mann 2010 — P. Turchin, I. Peiros, M. Gell-Mann. Analyzing genetic connections between languages by matching consonant classes // In this volume, pp. 117–126.

Vossen 1982 — Rainer Vossen. *The Eastern Nilotes: Linguistic and Historical Reconstructions*. Berlin, Dietrich Reimer Verlag.

Wichmann et al. 2009 — Søren WICHMANN, Eric W. HOLMAN, Dik BAKKER, and Cecil H. BROWN. *ASJP lexical similarity as a measure of language genetic relationship*. Available online at: http://email.eva.mpg.de/~wichmann/ASJPPerformance14.doc.

Winter 1973 — Werner WINTER. Areal linguistics: some general considerations // *Currrent trends in linguistics 11 — Diachronic, areal and typological linguistics.* Ed. by Thomas A. Sebeok. The Hague: Mouton, pp. 135–148.

Yeon-Ju & Sagart 2008 — Lee YEON-JU, Laurent SAGART. No limits to borrowing: the case of Bai and Chinese // *Diachronica* 25:3, pp. 357–385.

Статья посвящена методологическим аспектам создания глобальной лексикостатистической базы данных по всем языкам мира — одной из наиболее актуальных задач международного проекта «Эволюция языка» (Институт Санта Фе). Автор предлагает ряд существенных изменений стандартной лексикостатистической процедуры, как-то: замена традиционного стословного списка Сводеша на более компактный список из 50 «сверхустойчивых» лексических единиц; постулирование праязыковых реконструкций «низкого уровня» в качестве отправных узлов общего генеалогического древа; использование как обычного сравнительно-исторического метода, так и представлений о «фонетическом сходстве» для подсчета когнатов; и, самое главное, упор на максимальную точность семантической реконструкции и на жесткие ограничения синонимии.

Peter Turchin (University of Connecticut)
Ilia Peiros (Santa Fe Institute)
Murray Gell-Mann (Santa Fe Institute)

Analyzing genetic connections between languages by matching consonant classes[1]

The idea that the Turkic, Mongolian, Tungusic, Korean, and Japanese languages are genetically related (the "Altaic hypothesis") remains controversial within the linguistic community. In an effort to resolve such controversies, we propose a simple approach to analyzing genetic connections between languages. The Consonant Class Matching (CCM) method uses strict phonological identification and permits no changes in meanings. This allows us to estimate the probability that the observed similarities between a pair (or more) of languages occurred by chance alone. The CCM procedure yields reliable statistical inferences about historical connections between languages: it classifies languages correctly for well-known families (Indo-European and Semitic) and does not appear to yield false positives. The quantitative patterns of similarity that we document for languages within the Altaic family are similar to those in the non-controversial Indo-European family. Thus, if the Indo-European family is accepted as real, the same conclusion should also apply to the Altaic family.

Keywords: distant genetic relationship of languages; comparative linguistics; phonetic similarity; Altaic languages; quantitative methods in linguistics.

1. Introduction

Tracing "genetic" relationships between languages is sometimes a source of controversy in comparative linguistics. For example, within the linguistic community there is not universal acceptance of the Altaic family, i.e., the idea that the Turkic, Mongolian, Tungusic, Korean, and Japanese languages are genetically related (share a common ancestor) [Campbell & Poser 2008]. Even the recent publication of the *Etymological Dictionary of the Altaic Languages* [EDAL] did not put an end to this controversy [Georg 2004; Vovin 2005]. The critics claim that the observed similarities can be due either to chance resemblances or to "areal convergence"—borrowing resulting from cultural contacts (discussion in Ref. [Dybo, Starostin 2008]).

To demonstrate that languages belong to the same linguistic family it is best to trace them back to their common ancestor (= proto-language of this family), with known sound system, grammar, and partial lexicon. In most cases such proto-languages have to be reconstructed. According to the standard methods of comparative linguistics this can be done only if potentially related languages preserve a sufficient number of proto-language morphemes. Through analysis of such morphemes linguists establish a system of correspondences between the sound systems of daughter languages. For example, many German words beginning in /c/ (z in orthography) have the same meaning as English words beginning in /t/—*Zunge : tongue*, *Zahn : tooth*, etc., while initial German /t/ corresponds to English /d/, as in *trinken : drink* or *trocken : dry*. A set of such observations is used to reconstruct the phonological system of the

proto-language and the forms of its individual morphemes (phonological reconstruction). The meanings of the morphemes are reconstructed using much less rigorous methods. One problem here is that there can be a substantial semantic shift between two related words (cognates). An example is English *clean* and German *klein* 'small'—although these words are known to be cognates (the original meaning was 'neat, clean') they now have rather different meanings.

So far, proto-languages of only a limited number of language families have been properly reconstructed, thus demonstrating that the languages forming these families are related. As proto-languages of most proposed families are yet to be reconstructed, linguists still lack convincing evidence on possible relationships between languages. To compensate for the lack of information linguists use a variety of provisional methods ranging from inspection-based judgments to more formalized *lexicostatistics*. The assumption here is that if languages are related they should have lexical morphemes of common origin having identical meanings from the Swadesh 100-item list [Swadesh 1955]. Since no changes in meanings are accepted, semantic connections between the morphemes are straightforward. Still, phonological identification of relatedness is not based in this case on a system of correspondences[2] and therefore is not strict enough, with some similarities being possibly due to chance.

Here we propose a procedure based on lexicostatistics that does use strict phonological identification and permits no exceptions. This approach allows us to estimate the probability that the observed similarities between a pair (or more) of languages occurred by chance alone [Ringe 1992; Kessler 2001]. By design the proposed method is "conservative": we go to great lengths to minimize the possibility of false positives (concluding that languages are related when in fact they are not). Such an approach, which places a heavy burden of proof on anyone favoring a genetic relationship, is far from optimal, but we adopt it to avoid polemical controversies while applying our method to cases such as that of the Altaic family. The method is not a substitute for the more sophisticated approaches of comparative linguistics. Rather, it provides a procedure for testing hypotheses of genetic relationships without relying on matters of choice or judgment.

2. Results

2.1. Testing the Method on the Indo-European and Semitic Families

Before tackling the Altaic family, we test how well this Consonant Class Matching (CCM) method works on the well-studied Indo-European family. We distinguish between using modern languages for this purpose and using attested or reconstructed ancient languages. Applying the procedure to 21 modern Indo-European (IE) languages (additional tables are in *Supporting Information*) we find that it reliably identifies such branches as Indic, Slavic, Germanic, and Romance (SIs varying between 45 and 77%, all statistically significant at $P < 10^{-6}$). By contrast, similarity between languages belonging to different branches is much lower (between 1 and 21%). A particularly interesting comparison is between Germanic and Indic languages (Table 1). The SIs are very low, between 1 and 7%. Half of the comparisons are not significant at the 0.05 level, while all but one of the rest are weakly significant at $0.05 < P < 0.01$.

---

[2] Another application of lexicostatistics requires good knowledge of comparative phonology and etymologies and is used to generate linguistic families classifications, based on the amounts of etymologically identical words revealed by each pair of languages studied.

Both the Indic and the Germanic groups reveal themselves beyond any doubt, while the genetic relation between these two groups is not convincingly demonstrated by Table 1. We recall that the validity of the IE family was originally established not on the basis of modern languages but rather by comparing ancient ones, which are much closer to each other. The results of the CCM method (Table 2a) reflect the greater degree of similarity (all comparisons are significant at least at $P < 0.02$ level, and most at much higher significance levels). The SI between Old High German and Old Indian, in particular, is 14%. The probability of this overlap happening by chance is vanishingly small ($<10^{-6}$). When we apply the CCM approach to several ancient Semitic languages (Table 2b) we find that SIs for all comparisons are highly significant ($P << 10^{-6}$).

The improved resolution obtained with ancient languages is not surprising. The longer the period since the two languages diverged, the more opportunity there has been for roots in the 100-item list to "mutate" and become dissimilar (that is, cross into a different phonetic class) or to be replaced (as a result of a semantic shift). As time passes, the degree of similarity between any two genetically related languages should eventually decline to the point where in direct comparison it is indistinguishable from random noise. However, if we keep applying the procedure of reconstructing proto-languages we may be able to defeat that phenomenon.

The Indo-European and Semitic families are unusual in that they enjoy such a rich abundance of attested ancient languages. Does that mean that we cannot investigate genetic relationships when ancient written sources are lacking? As suggested just above, one possible approach to this problem is to use reconstructed proto-languages. When we apply the CCM method to the proto-languages of four IE branches, we obtain the same pattern as for attested ancient languages (Table 3a). For example, the SI between the Proto-Iranian and the Proto-Germanic languages is 13%. By contrast, in pairwise comparisons between five modern Germanic languages (German, English, Dutch, Icelandic, and Swedish) and two modern Iranian languages (Kurdish and Ossetian) it ranges between 5 and 10% (average = 7%).

Using reconstructed proto-languages can sometimes yield even better results than using attested old languages, as is shown in the Iranian–Germanic comparison. The SIs between Old High German and Avestan or Classical Persian are only 9–10%, whereas the overlap between Proto-Germanic and Proto-Iranian is 13% (and the statistical significance of the result increases by several orders of magnitude). This improvement is at least partially due to the greater age of Proto-Germanic and Proto-Iranian compared with Old High German and Classical Persian respectively.

It should be mentioned, however, that the main issue is not the age of the languages, but the degree to which they resemble their proto-languages. Ancient languages are usually more archaic in this sense, as they retain many features of their proto-languages, both in phonology and lexicon. At the same time some modern languages are also quite archaic, for example, Lithuanian. Therefore the role played by this language in Indo-European studies is similar to that of Ancient Greek, Latin and other ancient languages. In some cases a proto-language can be only a thousand years old, but because of its archaic character its relations with other (proto-)languages can be identified even by the CCM method.

2.2. Applying the methodology to the Altaic family

Next, we use the CCM approach to test the reality of the Altaic family. We have four independent reconstructions [EDAL; Mudrak 1984]: Proto-Turkic, Proto-Mongolian, Proto-Tungus, and Proto-Japanese (Korean dialects are too similar to one another to justify a reconstruction

of Proto-Korean). We also calculated the degree of similarity between these four languages and Proto-Eskimo, because O. Mudrak [Mudrak 1984, 2009] proposed that Eskimo languages are closely related to the Altaic family. Available Aleut data is not sufficient for the CCM analysis.

The SIs for the four Altaic proto-languages (Table 3b) range between 6 and 11% (average = 8.7%). This range of values is lower than that for the IE family. Nevertheless, the significance levels range between 0.01 and $<10^{-5}$, and this is strong evidence for historical connections among the four linguistic groups. Note that when we run the test on modern languages, the degree of similarity between them is greatly attenuated. For example, comparing five modern Turkic languages (Turkish, Tatar, Chuvash, Yakut, and Tuvinian) with two modern Japanese ones (Tokyo and Nasa) we detect a statistically significant relationship only in two out of ten cases (*P*-values are 0.03 and 0.01). The SI between the proto-languages, however, is significant at $P < 0.001$ level. This is the same pattern that we have already noted in the context of the IE family. Interestingly, we find support for the hypothesis of Mudrak that there is a relationship between Altaic and Eskimo (Table 3b; significant SIs in 3 out of 4 cases).

We can now reject the explanation that the observed similarities between Altaic languages are due merely to chance. What remains, however, is the second objection: that the proto-languages of these families could have acquired similar lexicons "due to a prolonged history of areal convergence" [Georg 2004]. One possible response to this alternative explanation is that borrowings into the basic lexicon (100-word lists) are rare [Starostin 2000]. Thus, we expect that languages belonging to different linguistic families will have low SIs, even when they have coexisted in the same region for a long period of time. We test this proposition empirically.

First, we looked at comparisons of languages belonging to different families that were located in spatial proximity: (a) Old Chinese vs. the proto-languages within Altaic; (b) Turkish vs. modern languages of people that inhabited the Ottoman Empire (1378–1914); and (c) Turkish vs. Classical Persian and Arabic (Table 4). The last comparison is particularly interesting because these three languages have coexisted in close cultural interaction at least since the Seljuk Sultanate (eleventh century), and many educated persons in the Middle East were trilingual.

The SIs in Table 4 are somewhat higher than expected under the null hypothesis: three out of eleven comparisons are significant at 0.05 level, and the maximum SI is 6%. What is important for our purposes, however, is that prolonged contact yields much lower SIs than those observed beween proto-languages within Altaic (such as the SIs of 11% observed in comparisons of Proto-Mongolic with Proto-Turkic or Proto-Tungus). This observation is contrary to the hypothesis that the observed similarities between Altaic languages are entirely due to borrowings.

More generally, in the 66 comparisons between Altaic and Semitic languages the SIs ranged between 0 and 5% and there were only two significant *P*-values (whereas we expect 3.3; more generally, the distribution of *P*-values is not significantly different from the uniform, $\chi_9^2 = 9.55$, $P = 0.39$). This pattern is precisely what should happen when languages are so distantly related that most "signal" has been lost and there were no cross-borrowings into the basic lexicon. In the 363 comparisons between Altaic and IE languages, however, there were 45 significant values (versus the expected 18). There is, thus, evidence for either some limited degree of cross-family borrowing or else deeper genetic connections between the Altaic and Indo-European families, as was proposed by Illich-Svitych in the context of his Nostratic superfamily [Illich-Svitych 1971–84], or both. The main point, however, is that the evidence for internal connections between the Altaic languages is orders of magnitude stronger. (To test the superfamily idea properly using CCM it will be necessary to compare the reconstructed proto-languages of

Indo-European, Altaic, and so forth.) The maximum observed SI in comparisons of modern languages or proto-languages within Altaic to those within IE was 8% (between Albanian and Nasa, no doubt caused by chance: the bootstrap-estimated probability of getting at least one SI=8% or better in the 363 comparisons is $P > 0.7$). By contrast, in the comparisons between the proto- languages within Altaic we observe SIs up to 11%. The bootstrap-estimated probability of getting *two* SIs of 11% in six comparisons (Table 3b) by chance is much less than 10^{-6}.

3. Discussion

In summary, the Consonant Class Matching approach classifies languages correctly for well-known families and does not appear to yield false positives. It gives reliable statistical inferences about historical connections between languages recorded a relatively short time (say 3,000–4,000 years) after their divergence. Greater time depth (say 6,000–7,000 years) can degrade the signal to the point where it is not detected by the method. We can circumvent this problem, however, using proto-languages, which act like attested ancient languages.

The quantitative patterns of similarity that we documented for languages within the Altaic family are somewhat similar to those in the noncontroversial Indo-European family. The evidence for the common origin of the Altaic languages, at least with respect to word-list comparisons, is thus nearly as strong as that for the Indo-European languages. If the Indo-European family is accepted as real, the same conclusion should also apply to the Altaic family.

However, we do not make a stronger claim that the Altaic languages we analyzed form a monophyletic group [Pagel 2009], because in order to do so, we would need to use the method to construct a phylogenetic tree for these languages. More generally, it should be strongly emphasized that the CCM method is not seen by the authors as a substitutee for the standard procedures of comparative linguistics. Properly reconstructed proto-languages remain the principal tool for demonstratung that the daughter languages are genetically related. In the absence of such reconstructions the CCM method can be used as a "short-cut" approach for finding non-random relationships among languages.

4. Methods

4.1. Linguistic data

The linguistic data (lexicostatistical lists of individual (proto-)languages) are taken from a collection of databases prepared by participants of the Evolution of Human Languages Project (Santa Fe Institute, USA) and the Tower of Babel Project (Moscow, Russia). We code each root (= main lexical morpheme) in the 100-word list for each language by replacing its first two consonants with generic consonant classes, following a suggestion put forward by Dolgopolsky [1986]. The table mapping consonants to the nine classes is given at the next page.

Bold capital letters refer to the class code. *r*-type consonants are grouped together with retroflexes in **T**. The ninth class (#) is formed by consonants which historically can be identical with ∅ (lack of a consonant): ʕ, *h, y*, etc. For this paper we treat all vowels as one single class. As a result, each root is represented as a sequence of consonantal classes: *gataw* > KT, *xuthip* > KT, *mars* > MT, and *arbil* > #T (# represents the missing initial consonant). Altogether there are 81 (9×9) possible forms of roots. We have performed this procedure for 53 Eurasian and North African languages.

	Front: labial, labiodental	**Central**: dental, alveoral, postal-veolar, palatal, retroflex	**Back**: velar, uvular
Nasal	**M***: m, m, ɱ*	**N:** *n, n, ɲ, ŋ*	*ŋ, N*
Plosive, implosive, ejective	**P:** *p, ph, b, p′, ɓ*	**T:** *t, th, d, ɖ, t′*	**K:** *k, g, k′, q,* G
Fricatives and approximants	*ɸ, β, f, v*	**S:** *s, z, ʂ, ʐ ʑ, ʃ, θ*	*x, ɣ,* R
Affricates		**C:** *c, ʒ, ʧ*	
Laterals		**L:** *l, ɬ, ʅ*	

The measure of similarity between two languages is the proportion of roots of the same meaning whose first two consonant classes match. For example, English *nose* and German *Nase* (both coded NS) are classified as similar while *dog* and *Hund* (TK versus #N) are classified as dissimilar. The German *Zunge*, coded CN, and English *tongue*, coded TN, are also classified as dissimilar, even though they are cognates. Our measure of similarity misses systematic sound correspondences that cut across our consonant classes. (In addition, it omits information contained in vowels and in any consonants other than the first two.)

4.2. Statistical Analyses

The next step after determining the proportion of matches between two 100-word lists is to estimate the statistical significance of this result. A naïve approach assumes that the probability of a match between the first consonants or the second ones is one in nine (the number of consonant classes) and the probability of both consonants matching is 9^{-2} or one in eighty-one. With this method we would expect, on average, a bit more than one match (100/81=1.2) in a list of 100 words. This approach is, however, flawed in several ways. First, some consonant classes are more common than others, and therefore the random chance of both consonants matching is, on average, greater than 1:81. Second, presence of a certain consonant in one position may affect the probability of finding another consonant in the other position. In other words, the assumption of independence may not be warranted. Finally, we must deal with such irregularities as missing or multiple words in some positions.

We use the bootstrap method (15) to estimate the statistical significance of the observed proportion of matches between word lists of two languages (the Similarity Index, SI). The procedure works as follows. We randomly select a root from List 1 and match it with a random root from List 2 (there are two alternative methods of random selection, see the next paragraph for the explanation). Repeating this step 100 times, we calculate the "bootstrap SI" (the proportion of matches between two random 100-word lists). Next, we replicate this procedure many times (e.g., 10,000 iterations) and use the 10,000 bootstrap SIs to approximate the probability distribution of the SI under the null hypothesis (that any matches are due to chance). Finally, we determine the proportion of bootstrapped SIs that is equal to or greater than the index calculated for the original lists. This gives us an estimate of the probability of observing this value (or a larger one) under the null hypothesis. The smaller this estimated probability, the greater our degree of belief that the proportion of observed matches could not arise by mere chance.

There are two ways to perform random selection: with or without replacement. In the first case (the classic bootstrap) after a word is chosen from the list, and matched with a word from

the other language's list, the word is put back. In other words, the same word can be chosen several times (and, therefore, some other words are never chosen). The alternative procedure (known as the permutation test) is to sample without replacement, so that each word is selected once. We repeated our analyses using both the bootstrap and the permutation test and obtained similar results. However, the permutation test was slightly more permissive (it gave a greater proportion of false positives), and therefore we report only the bootstrap results. We routinely used 10,000 bootstrap iterations to construct the probability distribution of the SI, but in cases where all bootstrapped SI were smaller then the observed one, we reran analysis with 1 million iterations. Thus, $P < 10^{-6}$ means that the observed SI was greater than *all* of 1 million bootstrapped SIs.

Our approach allows for missing words. Thus, the SI is the number of matches divided by the number of possible matches (subtracting observations with missing values). Missing values are handled during the bootstrap in exactly the same manner. That is, a bootstrapped SI may also have a number less than 100 in the denominator, if missing values happened to be chosen during the sampling process.

5. Appendices

Table 1. Similarity Indices (consonant class matches) within and between modern Indic and Germanic language groups. Below the diagonal: Similarity Indices (percentage of matches). Above the diagonal: the bootstrap-estimated probability of the observed SI or a larger one under the null hypothesis.

	Hindi	Beng.	Nep.	German	Engl.	Dutch	Ice.	Swed.
Hindi	–	$<10^{-6}$	$<10^{-6}$	0.02	0.6	0.2	0.3	0.04
Bengali	52	–	$<10^{-6}$	0.02	0.01	0.01	0.14	0.01
Nepali	56	49	–	0.005	0.3	0.08	0.7	0.01
German	6	6	7	–	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$
English	2	7	3	46	–	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$
Dutch	4	7	5	76	59	–	$<10^{-6}$	$<10^{-6}$
Icelandic	3	4	1	46	45	54	–	$<10^{-6}$
Swedish	6	7	7	64	57	76	72	–

Table 2. CCM results for ancient languages (SIs below the diagonal, P-values above the diagonal).

(a) Indo-European languages.

	OInd.	Avest.	CPers.	OH Germ.	Latin	OIrish	AGreek	Hitt.
Old Indian, 1000 BCE	–	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$	0.0003	0.02	$<10^{-4}$
Avestan, 600 BCE	42	–	$<10^{-6}$	0.001	$<10^{-4}$	0.01	0.01	0.01
Classical Persian, 1000 CE	23	40	–	0.0002	$<10^{-4}$	0.0002	0.004	0.001
O.H.German, 900 CE	14	10	9	–	$<10^{-6}$	$<10^{-6}$	0.0001	0.006
Latin, 300 BCE	19	15	13	17	–	$<10^{-4}$	$<10^{-4}$	$<10^{-4}$
Old Irish, 900 CE	9	8	9	14	13	–	0.001	0.009
Ancient Greek, 600 BCE	7	8	7	11	22	9	–	$<10^{-6}$
Hittite, 1500 BCE	16	8	9	8	16	8	20	–

(b) Semitic languages.

	Akkad.	Hebrew	Aramaic	Arabic	Ge'ez
Akkadian, 1600 BCE	–	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$
Hebrew, 700 BCE	42	–	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$
Aramaic, 300 CE	37	49	–	$<10^{-6}$	$<10^{-6}$
Arabic, 600 CE	24	33	34	–	$<10^{-6}$
Ge'ez, 400 CE	32	37	29	33	–

Table 3. CCM results for reconstructed protolanguages (SIs below the diagonal, P-values above the diagonal).

(a) Indo-European

	P-Iranian	P-Slavic	P-Baltic	P-Germanic
Proto-Iranian	–	$<10^{-6}$	$<10^{-6}$	$<10^{-6}$
Proto-Slavic	20	–	$<10^{-6}$	$<10^{-6}$
Proto-Baltic	13	35	–	$<10^{-6}$
Proto-Germanic	13	19	21	–

(b) Altaic and Eskimo

	P-Turkic	P-Mong.	P-Tungus	P-Japanese	P-Eskimo
Proto-Turkic	–	$<10^{-4}$	0.002	$<10^{-4}$	0.04
Proto-Mongolian	11	–	$<10^{-5}$	0.0008	0.004
Proto-Tungus	8	11	–	0.02	0.00003
Proto-Japanese	9	7	6	–	0.41
Proto-Eskimo	6	8	10	2	–

Table 4. CCM results for inter-family comparisons

(a) Altaic proto-languages vs. Old Chinese

	Proto-Turkic	Proto-Mong.	Proto-Tungus	Proto-Japanese
Similarity Index	2	2	5	6
P-value	0.28	0.21	0.02	0.03

(b) Languages of the Ottoman Empire vs. Turkish

	Kurdish	Serbian	Albanian	Greek	Armenian
Similarity Index	4	4	5	1	1
P-value	0.10	0.08	0.01	0.72	0.76

(c) Turkish vs. Persian and Arabic

	Persian	Arabic
Similarity Index	4	1
P-value	0.06	0.79

For technical reasons, tables containing raw data on consonant classes and their matching could not be included in the article itself. They can, however, be freely accessed on-line at http://cliodynamics.info/data/SuppInfo.xls

Literature

CAMPBELL L. & POSER W. J. (2008). *Language Classification: History and Method*. Cambridge University Press, Cambridge.

DOLGOPOLSKY A. (1986). A probabilistic hypothesis concerning the oldest relationships among the language families // *Typology relationship and time: a collection of papers on language change and relationship by Soviet linguist*, eds SHEVOROSHKIN V. & MARKEY T. Karoma, Ann Arbor, pp. 27–50.

DYBO A. V. & STAROSTIN G. S. (2008). In Defense of the Comparative Method, or the End of the Vovin Controversy // *Aspects of Comparative Linguistics* 3. Institute of Oriental Studies, Moscow, pp. 109–258.

EDAL — Sᴛᴀʀᴏsᴛɪɴ S., Dʏʙᴏ A., & Mᴜᴅʀᴀᴋ O. (2003). *Etymological Dictionary of the Altaic Languages*. Brill, Leiden.

Eғʀᴏɴ B. & Tɪʙsʜɪʀᴀɴɪ R. J. (1993). *An introduction to the bootstrap.* Chapman and Hall, New York.

Gᴇᴏʀɢ S. (2004). Review of Etymological Dictionary of the Altaic Languages // *Diachronica* 21: 445–450.

Iʟʟɪᴄʜ-Sᴠɪᴛʏᴄʜ V. M. (1971–84). *Opyt sravneniya nostraticheskikh yazykov (semito-hamitskij, kartvel'skij, indoevropeyskij, ural'skij, dravidijskij, altajskij). Sravnitel'nyj slovar'*. Vols 1–3. Nauka, Moscow.

Kᴇssʟᴇʀ B. (2001). *The Significance of Word Lists*. CSLI Publications, Stanford, CA.

Mᴜᴅʀᴀᴋ O. A. (1984). K voprosu o vneshnikh svyazyakh eskimosskikh yazykov [On the external relations of Eskimo languages] // *Linguistic reconstruction and the prehistory of the East*. Institute of Oriental Studies, Moscow.

Mᴜᴅʀᴀᴋ O. A. (2009). *Some Altaic Etymologies Found in the 100-word Etymostatistics List.* Manuscript.

Pᴀɢᴇʟ M. (2009). Human language as a culturally transmitted replicator // *Nature Reviews Genetics*. Published online 7 May 2009: 1–11.

Rɪɴɢᴇ D. A. (1992). On calculating the factor of chance in language comparison // *Transactions of the American Philosophical Society* 82: 1–110.

Sᴛᴀʀᴏsᴛɪɴ S. A. (2000). Comparative-historical Linguistics and Lexicostatistics // *Time Depth in Historical Linguistics*, eds Rᴇɴғʀᴇᴡ C., McMᴀʜᴏɴ A., & Tʀᴀsᴋ L. McDonald Institute for Archaeological Research, Cambridge, pp. 223–259.

Sᴡᴀᴅᴇsʜ M. (1955). Towards greater accuracy in lexicostatistic dating // *International Journal of American Lingusitics* 21: 121–137.

Vᴏᴠɪɴ A. (2005). The end of the Altaic controversy // *Central Asiatic Journal* 49: 71–132.

Гипотеза о генетическом родстве тюркских, монгольских, тунгусо-маньчжурских, корейского и японского языков до сих пор оспаривается многими лингвистами. Авторы статьи вносят свой вклад в разрешение этого вопроса, предлагая новую, относительно простую процедуру анализа возможных родственных связей между языками: метод отождествления форм по консонантным классам (Consonant Class Matching, сокр. CCM), работающий на материале стандартных лексикостатистических списков. Метод позволяет оценить вероятность случайных совпадений между двумя или более языками в этих списках. Калибрация процедуры проводилась на материале хорошо изученных семей (индоевропейской и семитской) и не привела к получению заведомо ошибочных результатов. При дальнейшем применении метода к языкам алтайской семьи оказывается, что степень схождений между ними в целом не сильно отличается от соответствующих результатов по индоевропейским языкам. Таким образом, серьезных оснований на то, чтобы отвергать алтайское родство (одновременно принимая индоевропейское), на самом деле не обнаруживается.

Pierre J. Bancel & Alain Matthey de l'Etang
Association for the Study of Language in Prehistory

Where do personal pronouns come from?[1]

The stunning preservation of 1st and 2nd person pronouns and possessives in low-level language families turns into a relative diversity within and between macrofamilies and phyla. However, the global stock of ancestral pronoun stems exhibit particularities hardly compatible with a completely independent origin. A tentative evolutionary explanation of these apparently contradictory facts is proposed here. In the evolution of language, pronouns may have appeared only with syntactic articulation, often linked to the acceleration of cultural evolution seen in *Homo sapiens* from around 100 kyBP on. Syntax itself must have evolved over a long timespan, and the emergence of pronouns from preexisting words — nominals that were the most frequent subjects and objects of verbs referring to the speaker and the hearer, though this reference indirectly depended from their original meaning — must have taken time as well. The multiple stems reconstructed for each person in macrofamilies (and, to a lesser degree, low-level families) might be a trace of a final stage of this evolution.

Keywords: Comparative linguistics, typology, personal pronouns, kinship terms, origins of language

The problem

In two centuries of comparative-historical linguistic research, it has become more and more evident that 1sg and 2sg pronouns and possessives are in nearly all language families like hard rocks standing in a plain, resisting erosion long after most other ancestral words have been swept away by the winds of time. Dolgopolsky (1964) finds 1sg and 2sg pronouns to be the first and third longest-lasting word meanings, respectively. Pagel (2000: 205) calculates the time necessary for words of ancestral languages to disappear from half their descendants — an idea adapted from particle physics —, and also finds the 1sg pronoun to be an extraordinarily enduring word, with a half-life of 166 ky.[2]

In an extensive study of *m-* and *t-* stems in the Eurasiatic[3] macrofamily (Bancel & al. forthcoming), we have calculated their loss rates in the Proto-Indo-European (PIE) 1sg and 2sg pronouns and possessives from nearly 500 IE languages and dialects. In the four paradigms, *m-* and *t-* have survived in 98.5% to 99.6% of IE languages. With an estimated age of 8,000 years for

---

[1] Mail should be sent to first author at pierrejbancel@hotmail.com.

[2] Pagel concedes that this figure "should not be taken literally, and most certainly do[es] not imply [a] time [depth] of 166,000 years or even 15,000 years for the Indo-European data." In fact, the method relies on an estimated age of the considered family, which is already embedded in the word's estimated loss rate from which half-life is calculated.

[3] We take the term "Eurasiatic" in Greenberg's (2000–2002) sense, rather than in that adopted by Gell-Mann & *al.* (2009), but it makes no real difference for our present purpose.

the IE family, these figures correspond to incredibly low loss rates per millennium of 0.05% to 0.24%.[4] These rates correspond to half-lives of *m- and *t- in the range of several hundreds of millennia.[5] And the situation is much the same in most other Eurasiatic subgroups.

With such inoxidizable pronouns and possessives, one would expect the situation to change very little as one proceeds back in time. By the preservation standards of PIE *m- and *t-, the pronouns and possessives of an ancestor language spoken 20 kyBP should be reflected in 96.1% to 99% of its daughter languages. Even the Proto-Sapiens hypothesis should receive quick confirmation from an expected near universality of pronouns and possessives. If Proto-Sapiens was spoken 100 ky ago, as one may reasonably estimate on archeological and genetic grounds, 1sg and 2sg pronouns and possessives should have been preserved in 82.7% to 95.1% of its descendant languages — i.e. all languages of the world — and in a still greater proportion of families, whose proto-languages by definition have had less time to evolve.

However, even at the incomparably younger Eurasiatic stage (often estimated in the 10 kyBP range), we are faced with much more diversity: Turkic, Korean, Japonic and Aleut entirely lost *t-, and in at least Korean *m- has vanished as well.[6] Enlarging our view to families more distantly related to Eurasiatic still worsens the picture. According to most Nostraticists, the families directly related to Eurasiatic are Kartvelian, Dravidian and Afroasiatic — unless it is rather Amerind, as is claimed by Greenberg (2002: 2–3). There are only scattered traces of 1sg *m- in Afroasiatic (Bomhard 2008: 274), which has however a 2sg *(n)t-. As to Amerind, we are faced with the uncomfortable situation where *m- is the stem of 2sg and 2pl pronouns (Greenberg 1987: 277–9, see also Nichols 2008) — though Ruhlen (1994a: 228–9) also posits an Amerind 1pl *ma. For its part, the Amerind 1sg stem *n- (Greenberg 1987: 272–5; see also Ruhlen 1994a: 192) is reconstructed in Nostratic as a 1pl (Bomhard 2008: 281–3), including in Indo-European (e.g. Latin nōs 'we, us,' Gothic uns 'us'), and with lesser reliability as a 2sg stem as well (Bomhard 2008: 287–9).

Finally, if one widens the scope unto the global level, as done by Ruhlen (1994a: 252–60), who compiled lists of 1sg, 1pl, 2sg and 2pl pronouns in the world's language families,[7] what one finds is an apparently desperate mess of *m- and *n- in the two persons and numbers, *k- 1sg and 2sg, *t- 2sg and 1pl, plus numerous erratic forms (Table 1). But is really the global diversity of pronouns a mess, and is it completely desperate? Not exactly.

First of all, phonetic diversity among pronoun stems is not as huge as it seems at first glance, with 40 stem phonemes in Ruhlen's list of 348 pronominal forms. Six consonants

---

[4] The loss rate per millennium r results from the formula $r = 1-(1-x)^{1/y}$, where x is the total loss rate over y millennia. Thanks to Sébastien Gaudry (Ecole Centrale Paris) and Sabine Bréchignac (Hôpital Avicenne, Assistance publique-Hôpitaux de Paris) for their contribution to this formula.

[5] With Pagel's formula (half-life $t_{50} = -\log_e(0.5)/r$, where r is the loss rate per millennium), a 0.05%/ky loss rate amounts to a 1,386 ky (= 1.4 My!) half-life; a 0.24%/ky loss rate "only" equals a 289 ky half-life. These results, though really indicative of a massive stability of pronoun stems, must be taken with a big grain of salt because of their sensitivity to the size of sample, an important difference with the original half-life method in physics, where all particles of the sample already exist in the beginning of the experimentation, while in language evolution they appear in the course of it with the successive divergences of the proto-language.

[6] We do not count Korean uli 'we,' whose u- is taken by Greenberg (2000) to be the final outcome of *mu > *bu > wu (wuli 'we' is attested dialectally) > u on the account of analogous *m > b evolutions in Uralic, Altaic and Chukotko-Kamchatkan, as a case in which the stem consonant *m- has survived. In our Eurasiatic tables, *m- and *t- are considered surviving only when the stem consonant left a clear phonetic trace of itself.

[7] The forms compiled by Ruhlen are either reconstructions (in families where the work was done) or best guesses about the most likely original forms (in each of the other families). Given the extraordinary stability of pronoun stems, there is little doubt that in the latter cases a phonologically informed inspection may allow to identify most original stems nearly as accurately as reconstruction.

alone — namely *m, n, t, k, s, j* — make up nearly two thirds of the sample (217 items, or 62.4%). All six are found on different continents in various distant families. Six more sounds are relatively common: these are *h, ʔ, ŋ, w, i, u* (61 items, or 17.5% of sample). All 28 other sounds occur very scarcely, with from 4 items to a single one each.

Table 1. **Number of occurrences of each stem phoneme in Ruhlen's (1994a) worldwide lists of pronouns.** In CV, CVC, VC and VCV forms, C_1 is considered the stem; in VV forms, V_1 is taken to be the stem. Alternate forms with different C_1 have been counted under each consonant, but alternate forms with the same C_1 have been counted only once. A few complex forms have been discarded from the count. Symbols *j* and *y* most of the time transcribe a palatal glide and have been subsumed under *j* in the table. For both *b-* and *p-* stem consonants, a subcount is given between parentheses of forms alternating with *m-* forms in the same family.

Stems	1sg	1pl	2sg	2pl	Total	% of total
m	11	19	19	17	66	19.0
n	20	19	12	5	56	16.1
t	5	7	10	8	30	8.6
k	17	6	11	5	39	11.2
s	4	1	7	3	15	4.3
j	7	3	0	1	11	3.2
Subtotal 1	**64**	**55**	**59**	**39**	**217**	**62.4**
w	4	1	9	0	14	4.0
ŋ	8	1	4	0	13	3.7
ʔ	5	1	3	2	11	3.2
h	7	1	2	0	10	2.9
i	6	0	1	0	7	2.0
u	4	0	2	0	6	1.7
Subtotal 2	**34**	**4**	**21**	**2**	**61**	**17.5**
p/b (*p/b* alternating with *m*)	0/0 (2/2)	0/1 (2/1)	2/0 (4/0)	1/2 (1/0)	3/3 (9/3)	1.7 (3.5)
p b v d z ð r tˡ sw ɬ l š ž č ǰ šʷ šʲʷ lž ch ñ g kh kw x xw ɢ ħ a	**18**	**22**	**16**	**14**	**70**	**20.1**
Average # of occurr. of 1/40 stems	2.9	2.0	2.4	1.4	8.7	2.5
Total	**116**	**81**	**96**	**55**	**348**	**100.0**

This is old news, in a way, for it has long been remarked that pronouns in most languages have a tendency to be based on a few stem consonants, which was attributed to a kind of functional convergence due to their huge frequency in discourse. Of course, the pronouns' overall shortness may be (and, in many languages, surely is) independently due to this functional constraint. Nevertheless, frequency cannot explain the massive convergence of pronoun *stems* on a handful of consonants at the global level, particularly with regard to the inalterable stability of stems in low-level families: if change had always been as slow as is observed in low-level families, there would be no phonetic convergence nor divergence of any kind to be expected. Preservation would be the only choice.

However, things are not that simple. A particular form of change may be observed already in low-level families, and this change almost exclusively consists in simplification: rather than innovating or borrowing pronoun stems, descendant languages may preserve only part of the stems reconstructed in their ancestral language. It may be observed for 1sg in the Indo-European family in our survey covering 500 languages, exactly a third (33.7%) of which lost any reflex of the PIE suppletive nominative *eghom 'I' (the whole Celtic group — save perhaps Gaulish, see Blažek 2008 —, plus parts of Romance, Tocharian, Iranian, Indic, and Anatolian). And almost no language having (often independently) lost *eghom did replace it by a new pronoun. Nearly all have generalized a form of the other PIE 1sg stem *m- instead.

At the Eurasiatic level, the 2sg PIE pronoun stem *t- is also generally attested in Uralo-Yukaghir, Mongolic, Tungusic[8], and Chukotko-Kamchatkan as a pronoun stem as well, so that there may be no doubt about its Eurasiatic ancestry. But there is another Eurasiatic 2sg pronoun stem *s-, found in Turkic, Tungusic, Korean, Japonic, Gilyak, and Kartvelian, also represented in the Eastern Itelmen 2pl *suze* 'you' (cp. 1pl *muze* 'we') and in the Eskimo 2pl subject marker of intransitive verbs -si (Greenberg 2000: 74–6). In PIE, it is also represented by a 2sg verb ending — and, since most personal verb endings derive from grammaticalized pronouns, there may be little doubt that the ancestor language of PIE had a 2sg pronoun stem *s-. Where has gone this Eurasiatic *s- pronoun stem in the Indo-European, Uralo-Yukaghir, Mongolic, Chukotko-Kamchatkan (save Eastern Itelmen) and Aleut lineages? It clearly underwent a severe loss rate, hardly compatible with those observed in low-level families.

This apparent multiplication of pronoun stems in ancestral languages as one goes farther back in time poses a strong typological problem, aptly spotted and exposed by Babaev (2008: 8): no known language possesses as many pronominal stems as are reconstructed for Proto-Nostratic. However, Babaev's explanation of this ancient variety as an artifact of reconstruction, resulting from innovations having piled up in descendant languages, remains puzzling, precisely because these too numerous Proto-Nostratic pronominal stems do not appear to have been innovated in each descendant language or family, most of them being found in several distant subgroups and being unlikely to have been borrowed. At the global level, with a half-dozen consonants gathering a large majority of low-level ancestral pronoun stems, one may only expect that the stock of pronoun stems in each of the most ancient macrofamilies will more or less be the same, though they will not match systematically with regard to person and number across macrofamilies.

Besides this distribution of pronouns stems over families and time, the global stock of pronoun stems also exhibits a phonetic particularity. As compared to dental-alveolar *t* and velar *k*, plain oral labials are amazingly underrepresented. To be sure, *p-* is not completely absent from Ruhlen's lists, nor is its voiced counterpart *b-* (18 items together, or 5.2% of sample), but exactly two thirds of them (12 out of the 18 items) appear to alternate with an *m-* form, e.g. in Ruhlen's Altaic 1sg forms *mi* ~ *bi*, where *bi* is the suppletive nominative of *mi* and certainly derives from it. This leaves us with only 6 occurrences (1.7% of sample) of undoubtedly original *b-* and *p-* stems (3 each, or 0.9%), to compare to the 66 occurrences of their nasal counterpart *m-*, and the 30 and 39 occurrences, respectively, of their dental *t-* and velar *k-* counterparts. Since plain oral labial stops are among the most widespread consonants in the languages of the world, their discretion among the global stock of pronouns would be a big stroke of luck if pronouns had arisen in complete independence from one another.[9]

---

[8] In Tungusic, 2sg *t-* is represented in the 1pl inclusive *miti,* literally 'I-thou' (Greenberg 2000: 72).

[9] Another distributional particularity in Ruhlen's list is the low number of voiced stops. With 3 *b-,* 4 *d-,* 2 *g-* and 1 *G-,* against 30 *t-* and 39 *k-,* they are nearly 10 times scarcer than their unvoiced counterparts. It may be (and

How to reconcile the extraordinary stability of personal pronouns in low-level families with their relative divergence within deeper-level families, while they however concentrate on very few stem consonants at the global level (though they do not match semantically), and display a typologically striking lack in their phonetic distribution? We will propose below a conjectural solution, deriving them from kinship appellatives like *mama, nana, tata, kaka, jaja*, etc., which must have preexisted them.

A solution

Before exposing our arguments, a warning is here in order. We are not reconstructing, with whatever method, be it "standard" or multilateral comparison, the ancestry of such or such pronoun stem, e.g. Eurasiatic *m-* and *t-*, as interpreted by Babaev (2009a: 142) in his review of Bengtson (2008), where our conjecture was first exposed.[10] We did not (nor do today) intend to claim that any particular pronoun stem descends from such or such kinship appellative. In particular, we do not claim that speakers of Proto-Eurasiatic (nor of any other known proto-language) had changed some of their kinship terms into personal pronouns. Rather, we wanted (and still want) to suggest that 1st and 2nd person pronouns *as a category* might — and, in our opinion, may only — have evolved from that of kinship appellatives, in the course of a radical transformation of the nature of language, namely the emergence of syntactic articulation, by far anterior to Eurasiatic and Nostratic (though some of its evolutionary consequences might have lasted up until their respective time periods).

Of course, this conjecture being correct would imply that most pronoun stems in the world's languages, and among them Eurasiatic *m* and *t,* would in all likelihood remotely descend from kinship appellatives. But the demand presented by Babaev (2009a: 142) of typological evidence for such a shift is impossible to satisfy, precisely since pronouns change so little in modern languages — and the situation is absolutely not the same as it was at the time where human language acquired pronouns, both linguistically (1st and 2nd person pronouns now exist in all languages) and sociologically (kinship must have then been the only mode of social organization). As to the comparative evidence required by Babaev, it is also impossible to satisfy, for the same reason — except collectively, with the fact that a great majority of pronoun stem consonants, known not to be innovations (at least within our comparative reach), also are the stem consonants of kinship appellatives, which in turn must have preexisted pronouns (a claim independent from the belief that modern appellatives descend from Proto-Sapiens, as we will see).

We use the results of linguistic comparison to try and gain a view of very ancient facts, which linguistic comparison alone could not attain. Our results may certainly seem less secure than those obtained through regular sound correspondences, but asking questions like "Of the phonetic and syntactic articulations, which one may have appeared first?" or "What does it take for a language to have personal pronouns?" also is historical linguistics, even if sound correspondences alone may never answer them. The reader is thus urged not to apply auto-

---

surely in some cases is) an artifact of comparison: since initial voiced consonants do not very often get devoiced, one is tempted, when faced with *p-* ~ *b-*, *t-* ~ *d-* or *k-* ~ *g-* correspondences, to posit preferentially an unvoiced original consonant. But, precisely since initial voiced consonants do not often get devoiced, if numerous families have had originally voiced pronoun stems, one should retrieve them in their descendant languages and not be tempted to posit an unvoiced original stem consonant.

[10] Babaev's mistake may in great part be due to the structure of our paper, most of which dealt with Eurasiatic pronouns, then shifted abruptly subject to this conjecture, and to our admittedly unusual method, as well as, and perhaps mainly, to lacunas in our argumentation, which we will try to mend here.

matically his/her knowledge of comparative linguistic procedures (though this and other knowledge may certainly be useful) in assessing our evolutionary arguments. Here they are.

As already mentioned, the six stem consonants (*m, n, t, k, s, j*) grouped in the first part of Table 1, totalizing 62.4% of ancestral pronominal forms worldwide, also are stems of globally-spread kinship appellatives, namely the five Proto-Sapiens words *mama, nana, tata, kaka* and *jaja* (Bengtson & *al.* 1994: 292–3; Ruhlen 1994*b*: 122–4; Bancel & *al.* 2002, 2005, in press; Matthey de l'Etang & *al.* 2002, 2005, 2008, in press), plus *ise* 'father,' widespread in Eurasiatic, Amerind and Niger-Congo. Most other stems listed in Table 1 may derive phonetically from one or another of these six consonants.[11] From a general phonetic viewpoint, this makes kinship appellatives unproblematic ancestors of personal pronouns. But why should they be the pronouns' ancestors? Why could not pronouns always have coexisted with them?

To answer these questions, we must leave the domain of strict linguistic comparison and enter those of general theory of language and human evolution. Human languages are known to be doubly articulated, phonetically and (morpho)syntactically (Martinet 1960: 13–5, 17–8). The phonetic articulation consists in meaningless elements, phonemes, combined into sequences to form simple meaningful elements, called *monemes* by Martinet, a term of his own coinage referring to both simple words and morphemes. In turn, the syntactic articulation consists in the combination of these elementary meaningful monemes into complex sentences.

Martinet orders these two articulations into a first and a second one, and finds that syntax comes first. His reasoning is based on a representation of language, viewed only from the speaker's side, in which the speaker has something to make known to someone else ("*tout fait d'expérience à transmettre, tout besoin qu'on désire faire connaître à autrui", ibid.*: 13). The speaker begins analyzing his initial, languageless (?) thought as a bunch of lexical units corresponding (?) to this thought of his,[12] which he arranges in the right order (syntactic articulation) and finally proceeds to convert this word sequence into a phoneme sequence (phonetic articulation). Thus, Martinet's order of syntactic and phonetic articulations exclusively relies on the assumption that a "thought" is entirely converted into an ordered word sequence in the speaker's mind before being passed to the phonetic component, in order to be converted into a phoneme sequence and uttered. With such a sequential processor, speakers should not be able to utter two sentences in a row without at least a marked pause between the two, since they would be able to begin to process the second one only after having finished to utter the first. Also, one never should see a speaker stopping short in the middle of a sentence, searching for a word not yet found in his internal lexicon. But many speakers are perfectly able to utter an indefinite number of sentences with no other pauses than for a short breathe, while everyone utters incomplete sentences everyday.

Instead of processing full thoughts/sentences through all components of their language processor one after another, real speakers must handle many different subparts in extremely short timespans, and we have as much as no understanding of this real-time language processing — albeit it is the only grammar deserving to be called *natural*. Within the timespan of a single sentence, speakers continuously think, spot words and morphemes corresponding to the theme and articulations of their thought (which words may in turn modify their thought, against which they must be checked back), organize them into groups and phrases (again with implications on and neces-

---

[11] Only the basic plain velar nasal *ŋ*, represented in the second part of Table 1, does not appear as a very likely descendant of any of consonants *m, n, t, k, s, j*. We leave the question pending, noting that (i) cases of evolution *m > ŋ*, though not common, are not exceptional, (ii) in our global database of kinship terms, there are relatively numerous instances of an appellative *(ŋ)aŋa* 'mother, grandmother, aunt,' mostly in African, Indo-Pacific and Australian languages, even though they do not make a very strong case for a regional etymology, while Proto-Niger-Kordofanian 1sg independent pronoun **ŋgai* exactly matches Proto-Pama-Nyungan 1sg ind. pr. **ŋgai* (Ehret 2007).

[12] This process, if it existed under the form assumed by Martinet, would be a third articulation of language.

sary checking with their initial thought), process bits of morpheme sequences in the morpho-nological component, then in the phonological, then send them to the motor component to utter the corresponding sounds, and control *a posteriori* what they have just said with regard to pho-netic, syntactic, lexical and logical accuracy, while keeping a pragmatic eye on the interlocutor and his/her reactions. The existence of all these subprocesses is *a contrario* warranted by the most com-mon lexical, syntactic, morphological, phonological and phonetic speech errors (for an example of real-time morphological speech error in children, see Pinker 1999: 220–3).

As for hearers (because hearers are a necessary ingredient of language, and they cannot decently be supposed to begin decoding with syntax before having heard and identified pho-neme sequences, and found corresponding words in their inner lexicon), they continuously decode the acoustic signal hitting their eardrums, while processing what they have just heard on both lexical and morphosyntactic levels, controlling the grammaticality of their interpreta-tion as well as its semantic, logical and pragmatic relevance on both levels of discourse and external circumstances, and preselecting the most likely continuations at the phonetic, lexical (e.g. an animate noun after adjectives such as *sympathetic* or *loath,* etc.), syntactic (e.g., in an SVO language, verbs after a subject nominal, direct objects after a transitive verb) and semantic levels to speed up interpretation of the oncoming speech flow, keeping track in a permanently readjusted short-term memory of the few preceding sounds in order to rectify a possible auditory or parsing error, while they keep an eye on possible cues warning them that their speech turn is coming soon and they have to prepare to answer, or to emit some approbative grunt urging their interlocutor to speak on.

How many times these subprocesses are run during a sentence, whether they are run in parallel or not, and if so how they are synchronized, all these questions exceed our under-standing today, except that one may be sure that there is a lot of comings and goings between the different components of language within the time of a sentence in the minds of speakers and hearers. As a result, from the vantage point of speech act, not only syntax certainly is not the first articulation of language but ordering the two articulations is wholly devoid of reality.

Nevertheless, it seems that another ordering of the syntactic and phonetic articulations is possible from the phylogenetic viewpoint. Many arguments converge in support of the idea that syntactic articulation must have emerged late in the evolution of language.[13] The first line of support comes from studies on language acquisition by children, who at the age of 11–12 months start uttering isolated words, then begin (at 15–18 months) to use two- or three-word combinations, and finally begin (around 20–24 months) to acquire morphological and syntactic rules (Brigaudiot *& al.* 2002): children clearly acquire the phonetic articulation first.

It is also confirmed by observations from apes trained to manipulate symbols, either chimps (e.g. Gardner *& al.* 1989), gorillas (Patterson 1987), or bonobos (Savage-Rumbaugh *& al.* 1994). They are able to learn and to relevantly use up to several hundred symbols, but most of their utterances consist in a single symbol, even if the most gifted pupils may occasionally combine two or three of them, exceptionally four, though mostly without determined order.[14] For chimps using symbols, syntax remains beyond their capacities.

---

[13] Bickerton's (1990) theory of protolanguage, a misleading name for a primitive stage in the evolution of human language ability without syntactic articulation (and not the ancestral language of any given family), al-ready claims that syntax should have appeared in a relatively recent stage.

[14] For both apes and babies, 1-word utterances are sentences (specialists in language acquisition coined the phrase *holophrastic word* "whole-sentence word" to qualify them), and may convey complex meanings, often with heavy contextual reference, but it is not the point here. The point is that these sentences are not syntactically ar-ticulated — if they possibly are semantically, a component neglected by Martinet as if it were not part of language but contained in an extralinguistic "thought," still another dubious axiom.

Finally, the posteriority of the syntactic articulation is supported by mere commonsense: before gathering words into complex sentences, one must have words at one's disposal, which in all languages are made from phonemes. For this reason, any modern speaker must begin by the phonetic articulation in order to build words, and syntactic articulation has to come next.

How could archaic humans have built a syntactically articulated language before having invented the phonetic articulation and progressively built not only two or three articulate signs, but dozens or, more likely, several hundreds of strongly individualized words — otherwise, combining them would have been of little interest? And this initial process may not have been completed overnight. It is unlikely that the first phonetically articulate sequences also bore a truly symbolic meaning, as do modern words and morphemes — otherwise, it would have been like discovering at the same time the law of universal gravitation and the quarks, or the existence of microbes and the DNA. Rather, we would expect them to have fulfilled functions identical or close to preexisting animal vocalizations. Giving them a symbolic value must have been the result of a long subsequent evolution, as more phonemes became utterable with the progressive transformation of the human vocal tract, allowing to enlarge the lexicon enough to specialize some signs to designate clearcut classes of beings, things or actions — i.e. evolving them into words. Both these phonetic and semantic evolutions also must have long been dependent on the growth of brain size and processing power, as well as on such apparently hardwired behavioral evolutions as the emergence of spontaneous attention to articulate speech, the development of babbling in babies — a universal training stage, which may have appeared and spread only after mastering some degree of phonetic complexity had become a selective advantage —, or the tendency to react to speech with speech rather than directly with other acts. As a result, this initial evolution of phonetic articulation must have been anchored for most of its duration to biological evolution, whose pace is much slower than linguistic or cultural evolution.

Thus, there is an order in the two articulations of language, after all, which is historical in nature — and this order is the opposite of that found by Martinet. Phonetic articulation must have come first, and syntax only much later.

In human history, acquiring the second, syntactic articulation may not have been a small event. With syntax, you become able to tell stories, to describe precisely how to design and build any artifact, and to form complex thoughts about new ones. It is a fantastic universal tool for both innovation and transmission — technical as well as social, intellectual and religious. It must have revolutionized the life of the communities where it developed.

It happens to be the case that such a revolution has long been perceived in human prehistory. André Leroi-Gourhan (1964) studied the evolution of technical ability in humans, which he measured in meters of blade obtained per kilogram of rough silex knapped. He found that, since the earliest stone tools, ca. 2 MyBP, it had grown in direct correlation with the growth of endocranial volume, and hence brain size, until around 50 kyBP, at which point skull capacity stopped to grow while technology took off in a way silex blade length could not measure anymore. This 50 kyBP crossroads where cultural evolution finally diverged from the biological was termed the "Sapiens explosion," since new techniques of all kinds seemed to have suddenly appeared, including seafaring with the first settlement of New Guinea and Australia across at least 100 kilometers of sea (Coupé & al. 2005). For around the same time, our Sapiens ancestors had left their African homeland to colonize the whole Old World, where they quickly supplanted the various human species having evolved there separately since hundreds of millennia, like the European Neandertals.

This cultural explosion must today be relativized with regard to its alleged instantaneity, since it now appears to have been preceded by an evolution in the African homeland of *Homo*

sapiens, as shown by the discoveries at South African sites Klasies River Mouth (Singer *& al.* 1982), Blombos Cave (Henshilwood *& al.* 2001, d'Errico *& al.* 2005, 2009) or Pinnacle Point (Marean *& al.* 2007). There one finds, as early as 80–130 kyBP (and even 160 kyBP at Pinnacle Point), clear traces of culturally modern behavior: the early *Homo sapiens* who occupied these sites cooked meat and plants on fire, fed on marine resources, made microlithic and polished bone tools, and, at Blombos in layers dated to around 80 kyBP, carved symmetrical geometric patterns on regular parallelepipeds of red ochre, and pierced shell beads (found in clusters which must have been worn in necklaces).[15] All complex behaviors which archeologists rightly link with the necessary use of a form of symbolic language close in complexity to those used by contemporary humans.

Thus, as the consensus[16] grows, the Sapiens cultural "explosion" or, rather, acceleration, would be the archeological landmark left by the apparition and evolution of syntactic articulation in human language. A process which certainly took time itself, because of the quickly growing complexity of the real-time encoding and decoding processes evoked above. And if we may consider that it was already underway around 150 kyBP, had continued to develop around 80–100 kyBP and had still made more progress at 50 kyBP, we have no idea of when it was completed (nor even, to be provocative, whether it is completed today).

Well and good, but what has this discussion about syntactic articulation to do with the origin of pronouns? Simple. The existence of pronouns and person markers directly depends on syntax. Without syntax, they are not only useless but even inconceivable.

Imagine a language without syntactic articulation — with 1-word utterances only for a very long time, and then with 2 or 3 juxtaposed words. There are no subjects, no verbs. There may be calls, and names are useful for this use as they allow to call a particular person. Other symbolic words are used as whole sentences, with the help of context and gestures. What use would be *I* and *thou*? And, above all, how could have appeared these extremely weird words — whose essential semantic feature is to change reference with the speaker? It is the very essence of symbolic language to share symbols which refer to the same objects for all users, and in all languages all words — save person markers — share this precious property, whose acquisition gives babies the key of spoken language. Only 1st and 2nd person pronouns and markers have the exotic particularity that their only meaning is to change reference with the speaker.[17] I am my own and nobody else's I. And so is each of you all — his/her own and nobody else's I. Conversely, each of you is one of my thous, which he/she is not with regard to him/herself, while I am one of your thous, which I am not for myself.[18]

1st and 2nd person pronouns and markers are highly useful tools in conversation, and no attested human language seems to lack them. However, even with syntactically fully articulate languages, they are not absolutely necessary. It is always possible to speak in the 3rd person, Benveniste's (1946) non-person, occasionally using personal names to disambiguate who is doing what to whom: *Pierre and Alain tell Readers.* In the beginning of syntactic articulation, when people pro-

---

[15] Also, the time where the human brain reached its present size must be somewhat relativized, since early *Homo sapiens* like those of Shkul and Qafezh (ca. 90 kyBP) and even earlier *Homo neanderthalensis* (from ca. 300 kyBP on) already attained skull capacities within the range of contemporary humans.

[16] Notably expressed in several papers of Botha *& al.* (2009; e.g. d'Errico *& al.*), and in Bickerton (2009).

[17] Other words may include reference to the speaker or the hearer, like *here* 'around the place where I am,' *now* 'at the moment I am speaking,' or *this* 'the known or shown thing near me,' but only 1st and 2nd person pronouns exclusively consist in a reference to the speaker or the hearer.

[18] To be completely true, it may occur in the inner speech that one addresses oneself as a 2nd person — *"Pierre, what did you say?"* This mild symptom of a split personality reflects the fact that self-consciousness amounts to place oneself at a remove from oneself. However, talking about oneself as a 2nd person to somebody else would be considered a symptom of a serious speech or psychic disorder.

gressively became more and more able to combine words and to answer other people's utterances (something which must have been difficult and rare with 1-word utterances), 3rd non-person was certainly the only way to have a subject and a verb, as well as a verb and an object.

How may have appeared 1st and 2nd person pronouns, then? It would be absurd to suppose that they were intentionally invented by people having realized how useful they would be if they existed. Rather, they must have evolved from preexisting words. And the category these words must have belonged to is easy to identify. It is that of nominals which were used to refer to the speaker and the hearer — and hence to human beings —, whose most frequent members may have been turned into pronouns under a shortened phonetic form, as the development of syntactic articulation and the parallel rise of conversation made more and more often necessary to specify who was doing what to whom.

Among these nominals referring to humans, several subcategories do not qualify as the potential ancestors of personal pronouns. It would be very difficult to conceive how ordinary common nouns (like *hunter* or *girl*) or proper nouns (like *Jehan* or *Little Big Woman*) could have given rise to pronouns and acquired the property to switch reference: most common and proper nouns refer to the same object whoever is speaking, and are thus separated from pronouns by an apparently impassable semantic wall.

Moreover, if such ordinary common or proper nouns were the ancestors of pronouns, the global phonetic picture of present-day pronouns would be very difficult to explain in all cases. On the one hand, if all modern pronouns shared a common origin, and descended from a subset of common or proper nouns in a single ancestor language, how could one explain that it is impossible to assign any of the modern pronominal stem consonants to a common global origin? It would be at odds with the exceptional preservation of pronouns in low-level families. On the other hand, if present-day pronouns descended from a subset of proper or common nouns in several different ancestor languages, how could one explain that their stems converge so massively towards a handful of stem consonants, whatever the language family they belong to, while very few seem to have been innovated in the last 10 to 15 ky?

Among nominals likely to refer to the speaker and the hearer, only kinship terms, and in particular kinship appellatives like *mama, nana, tata, kaka, jaja,* etc., appear as likely ancestors of personal pronouns. First of all, kinship appellatives definitely are of Proto-Sapiens ancestry — because of their ubiquity and the impossibility, contrary to the widespread belief following Murdock's (1957, 1959) and Jakobson's (1960) famous papers on 'Why Mama and Papa?,' that they had resulted from convergent innovations (Ruhlen 1994*b*: 122–4; Bancel & *al.* 2002, 2005, in press; Matthey de l'Etang & *al.* 2002, 2005, 2008, in press).

Kinship appellatives must even be much more ancient than Proto-Sapiens, and certainly played a major role in the emergence of phonetic articulation in Proto-Human. The first phonetically articulate words, uttered by mouths and tongues that had not been designed for speech by evolution, must have been built from the simplest consonants cast into the simplest syllable structures (Lieberman & *al.* 1972, Lieberman 1992) — which kinship appellatives still are today, with their typical CVCV, VCV or CVC reduplicative structure and their basic plain stops and vowels. Rather than meaning anything in the modern sense, they must have fulfilled some of the functions of prelanguage vocal communication, like calls — which kinship appellatives still are today, and even exclusively in the first uses of 1-year children (Grégoire 1937, approvingly quoted by Jakobson 1960), to only progressively acquire a referential value, thus opening children the door to symbolic representation and meaning.[19] The first phonetically

---

[19] This succession in the acquisition of language by children is another indication that phonetically articulate sequences are likely to have emerged before symbolic representation.

articulate words also must have been easy to transmit from generation to generation through mouths, brains and ears lacking specialization for language, so that this invention did not get lost — and kinship appellatives, thanks to their particularly simple phonetic structure and functional usefulness, have not get lost until today. All these conditions are fullfilled by nursery kinship terms, and by them only. Finally, as said in the warning beginning this section, even those who think that modern kinship appellatives have not been inherited from Proto-Sapiens, but are innovated by children every now and then, could hardly argue against their anciety as a category. Since their acquisition by babies is — thanks to their unique phonetic and functional properties — a crucial initial step in the transmission of articulate speech and symbolic representation in all human communities of the world, arguing that kinship appellatives appeared recently would require to explain how babies (and more generally humans) managed to acquire articulate language before.

In the Paleolithic, all humans were hunters-gatherers, a lifestyle implying to live in small bands of a few dozen individuals, most of which are related. All historically known groups of hunters-gatherers have lived this way, and such was certainly the case of all groups since the very origins of the human lineage, as testified by the parallel lineages of bonobos and chimpanzees, who also live in small foraging bands of related individuals — and these bands display primitive features of a kinship-based social organization (De Waal 1982). More generally, evolutionary biologists classically explain how cooperation may have evolved among closely related individuals,[20] which is the case of all cooperating animals, whether insects or vertebrates (Hamilton 1963). John Maynard Smith (1964) even coined the now classical cover term of *kin selection* to refer to this branch of evolutionary theory. It is thus a safe bet to assume that, in archaic humans, language and kinship-based social organization, two highly cooperation-oriented institutions, must have evolved together from start.[21]

For these reasons, kinship appellatives must have been around long before the appearance of pronouns and person markers. They must have been in daily use as calls and address terms between Paleolithic hunters-gatherers, as they still are in contemporary societies by children towards parents, and in more traditional societies towards any person, which may be addressed according to age and status as 'son/daughter,' 'brother/sister,' 'cousin,' 'father/mother,' 'uncle/aunt,' or 'grandfather/grandmother.' It is extremely likely that kinship terms have become, in the early times of syntactic articulation, the choice tools to disambiguate the human subjects and objects in sentences, since all humans known to any speaker and likely to be told to and/or about belonged to his kindred.

---

[20] It essentially relies on the fact that related individuals share a great part of their genes, so that a mutation resulting in greater cooperation, even detrimental to an individual's reproduction, may be selected if it enhances reproduction of its relatives, which are likely to share this mutation and hence to propagate it. Bickerton (2009: 113–5) makes the point that high predation pressure on australopithecines in the savanna must have led to the reduction of "within-group competition (and, ultimately, the birth of cooperation)."

[21] In this respect, evolutionary theorist Richard Dawkins, in his world-famous book *The Selfish Gene* (1976), remarked that a child's mother's brother is the closest male ascendant with whom the child may be sure to share a maximum of genes, and as such is a choice subject for kin selective processes. Dawkins asked anthropologists whether the mother's brother would not have played a role in some human societies. In a footnote to the 2nd edition of his book, he mentions to have received volumes of mail from readers telling him that the mother's brother was a central subject for social anthropologists since more than a century, because of its prominent role in a great many societies worldwide. The globally-spread kinship appellative *kaka* 'mother's brother, grandfather, elder brother' (Ruhlen 1994b : 122–4; Bancel & al. 2002, in press; Matthey de l'Etang & al. 2002, in press) might be the earliest trace of a kin selective process having led to the rise of the mother's brothers' role in the development of human societies.

As for the semantic plausibility of the evolution of kinship appellatives into 1st and 2nd person pronouns, and especially with regard to the switching reference of pronouns, it may be remarked that kinship appellatives are the only other class of nominals to partly share this property. Indeed, they share referential properties with all the major classes of nominals, thus all the more qualifying as the ancestors of the entire category of nominals, beyond their internal features pointing towards their primeval anciety. Such is the case, for instance, of English *dad*. If I ask *'Where is your dad?,'* dad is a common noun, but if my interlocutor answers *'At the moment, Dad is out for angling,' Dad* is a proper noun referring to a single person — a specificity rendered in writing by the initial capital. But this proper noun, precisely due to the relational nature of kinship terms, is again specific. I am supposed to understand that *Dad* is in fact *my interlocutor's* father, and if I reply *'Oh! That's why Dad went out so early, they must have gone together,'* he in turn understands that I am referring to *my own* father. When used as proper nouns, i.e. referring to a determined person, kinship appellatives share with pronouns and person markers the particularity to switch reference with the speaker (though in the case of *Dad* the reference is not to the speaker or the hearer himself, but to a person considered as "inalienable property" of the speaker). Moreover, some kinship terms are reciprocal, i.e. they are likely to be used towards each other by two interacting speakers, like in English *brother* and *sister.* Any male whom I may call *Brother* may call me *Brother* in return if I am a male, and if I am a female any person I may call *Sister* may also call me *Sister.* This switching reciprocal reference of *Brother* and *Sister* is still closer to that of personal pronouns (though it fails to differentiate the two interlocutors in each one's speech). Thus, kinship appellatives intrinsically share referential properties with all three nominal categories of proper nouns, common nouns, and pronouns. Like common nouns, they can refer to a class of beings, defined by common properties of these beings (in the example, the category of *dads*). Like proper nouns, they can refer to a particular individual (the speaker's *Dad*). And, in this proper noun use (but contrary to all other proper nouns), they switch reference, like pronouns, from a particular individual to another as the speech turn passes.

In the stage of Proto-Human language that preceded the apparition of pronouns, kinship terms such as *mama, tata, nana* or *kaka* may have been the most frequent way to address people, so that they might easily have given rise to a 2nd person pronoun. It may seem less straightforward for the 1st person pronoun, since by definition there is no kinship term referring to oneself. However, just like for 2nd person, the 1st person pronoun must have emerged from an earlier nominal used by the speaker to refer to himself, and no other nominal category possesses such a word. It is perfectly conceivable that, in the stage before the emergence of personal pronouns, speakers referred to themselves by the kinship term used towards them by the addressee. In modern languages with personal pronouns, such practice would seem weird, but is occasionally used when speaking to children who do not master the use of personal pronouns, as in *'Mum wants Sonny to eat up those peas.'*

From such uses, which may have been general in the first stages of emergence of syntactic articulation, may have arisen an intermediate class of "pronominoids," made of shortened forms of the most frequent kinship appellatives, able to refer to either the speaker or the hearer (and hence used as both 1st and 2nd person according to circumstances). Their exact status we must admit to ignore, even though it seems likely that the choice among the series was initially determined according to the kinship relation between interlocutors.

In a subsequent phase, each of them would have specialized as a 1st or 2nd person, while they lost any semantic connection with kinship appellatives. If we assume that the most ancient language phyla split up during this period (which may have lasted up to several dozens of millennia), it would explain why all of them have pronoun stems chosen from a very small

consonant set, which appears to coincide with that of the most frequent kinship appellatives. It would also explain why, in spite of this striking convergence, pronoun stems do not match semantically in the different phyla — since in each phylum they would have been selected independently as 1st or 2nd persons, all of them having originally had the two values. In the following millennia, their multiplicity in each language phylum would have naturally led to continuous simplification, explaining why more stems are reconstructed in more ancient ancestor languages than in recent ones, in the frequent absence of innovations in their descendants. The independent simplification processes in different phyla would also explain why not all of them have exactly the same stock of stems.

The strange lack of representation of labial oral stops among the global stock of 1st and 2nd person pronoun stems could also find a plausible explanation. Among kinship appellatives, *papa ~ baba* 'father, grandfather, brother' is one of the most widespread (it is reflected in about 70% of the some 2,200 languages in our global database of kinship terminologies). As such, if our hypothesis is correct, one would also a priori expect *p-* pronominal stems initially derived from *papa* to be widely represented. However, there is another kinship appellative *tata ~ dada,* which at present cannot be distinguished semantically from *papa ~ baba,* and is nearly equally well represented worldwide. It is well known that true synonyms cannot coexist for a long time in the same language, and the survival of both *papa* and *tata* in many languages ensures that there must have been a difference between them, whether in their respective meaning or connotation. Perhaps this difference led to preventing *papa* from being used as a pronominoid, so that today the global pronoun stem stock still exhibits this typologically unlikely dearth in labial oral stops.

These are the reasons why we think that the very particular word class of 1st and 2nd person pronouns must have descended from preexisting words, and that kinship appellatives are the only possible ancestral class. While it is certainly beyond our proving and disproving capacities, we do not see another, more consistent evolutionary way through which personal pronouns might have appeared in human language.

Conclusion

In the conjecture presented here, not everything is of equal value. Consistently explaining the multiple reconstructions of pronoun stems in deep-level families, converging onto a handful of stem consonants at the global level, in the near absence of innovated pronouns in low-level families, seems to us to be one of its greatest strengths.

Other general points regarding the early prehistory of language, like the anteriority of kinship appellatives with regard to pronouns, and the phylogenetic reordering of the two articulations of language, we consider as pretty well supported by ontogenetic and evolutionary arguments.

The weakest point, in our opinion, certainly is the transition between kinship appellatives and pronouns through the speculatively assumed "pronominoid" stage, no evidence of which we may propose to the reader. More thought is needed about this stage, but not thought only, and if this point is by now the weakest it also might in the future prove the most fruitful. Here we are getting closer, both in the time sequence and the matter dealt with, to what most readers of *VJaR/JLR* are accustomed to: reconstructing ancient languages.

Our conjecture essentially relies on the observation of reconstructed pronouns in the Eurasiatic and Nostratic macrofamilies, as well as on a statistical observation of the low-level ancestral pronouns at the world level. Generalizing the Nostratic case is thus predictive. And

the prediction it makes may be confirmed or belied by the historical behavior of pronoun stems in other macrofamilies: in these ones as well, one should find very few innovated stems in the member subgroups, and most of the changes in pronouns should be restricted to loss of some of the macrofamily's ancestral stems. (Of course, this prediction is meant as a general rule, and may encounter counter-examples.)

A prediction it does not make, as highlighted in the beginning of the second section, is that each macrofamily would have changed some of its kinship appellatives into pronouns or pronominoids. However, since it assumes that the pronominoid stage — or stages — could have lasted for dozens of millennia and perhaps until recently, up into the Proto-Nostratic period, it should not be lost of sight that in some cases there might be after all a closer relation between each macrofamily's particular kinship terms and pronominal stems. Should it prove true in some cases, as the partly parallel distributions of 1sg *ŋgai and kinship appellative (ŋ)aŋa (see note 10 above) might suggest, it could lead to refine the present conjecture.

As a consequence, advances in the comparison of other ancient language families worldwide are likely to provide us with crucial insights allowing to validate or reject our conjecture.

Finally, let us hope that some readers will have enjoyed the tour, however risky it was, and will take us one day for another visit of their own to the earliest prehistory of language.

German V. Dziebel
Hill Holliday, Boston/Great Russian Encyclopedia, Moscow

On the Co-Evolution of Kin Terms and Pronouns

As Bancel and Matthey d'Etang (BME in the following) are perfectly aware of themselves, their conjectural model of the evolution of personal pronouns from kinship terms around 100,000 years ago is very hypothetical and speculative at this point. Hence, I don't intend my comments below to be a systematic critique of their ideas. They need to be commended for the bold attempt to tie several divergent lines of inquiry in order to reconstruct the beginnings of a historical process by which languages acquired such specialized lexical classes as kin terms and pronouns. Remembering my rather vociferous e-mail exchanges with Pierre and Alain over the evolution of reduplicated kin terms in the early 2000s, I was rather surprised to see us converge, by 2010, on the issue of the origin of pronouns from kin terms. Back in 2001, in my Russian book "The Phenomenon of Kinship" (Dziebel 2001), I drew on anthropology, archaeology, linguistics and population genetics to begin developing a very similar thesis, which can be summarized as follows. Due to the critical importance of social intelligence and kin-structured production in the evolution of *Homo sapiens sapiens* and the Middle-to-Upper Pa-

leolithic transition (see Moyer 2004), the recent evolution of modern humans from an original small deme characterized by a limited level of genetic diversity and a high propensity for kin-structured fissions and fusions (Neel & Salzano 1967; Weaver & Roseman 2005), the pervasive importance of kinship in foraging societies, the systematic nature of the historical transformations of kin terminologies, the undifferentiated referential properties of kin terms, the grammatical peculiarities of kin terms widely attested cross-linguistically (Jonsson 2001; Dahl & Koptjevskaja-Tamm 2001] and the empirical cases of crossover between kin terms, on the one hand, and other lexical classes, on the other (e.g., "kinship verbs" in some Australian, North American Indian and Khoisan languages [Evans 2000; Ōno 1996], "kinship pronouns" and "kintax" in Australian languages [Evans 2003], "kinship zoonyms" in Indo-European languages [Alinei 1985], etc.), kin terms may be thought of as a "language within a language," a phenomenological "proto-language" or an symbolic calculus, from which other lexical classes, including pronouns, common nouns, numerals, verbs, body part terms and proper

nouns could have evolved in the course of the evolutionary transition from hominin vocalizing to fully articulate sapient language. Although back in 1996 and 1997 I had toyed with various long-range etymologies possibly attesting to the evolutionary transition from kin terms to pronouns, body part names, etc., I obtained no solid results and restricted my claim to a strictly synchronic and philosophical level.

I have always thought that these kinds of speculations, albeit interesting and important, shouldn't strive too hard to become theories, as the level of uncertainty will always stay high regardless of the data at hand. Hence, I omitted this level of analysis from my subsequent English-language monograph (Dziebel 2007), although the overall focus on the peculiar linguistic structure of kin terms and their systematic historical transformations remained unchanged. As BME correctly note, kin terms share properties with common nouns (as in *Where is your Dad?*), proper nouns (as in *At the moment, Dad is out for angling*) and pronouns (as in *Mum wants Sonny to eat those peas*). Kin terms also show predicative qualities as, for example, in the expression *John and Bill are brothers* or in languages in which kin terms can be used only with appropriate pronominal markers of inalienable possession. As Gruber (1975, 40) wrote, "inalienable nouns are themselves underlyingly predicates." Ontogenetically speaking, kin terms are relational products, which, unlike elements of personal and spatial deixis, derive their meanings not from the acts of speech but from the acts of language acquisition when a child needs specialized cognitive tools that would allow him to correctly sort incoming stimuli into human and non-human, generic and specific, true and symbolic, subjective and objective "buckets" (see Hirschfeld 1989; 2001).

As an example of such an undifferentiated nature of kin term reference, I'd like to recount a personal story. I'm married to an American woman from a family with a long history of living in New England and with rather tight British, German and Dutch roots. Unlike the majority of Americans, they call 'father' not *dad* or *daddy* but *pop* or *papa*. My father-in-law is *papa* to his daughters and to his grandson (the son of my wife's sister). His wife, my mother-in-law, also refers to him as *papa* when speaking about him to those relatives who address him as *papa*. (His official name is Edgar but he prefers to be called Tony by those like myself who would call him by his first name.) In this one American family, a kin term, therefore, has become a personal name, a family nickname or an honorable title. My in-laws do not like to be called *babushka* and *dedushka*, when I speak to my daughter in Russian. My wife does not like it either. They would

like to be called *granma* or *grammy* and *granpa* or *granpy*. Only my Russian parents are *babushka* and *dedushka* to my daughter. Again, kin terms are treated as proper names, as they become rigid designators of specific individuals. At the same time, my wife and my in-laws realize that my father-in-law cannot be *papa* to my daughter, as in Russian 'father' is *papa*. My in-laws frequently make mistakes, correct themselves, apologize profusely and attract scolding from my wife who cannot understand why it is so difficult for them not to refer to my father-in-law as *papa* of my daughter. This reminds me of an anecdote quoted by Roman Jakobson (1971) in which one child forbids the other to apply pronoun "I" to himself: "don't call yourself I, only I am I." It has nothing to do with the word *papa* being a Russian word. English *papa* and Russian *papa* are the same word but in our English-speaking family an American father-in-law and a Russian son-in-law lay different claims to it. In the end, I am *papa* to my daughter but my father-in-law is *papa* to my wife, my sisters-in-law and my nephew (wife's sister's son) and *granpa* to my daughter. Forms *-pa* and *papa* here are the logical opposites of proper names. They are Jakobsonian shifters or deictic elements that change their reference depending on kinship grade, family status and speech role. In our family parlance, *papa* sometimes groups together me and my daughter and sometimes my father-in-law and his daughters creating collective shifters.

BME's paper contains a promise that high-level hypotheses and hunches pertaining to the origin and evolution of human language may, in fact, eventually become testable. In the meantime, a few critical remarks are in order.[1]

1. In all their writings, including the present one, BME treat kin terms as a self-evident class of nouns. Although it's true that kin terms are recognizable as such in every language, internal semantic, pragmatic and formal variation within this set is extraordinary. In 1871, Lewis H. Morgan pioneered the field of kinship studies with his famous *Systems of Consanguinity and Affinity of the Human Family*, in which he divided human kinship terminologies into three types (Turano-Ganowanian, or Iroquois, Hawaiian or Malayan

---

[1] A caveat should be made here: BME take for granted the existence of deep-level language families (macrofamilies, superphyla) such as Nostratic, Eurasiatic, Dene-Caucasian, Austric, etc. Although I'm very critical of these specific high-level groupings and the way in which some of long-rangers treat kin terms (see Dziebel 2008; 2009), I do not deny that low-level families are somehow related and will not touch on these vexed issues in this review.

and Descriptive or English), pointed to the different patterns of their geographic distribution and suggested a solution to the origin of American Indians (see Morgan 1871). Notably enough, in 1852, German linguist Johann Buschmann, also a student of American Indian languages, published a cross-linguistic study of reduplicated parental kin terms in which he divided them into labial and dental classes and argued that labial classes (*pa*, *ap* for father and *ma*, *am* for mother) were more prominent in the Old World, while dental classes (*ta*, *at* for father, *na*, *an* for mother) in the New World (Buschmann 1852). There is a striking parallelism between Morgan's focus on the semantic structure of kin terms and Buschmann's focus on the phonetic structure of kin terms. There is also a striking similarity between Buschmann's broad generalizations for kin terms and Joseph Greenberg's broad generalizations regarding Amerind and Eurasiatic pronouns.

In the 20th century, anthropologists developed various evolutionary models to explain the observable diversity of kin terminologies. Following in the footsteps of Morgan and others, I have been operating with a database of 2500 languages and published an updated historical typology of kinship terminologies on its basis (Dziebel 2001; Dziebel 2007). In addition, I amassed a comprehensive bibliography of kinship studies in anthropology, linguistics, logic, psychology, and other disciplines (see www.kinshipstudies.org). BME are apparently unfamiliar with this huge body of theoretical literature, as their studies pertaining to the evolution of kin terms do not reference it. This is a very odd oversight, which may have a negative impact on their theories. For instance, when they reconstruct meaning 'mother's brother' (other possibilities being 'grandfather' and 'older brother') for their proto-Sapiens etymon KAKA (Matthey d'Etang & Bancel 2002), they may find it surprising that 'mother's brother' as a separate meaning may not have existed in early human kinship systems, as it may not have differentiated yet from such meanings as 'man's sister's child' and/or 'spouse's father.' Similarly, 'grandfather' as a category subsuming 'father's father' and 'mother's father' but distinct from grandchild classes is likely a secondary evolutionary development absent from the earliest human kinship systems. Finally, 'older brother' may have existed only as two separate categories, namely 'man's older brother' and 'woman's older brother.' The evolution of human kinship systems is a history of categorical splits and mergers, and the reconstruction of reduplicated kin stems all the way down to the proto-Sapiens level without correlating them with the known global patterns of semantic change makes their whole exercise rather outlandish and hard to relate to.

2. In their writings, BME rely heavily on the argument that the omnipresence of formally reduplicated kin terms such as PAPA, MAMA, TATA, KAKA, etc. in world languages suggests their antiquity. First of all, they lump together CVC, CVCV and VCV kin terms as if they were interchangeable formations. Meanwhile, a quick look at Indo-European languages raises a doubt: while ancient Indo-European languages leaned onto the VCCV/VCV structure (Hitt *atta* 'father', *anna* 'mother', Goth *atta* 'father', Slav **otĭcĭ* 'father' with the regular loss of gemination), modern Indo-European languages tend to have CVCV (Russ *mama*)/CVC (Eng *dad*) structure, with no continuity between the ancient and modern reduplicative sets. The ancient set is comprised of basic terms (corresponding to modern *mother*, *father*, etc. that display complex morphology),[2] while the modern one encompasses hypocoristics co-existing with morphologically complex forms, which function as basic terms. The reduplicative shape of hypocoristics may be purely accidental (comp. Eng *bud* as contracted *brother*, without reduplication) and derived from various registers of adult speech. The reduplicative shape of ancient basic terms may be plesiomorphic and related to language acquisition by children.[3] It seems likely, therefore, that BME's sweeping approach lumps together several functional types of reduplicative formations. For long-range comparison it is important to go beyond similarities in sounds and meanings and identify exactly what *kind* of sound and what *kind* of meaning are in front of us.

BME also overlook the fact that reduplication is only one of several available surface strategies of organizing the phonetic and semantic content in kin terminological systems. Kin terms in general show a strong tendency to develop not only recurrent semantic patterns (Bifur-

---

[2] In the Gothic Bible *fadar* is used only once (Gal. 4,6) as an address form for God; the standard word for father is *atta*. The nominative form of *fadar* is unknown (Stiles 1988, 136, n. 3).

[3] Although BME argue strongly against Jakobson's (and Buschmann's, for that matter) theory that parental kin terms derive their phonetic properties from baby talk and are therefore convergent innovations, there is little doubt that kin terms guide various linguistic exchanges between adult relatives and children pertaining to the acquisition of language. For instance, among the Hopi the grandfather often calls his grandson *ikwa'a*, lit. 'grandfather' in the effort to teach him kin terms (Titiev 1967). In this case, we may be dealing with "semantic reduplication" as the role distinctions between grandfather and grandson are neutralized. Or, consider a common practice among the speakers of Arabic (or even English-speakers of Middle Eastern origin) to refer to their young sons as "fathers" and their young daughters as "mothers" (Littmann 1902, 134, n. 1) As children grow, these habits of speech fade away.

cate Merging, Generational, Crow-Omaha, Sliding Generational, Alternate-Generational-Self-Reciprocal, etc.) but also recurrent morphosyntactic patterns.

A special class of kin term formations called "Descriptive" or "Cumulative" literally describes a kin relation. Any language can produce descriptive compounds but some languages, e.g., Swedish, use these compounds as the only way to denote a relationship (*farfar* 'father's father', *mormor* 'mother's mother', *farmor* 'father's mother', *morfar* 'mother's father', etc.). Notably, *farfar* and *mormor* are reduplicatives, but, unlike PAPA and MAMA terms, reduplication occurs on the level of a syllable and represents an iconic representation of cumulative kin grades. These descriptive constructions tend to denote collateral and second-generation and higher kinship categories, but in some Niger-Congo and Nilo-Saharan languages, they are employed to generate terms for siblings as well (e.g., Mbay *nggonkom-* lit. 'child of my mother', *nggonbom-* 'child of my father') (Harvey 1991).

Another recurrent morphosyntactic pattern involves attaching derivational morphology to a basic kin term to arrive at a genealogically derived kin category (e.g., Lat *amitīnus* 'father's sister's son'/*amitīna* 'father's sister's daughter' from *amita* 'father's sister'). A subset of this compounding technology is the morphosemantic pattern whereby relational adjectives "little" and "big" are attached to a simple kin term with the resulting effect of connoting genealogical distance. This is the function of Eng *grand* and *great* in (*great-*)*grandfather*, (*great-*)*grandmother*, (*great-*)*grandson*, (*great-*)*granddaughter*. In many genetically unrelated languages one can encounter expressions "little father" for 'father's brother' and "little mother" for 'mother's sister'.

The pairing of semantically related kin categories is another typical morphosyntactic pattern. In such languages as Spanish, kin categories paired by sex tend to share a stem to which grammatical gender markers are applied (*hermano* B, *hermana* Z, *tio* uncle, *tia* aunt, *hijo* S, *hija* D, etc., with earlier Latin and Greek antecedents). Finally, some languages, e.g., Russian, have developed a whole slew of reduplicated kin terms *diadia* 'uncle', *ded* 'grandfather', *mama* 'mother', *papa* 'father', *tiotia* 'aunt', etc. that are not part of proto-Slavic kinship inventory and are recent formations driven by category mergers (*diadia* replaced the earlier pair *ujĭ* 'mother's brother' and *stryĭ* 'father's brother') and lexical diffusions (*mama* and *papa* were borrowed from French by Russian aristocracy in the late 18th century and then trickled down into general usage).

In many Australian languages, reduplication competes with compounding as two paradigmatic strate-

gies for describing kin relations (Harold Koch, pers. comm., 2009). Different morphosyntactic strategies for expressing kin relations frequently blend together. Consider Rus *babushka* 'grandmother', which contains a reduplicative stem *bab-* and a derivational morpheme *-ushka* with a diminutive meaning. Comparison between Rus. *tiotia* 'aunt' and Span *tia* 'aunt'/*tio* 'uncle' shows that two languages within the same family chose two different ways to utilize reduplication: across two segments of a word in the case of Russian and across a pair of semantically close categories in the case of Spanish. Finally, compounded kin terms typically undergo contraction and simplification, so that the original complex nature of these terms fades away.

Although it's true that reduplicated kin terms are very frequent and pervasive in languages, so are other morphosyntactic patterns. Moreover, data seems to indicate that some of these morphosyntactic patterns, including reduplication, descriptive compounding and pairing are relatively recent in the history of the low-order language families. It appears, therefore, that reduplication, compounding and pairing are recurrent morphosyntactic patterns that, at different points in time, had wide but always-different geographic distributions because they tend to emerge, compete with each other and disappear from languages under specific historical conditions. If a proto-Sapiens language had reduplicated kin terms, it may have also had compounded and paired kin terms. It must have also had unreduplicated, uncompounded and unpaired kin terms. If these surface patterns tend to recur in extant languages, then they must have recurred in ancient languages and in the incipient languages of our hominid ancestors. If reduplication is somehow more primitive than pairing and compounding, BME have never demonstrated it, and languages for which we have direct diachronic data suggest that reduplication, at least of the CVCV type, is a relatively recent strategy. BME's narrow focus on these high-order surface strategies (which they apparently perceive as basic) make BME's grandiloquent theories look rather trite. Finally, if reduplicated kin terms preceded unreduplicated terms, then how did the latter emerge?

In the context of BME's current paper, it remains unclear why, even if pronouns evolved from kin terms, should they necessarily evolve from reduplicated kin terms. BME acknowledge the lack of transitional forms between kin terms and pronouns and postulate hypothetical "pronominoids" to fill in the gap. The chances of filling the gap between reduplicated kin terms and pronouns are rather slim, since, to the best of my knowledge, pronouns are rarely, if

ever, reduplicated. The existence of recurrent morpho-syntactic patterns of kinship expression also calls into question BME's claim that human language evolved through two discrete stages — phonetic articulation and syntactic articulation. As far as kin terms are concerned, phonology, semantics and syntax are tightly intertwined, and there seems to be no reason to arrange these structural orders into an evolutionary sequence from simple to complex. The syntactic articulations of kin terms may be different from the syntactic role pronouns play, but they seem to have been with human kinship expressions since very ancient times.

3. BME report on the extraordinary stability of personal pronouns in low-level families and their relative divergence and progressive multiplication within deeper-level families. They perceive it as a serious contradiction. Meanwhile, why could not we solve this problem by reconstructing more complexly structured pronoun sets for deeper-level families and then postulate different paths of parallel devolution leading to more simplified pronoun sets in low-level families? There are strong reasons to believe that the evolution of kin terminologies from the Late Pleistocene to the present involved the progressive collapse of a great number of categories produced by various intersections of such variables of relative age, relative sex, self-

reciprocity, etc. (see Dziebel 2007, with further literature). During the same period of time new categories based on genealogical grades have emerged. If the analogy between kin terms and pronouns, for which BME advocate, is valid, then we could expect to find similar categorical reduction in pronoun sets, with such underlying deictic variables as physical distance, social distance, kinship grades, marriage classes, ethnic groups disappearing from more recent language families and branches (comp. familiar vs. respectful 2d person pronouns *tu* and *vu* preserved in French, *ty* and *Vy* in Russian and their loss in modern English, etc.). Kin-sensitive pronoun systems described for about 20 Australian languages are a good example. In Lardil, there are two sets of free pronouns in the non-singular: a harmonic set, for referents related in even-numbered generations, such as siblings, spouses, or grandkin; and a disharmonic set, for referents in odd-numbered generations, such as parents and children (Evans 2003, 24). Pronouns and kin terms may have been evolving independently as distinct sets with their own structure for a long period of time, overlapping in peripheral areas such as plural forms, before finally coalescing at the proto-sapiens or pre-proto-sapiens level when neither pronouns, nor kin terms looked anything like the linguistic items that the speakers of modern languages are used to.

Kirill Babaev
Russian State University for the Humanities, Moscow

Reply to Pierre Bancel and Alain Matthey de l'Etang

The origins of personal pronouns in the world's languages is certainly one of the most puzzling questions in diachronic linguistics. As well as the origin of language itself, this issue has seen a lot of most fantastic theories of the genesis of person marking. In the middle of the 19th century, Rudolf Westphal tried to explain Indo-European personal pronouns as the result of further development of personal verbal affixes *-m, *-t* and others which, in their turn, emerged in human language from nowhere, just to determine syntactic meanings. Nasal sounds, as Westphal put it, 'lay most closely' in the human speech apparatus, and that is why *sta-m* was the first personal form to appear. Next was the third person *sta-t*, with a person marker 'ly-

ing further' in the dental domain [op. cit. Дельбрюк 1904]. Later, in the beginning of the 20th century, Hermann Hirt believed that personal pronouns have common roots with nominal case markers and other affixes: the 1st person singular pronoun *me* was compared with the accusative *-m*, the instrumental *-mo-*, the dative / genitive *-om*, and the derivative suffix *-mo-* [Hirt 1932]. This could look promising upon first glance, and, remarkably, no Indo-European language presented any data to prove that this was wrong.

These are certainly not the most absurd versions of pronominal genesis even if we confine ourselves to Indo-European. Personal pronouns, in most languages, are so short that their comparison with any

other lexical or morphological item is usually limited to a single consonant and/or vowel. And, as long as the early stages of development of any proto-language have not yet been studied properly, there is always enough room for all possible speculations.

The only way to avoid such speculative hypothesis, however, is to base any discussion on the solid platform of knowledge.

Unless we can prove or deny a hypothesis by means of the comparative analysis, linguistic typology remains our most powerful weapon. Among the more than 6,000 languages of the world, many have undergone the genesis of personal pronouns within historically attested periods of time. Having analyzed many of them, one can make general and relatively exact conclusions about the most typical sources of personal pronouns. A detailed and comprehensive description of these sources is given, for instance, in [Helmbrecht 2004], where extensive examples are given from various languages of all continents. According to this and other analytic papers (including a brief survey in [Babaev 2009b]), personal pronouns can indeed develop from nominal items with 'human' meanings, such as 'man', 'speaker', 'body', 'slave' (for the 1st person) or 'master' (for the 2nd person), 'other' or 'the rest' (for the 3rd person). They can develop from deictic particles, including demonstrative pronouns, especially (but not exclusively) in the 3rd person. They also develop from composite constructions of various nature, such as 'it-is-me-who', 'my-body' or 'your-honour'. But almost nowhere do they develop from terms of kinship[1].

For instance, bearing in mind the numerous examples of genesis of the 1st person singular pronoun from the word 'man, person', we can subsequently compare Indo-European *me-, genitive *mene, with its Uralic, Altaic, Kartvelian and other cognates, with Nostratic *män(u)- 'man, male' postulated by Dolgopolsky [2008: no. 1422]. This version would enjoy phonetic support (CVC structure roughly the same) and make a lot of logical sense. Even if there is not enough internal comparative evidence from Nostratic languages to confirm the idea, typological evidence makes it at least worth digging further.

---

[1] For the sake of being precise, there are examples when kinship terms act as quasi-pronouns, in the languages of Southeast Asia: e.g., Vietnamese anh 'thou' or em 'me' literally mean 'elder brother' and 'younger suster', respectively [Cooke 1968]. But these are in fact used for both the 1st and the 2nd persons, and thus, do not carry exact person meanings. Colloquial Russian uses such words as отец 'father' and брат 'brother' for addressing people in the street, but this does not make them personal pronouns.

The disregard of typological data seems the key disadvantage of the analysis provided by Bancel & Matthey de l'Etang. They explain it with the note that typology might have been totally different at the stage when the human language was only forming. But we do not have any facts confirming this, and in the absence of such, typological verification is one of the few means to support such a hypothesis.

No sufficient logic is presented for the process by means of which kinship terms were transformed into personal pronouns for the first two persons, and the authors emphasize that this should be regarded as their weakest point. They suggest a 'pronominoid' stage when apellative kinship terms were shortened to be used as pronouns. Actually, this is a good point, because it is precisely the way that older nouns turn into pronouns in many languages of the world (whose history can be traced back). The well known Polish pan 'mister' and pani 'miss' is just such a 'pronominoid', in terms of the authors: it is extensively used as the 2nd person plural pronoun in polite and official speech. Spanish Usted 'you' < vuestra Merced 'your mercy' and Portuguese voce < vossa Merce also stand in the midway between being nominal constructions and pronominal forms. They still require that the verb be in the 3rd person, thus reminding us of their origin. The next step forward is seen in Romanian dumneata 'thou' which is followed by the verb in the 2nd person. However, again, no pronominal construction like that is based on former kinship terms.

Another shortcoming of the paper, which the authors are well aware of, is the focus on the languages of Eurasia that belong to only one macrofamily (Nostratic or Eurasiatic). Evidence from many other language families of the world is only touched upon briefly. However, without a more detailed analysis of pronominal system development outside Eurasia conclusions on the tentative human proto-language remain unbalanced. Indeed, in most branches of Nostratic personal pronouns are quite ancient and seem to have been immune to replacement for millennia. However, this is absolutely not the case in languages of East and South East Asia and the Pacific, where pronouns tend to appear and disappear quite often. These hundreds of languages are good counterexamples to the authors' hypothesis that personal pronouns are always stable in the language. In fact, they are in most cases subject to shortening, analogical change, and replacement by newly-formed pronominal paradigms or even, in rare cases, borrowed items.

Bancel and Matthey de l'Etang's hypothesis that such proto-human kinship terms as *mama, *tata, *nana

etc., could have been sources for personal pronouns in the world's languages may seem appealing because it is comprehensible accessible access. It will be much harder — in our view, quite impossible — to prove it. In the 1920s, Nikolay Marr, a well known Russian linguist, created the 'new doctrine of language' stating that all languages of the world originated from four 'basic elements': *sal, yon, ber, rosh.* The hypothesis was just as accessible, and plenty of works had been written to substantiate it. But it was never supported by linguistic evidence, and, in the end, was refuted by the scientific community.

Of course, we cannot say that personal pronouns *cannot* be derived from kinship nouns. We only note that the only solid proof to support this hypothesis will be reliable linguistic data.

Václav Blažek
Department of Linguistics and Baltic Studies,
Masaryk University, Brno

Reply to Pierre Bancel and Alain Matthey de l'Etang

The present contribution of Pierre J. Bancel and Alain Matthey de l'Etang continues their series of articles devoted to personal pronouns and kinship terms in a global perspective. They mention that the most frequent consonants forming personal pronouns in most of the world's language families are quite limited in quantity. Considering the fact that the so-called "nursery" kinship terms with the canonical shape *CACA & ACCA* usually consist of a similar set of consonants, they conclude that it is precisely these *Lallwörter* that could have been the source of personal pronouns. This is a very courageous hypothesis, with very little chance, however, to be proven.

One fundamental difference is the fact that pronouns generally conform to standard sound rules, contrary to nursery terms, e.g. Gothic *þu* "thou" vs. *atta* "daddy" respectively. Another objection is that the authors, trying to explain why some of consonants are so "favoured" in nursery terms, e.g. *B* in *BAB(B)A & AB(B)A* ±"father", but not in pronouns, blame this on the existence of the competing synonym *TAT(T)A & AT(T)A*. But the same could be said about the term *NAN(N)A & AN(N)A* ±"mother" that frequently "competes" with the synonymous *MAM(M)A & AM(M)A* — and yet, both *n-* and *m-* are quite frequently met in pronominal stems.

However, trying to analyze processes that the authors themselves date to around 100.000 years BP is only a virtual exercise of the same type as calculating the number of angels on the point of a needle. It would be more useful to evaluate the methodological approach. The method of 'mass comparison', promoted especially by Joseph H. Greenberg and his follower Merritt Ruhlen, is also dominant in the works of Bancel and Matthey de l'Etang. Naturally, it may be valuable to summarize rich lexical or grammatical data from many languages or various language families. It may also be useful to evaluate them statistically. Unfortunately, the method of 'mass comparison', gives us ample opportunities for postulating mistaken pseudo-cognates. This can be illustrated by several examples from other articles of Bancel & Matthey de l'Etang (in the present article we find only minimal concrete data, mostly replaced by 'impressionistic' quasi-reconstructions):

2008a, 435: Old Avestan *tā* "father" appears in *Yasna* 47.3, besides *ptā* in Yasna 45.11, 47.2, both from the IE nom. sg. **pətēr* (Barthomae 1904, 905; Hoffmann & Forssman 1996, 151). This means that it is **not** a nursery word.

2008b, 441: Kashmiri *bi* "I" does **not** belong among the *m*-forms where the authors place it. In reality, it reflects the IA plural pronoun **vayam* "we" = Shina *be* "we". The same shift from pl. to sg., but in the oblique stem, appears in Bhojpuri & Maithili *ham* "I" from **hamm < *amhᵒ < *asmᵒ* "us" (Masica 1991, 252–53).

2008b, 444: Nuristani forms of the 1st person of plural: Kâmviri *imo*, Kâtaviri *imu*, Sañuvîri *ima* are classified as examples of the 1st person *m*-pronoun. But historically they reflect the oblique pronoun **asmᵒ*, more precisely — such Middle Indo-Aryan case forms as **asmē, *asmānaṁ, *asmēhi* (Turner 1966, #986), derivable from **ṇs-(s)mᵒ*, where the main morpheme carrying the 1st person plural meaning is **ṇs-*. A similar

situation is seen in modern Iranian languages, beginning already from the Middle Iranian period (with the exception of Khotanese, where the original nom. *buhu* is preserved). The original nominative was replaced by the genitive **ahmākam*, still preserved as the gen. in Young Avestan *ahmākəm*, Old Persian *amāxam*, but used already as the nom. in Bactrian (*α*)μαχο, Sogdian /*māxu*/, Partian /*amāx*/, Middle Persian /*amā*/, and in all modern Iranian languages, e.g. Persian, Baluchi *mā*, Azari *āmā*, Kurdic Sorani *(h)ême*, Sivandi *hame*, Bashkardi *yamah*, Ossetic *max*, Yaghnobi *mox*, Ormuri *māx*, Shughni *māš*, etc. etc.

On the other hand, the authors would probably welcome the derivation of the IE pronominal root in **w-*, attested in the 1st person of pl. & du., from the primary cluster **mw-*, following Rasmussen (1999, 266). Although it is not generally accepted (maybe for the reason that this article was published in a journal which is not easy accessible), it represents a legitimate solution.

Summing up, I find useful the summarization of data collected by authors in some of their previous studies. But their present attempt at generalization is quite premature and misleading, since it is not based on the standard historical-comparative method, verified by typological data.

Pierre J. Bancel & Alain Matthey de l'Etang
Association for the Study of Language in Prehistory

Reply to replies

German Dziebel seems to find some convergences between his own work and ours. One may find interesting that another scholar having devoted so much work to kinship systems has independently stumbled on the conceptual links between kinship appellatives and personal pronouns. Many interpretations of our present and previous work made by Dziebel in his points 1, 2 or 3 are nevertheless abusive, and sometimes counterfactual.

For instance, in his point 2, he asserts we would claim that "formally reduplicated kin terms such as PAPA, MAMA, TATA, KAKA, etc. in [the] world['s] languages suggest their antiquity," which he apparently opposes on the ground that incompletely reduplicated terms [like APA, AMA, ATA, AKA and the like] seem to him more ancient: "[W]hile ancient Indo-European languages leaned onto the VCCV/VCV structure (Hitt *atta* 'father,' *anna* 'mother,' Goth *atta* 'father,' Slav **otĭcĭ* 'father'), modern Indo-European languages tend to have CVCV (Russ *mama*) / CVC (Eng *dad*) structure."

In reality, we never claimed that either syllabic structure ought to be more ancient. Rather, we used *(P)APA, (M)AMA, (K)AKA* — often though not always with the initial consonant between parentheses, sorry for this graphic inconsistency aiming at better readability —, and the like, as cover labels for CVCV, CVCCV, CVC, VCCV, VCV and even CV or VC attested terms, which cannot be distinguished from one another semantically:

only the root consonant makes a difference. Recall that we do not make reconstructions (see below our reply to Babaev and Blažek), but align words from a huge number of languages with comparable phonetic forms and meanings, and which must thus have some etymological relationship once hypothetical convergence due to babbling is cleared away, as we explained in several papers, some of which are freely accessible on the Nostratica website (www.nostratic.ru).

Moreover, the Indo-European data offered by Dziebel to support the greater anciety of VCV/VCCV forms are partial: besides his examples above, one finds Sanskrit *tatá*, Greek *mammê* and *pappa*, Latin *mamma*, *pappa* (the latter perhaps borrowed from Greek) and *tata*, and Old Breton and Welsh *tat* and *mam*. One thus has all the possible syllabic structures attested in writing since the most ancient times. Possibly this relative inconsistency in syllabic structure is partly due to interferences from babbling babies, even if most modern forms are individually traceable to very remote etymons or ancient borrowings — there may be little doubt, for instance, that the Modern English form *dad*, isolated within Germanic, was borrowed from Brythonic Celtic, whose modern members all preserved *tad* or *tat* (Vannetais Breton even has a 2sg possessive form *da dad* 'thy dad') since centuries.

Still in Dziebel's point 2, we evidently did not claim that personal pronouns should descend from fully re-

duplicated appellatives rather than from simpler ones! This interpretation must again result from Dziebel's own tendency to force the distinction between terms with different syllabic structures. Whatever the initial form of the concerned appellatives, they must have been reduced early to monosyllables — in this case, for obvious functional reasons —, since non-reduplicated pronouns are the overwhelmingly majority worldwide.

Our two other reviewers, Václav Blažek and Kirill Babaev, have overlooked our warning that our conjecture should not be assessed through an automatic application of comparative procedures. Rather, they essentially address our lack of regular sound correspondences and our use, instead, of Greenberg's and Ruhlen's multilateral comparison method. We certainly are Greenbergians, and since Greenberg (e.g. 1987, 1995) and Ruhlen (e.g. 1991, 1994b) themselves have abundantly and successfully defended multilateral comparison, we will not answer this critic in detail here. Let us only remind our readers that Greenberg's method allowed him to successfully classify several thousands of languages of Africa, the Americas, Oceania, and Eurasia.

The negative reaction of our critics is all the more difficult to understand as our article does not really rely on multilateral comparison. We have made explicit that our conjecture does not need that modern kinship appellatives descend from Proto-Sapiens — it only needs that kinship appellatives have existed at the time, which seems pretty unescapable given their present global distribution and their crucial role in language acquisition by babies. Nor does our account of pronominal roots rely on multilateral comparison: we have mainly dealt with pronoun roots reconstructed by Nostraticists themselves. Only the global statistic study of low-level ancestral pronoun roots *partly* relies on multilateral comparison (Ruhlen's lists quote reconstructed pronouns in language families where there are reconstructions), but, as stated in our paper, the stability of pronoun roots makes very unlikely that regular reconstruction would change much the statistic picture at the heart of our conjecture.

In this respect, Babaev's objection that in languages from Southeast Asia and the Pacific personal pronouns often are subject to much more change is true and interesting, but certainly does not constitute a major problem. As we underlined, language may function without 1st and 2nd persons, and these languages essentially do so, as had to do all existing languages before pronouns were invented — because they must have been invented at some point in language evolution, even if no doubt involuntarily and progressively. In the languages alluded to by Babaev, person marking is considered unnecessary and even

crude, and hence avoided as much as context allows. In cases demanding disambiguation, they are often rendered by periphrases. Let us note with Babaev that such periphrastic markers may end up as true pronouns, like Spanish *Usted* 'thou (honorific)' < *Vuestra Merced* 'Your Mercy,' or Romanian *dumneatá* 'thou (hon.)' < *Domnia Ta* 'Thy Lordship.' Paleolithic hunters-gatherers, however, may not be suspected to have used such appellatives as 'Your Mercy' or 'Thy Lordship' — they undoubtedly were, like all their modern counterparts, light years away from such grotesque obsequiousness. The only appellatives they may have used to disambiguate 1st and 2nd persons in discourse were those kinship appellatives still found in traditional agricultural and hunting-gathering societies, like *(T)ATA* 'Father/Grandfather,' *(M)AMA* 'Mother / Grandmother,' or *(K)AKA* 'Uncle/Elder Brother,' whose implications regarding generational position and relative age amount to acknowledge the (even unrelated) hearer's social status.[1] Babaev's objection finally reinforces our point.

There must have existed somewhere in Southeast Asia an ancestral language (or several of them), which, in the transitional period where 1st and 2nd person markers were in formation, decided for cultural reasons to avoid person-marking — just like several modern IE languages, including English, lost any trace of PIE 2sg *t- after having generalized the 2pl to address a single hearer. These ancestral languages transmitted person-marking avoidance to their descendants, and probably also influenced several neighboring languages such as Korean or Japanese — which, in spite of their clear Eurasiatic membership, also avoid marking 1st and 2nd persons and hence are the Eurasiatic languages with the poorest ancestral pronoun record. Because the stability of pronouns is nothing magical, but has been shown by Pagel & al. (2007) to be narrowly correlated with their huge frequency in discourse — in languages families where they are stable.

We cannot accept, in turn, Babaev's idea to derive Eurasiatic genitive forms *menV-* 'of mine' from Nostratic *män(u)* 'man.' Forms *menV-* 'of mine' are clearly built on *me-* '1st person' + *-nu* 'genitive,' a particle widely represented in Eurasiatic (Greenberg 2000: 130–7, Bomhard 2008: 283–6). Deriving the diverse meaningful elements of a compound form from meaningless parts of a simple root should be at odds with Babaev's own principles, as it is with ours.

---

[1] Old French still used *oncle* (< Lat. *avunculus* 'mother's brother' < PIE H_2ewyH_2- 'grandfather, mother's brother,' a likely derivative of *KAKA*) to address unrelated individuals. Reynart addresses Isengrim with an ironical respect as "mon oncle."

Václav Blažek found three "errors" in our data, not in the present paper but in previous ones. The first one is not an error: our interpretation of the data differ. Matthey de l'Etang *& al.* (2008) should not have allowed Avestan *tā* 'father' into their TATA series, Blažek argues, since it must derive from *ptā*, itself derived from PIE **patēr* and found in another Yasna. However, the coexistence of *tā* and *ptā* in contemporary Old Avestan texts is a weak clue that the former descends from the latter, and it certainly does not preclude the possibility that *tā* preexisted — as the two parallel Vedic forms *pitā* 'father' and *tatá* 'daddy' seem to show.

The two other errors found by Blažek merely result from his having misread us. 1) Contrary to his claim, we did not claim that Kashmiri *bɨ* 'I' derived from **m-*. In Bancel *& al.* (2008b: 443, last §), we specified that "Gujarati and [...] Kashmiri preserve the alternation between subject and non-subject forms (but replaced the derivative of the Sanskrit subject form *ahám* by new forms)," *contrary* to other Indic languages (Hindi, Punjabi and Marathi) we had just mentioned, which generalized a *m-* form in the nominative. It unambiguously implied that the Kashmiri subject form *bɨ* has nothing to do with **m-*. 2) As to the Nuristani *imu ~ imo ~ ima* forms, Blažek claims they derive from the PIE 1pl instead of 1sg. Undoubtedly — and here again, it is exactly what we said (Bancel *& al.* 2008b: 444, § 1).

Beyond that, we cannot suspect Blažek of trying to induce readers to think that there is no such thing as a 1sg marker **m-* in either Indo-European, Eurasiatic or even Nostratic. However, if one was to produce regular correspondences between attested forms to prove its existence, one could as well give up immediately the whole idea of a PIE 1sg **m-*. In all descendant languages, the 1sg pronoun paradigm is full of analogical replacements and reductions of the case and number subparadigms, so that not a single IE language preserves a series of forms directly derived from strictly equivalent PIE forms. The only formal element escaping these innumerable reshapings is precisely the root consonant *m-*, surviving unchanged in 99.6% of 494 IE languages we investigated. As such, its survival does not provide us with a series of regular correspondences, but with a list of highly differentiated words, all essentially having in common their initial *m-* — not a regular correspondence but an isolated identity, nonetheless revealing thanks to its generality. Thus, to sound correspondences ultras, **m-* '1st person,' the best preserved instance of common inheritance in IE languages, should be considered a hoax of lumping crackpots. In the particular case of 1st and 2nd person pronouns, Blažek's insistence on regular sound correspondences seems exaggerated.

The paper discussed here was written to illustrate how comparative-historical linguistics could contribute to the movement which has developed since two decades in search of the origins of language ability in humans — a question which has repeatedly been called the hardest scientific problem of our time. Respected archeologists, geneticists, primatologists, cognitivists, and synchronic linguists of several obediences and specialties actively work in this field, and have already come to exciting results. An apt summary of these outcomes takes for granted that the comparative method is limited by the usual 5,000-year ceiling (Kenneally 2007: 166–7), revealing the complete absence of long-range comparison from this fundamental debate. We are however persuaded that human language evolution cannot be understood independently from how actual words, morphemes and grammatical categories evolved, something which only language classification and etymological reconstruction may eventually tell.

Literature

Alinei, Mario. 1985. Evidence for Totemism in European Dialects // *International Journal of American Linguistics* 51 (4): 331–334.

Babaev K.V. 2008. *Proisxoždenie indoevropejskix pokazatelej lica.* Moscow, Eidos.

Babaev K.V. 2009a. Review of Bengtson J.D. (2008) // *VJaR/JLR.* 2, 139–43.

Babaev K.V. 2009b. О происхождении личных местоимений в языках мира // *Вопросы языкознания*, 4, 119–138.

Bancel P.J. & A. Matthey de l'Etang. 2002. Tracing the Ancestral Kinship System: The Global Etymon KAKA. Part I: A Linguistic Study // *Mother Tongue* 7, 209–44.

Bancel P.J. & A. Matthey de l'Etang. 2005. Kin Tongue. A Study of Kin Nursery Terms in Relation to Language Acquisition, with a Historical and Evolutionary Perspective. *Mother Tongue* 9, 171–90.

Bancel P.J. & A. Matthey de l'Etang. 2008a. The age of Mama and Papa // Bengtson (ed.), 417–438.

BANCEL P.J. & A. MATTHEY DE L'ETANG. 2008b. The Millennial Persistence of Indoeuropean and Eurasiatic Pronouns, and the Origin of Nominals // BENGTSON (ed.).

BANCEL P.J., A. MATTHEY DE L'ETANG & J.D. BENGTSON. In press. Back to Proto-Sapiens (Part 2). The Global Kinship Terms Papa, Mama and Kaka // MILICIC & JONES (eds).

BANCEL P.J., A. MATTHEY DE L'ETANG, K.V. BABAEV, V. BLAŽEK, J.D. BENGTSON & M. RUHLEN. Forthcoming. Reconstructing the Early History of Languages. The Eurasiatic Person Markers *m and *t.

BARTHOLOMAE, Christian. 1904 [1961]. *Altiranisches Wörterbuch*. Berlin: Walter de Gruyter.

BENGTSON J.D. & M. RUHLEN. 1994. Global Etymologies // RUHLEN 1994a, 277–336.

BENGTSON J.D. (ed.). 2008. *In Hot Pursuit of Language in Prehistory. Essays in the Four Fields of Anthropology in Honor of Harold Crane Fleming.* Amsterdam/Philadelphia, John Benjamins.

BENVENISTE E. 1946. Structure des relations de personne dans le verbe // *Bull. Soc. Ling. P.* 43–1. [Repr. in E. BENVENISTE, *Problèmes de linguistique générale,* vol. 1, 225–36, Paris, Gallimard.]

BICKERTON D. 1990. *Language and Species.* Chicago, Chicago Univ. Pr.

BICKERTON D. 2009. *Adam's Tongue. How Humans Made Language, How Language Made Humans.* New York, Hill & Wang.

BLAŽEK V. 2008. Gaulish Language // *Sborník prací filozofické fakulty brněnské univerzity / Studia minora facultatis philosophicae universitatis brunensis,* 38–65.

BOMHARD A.R. 2008. *Reconstructing Proto-Nostratic. Comparative Phonology, Morphology, and Vocabulary*, Vol. I, Leiden, Brill.

BOTHA R. & C. KNIGHT (eds.). 2009. *The Cradle of Language. Studies in the Evolution of Language.* Oxford, Oxford U. Pr.

BRIGAUDIOT M. & L. DANON-BOILEAU. 2002. *La Naissance du langage dans les deux premières années.* Paris, P.U.F.

BUSCHMANN, Johann C. E. 1852. Über den Naturlaut // *Philologische und Historische Abhandlungen der Königlichen Akademie der Wissenschaft zu Berlin* 3: 391–423.

COUPÉ C. & J.M. HOMBERT. 2005. Les Premières Traversées maritimes. Une fenêtre sur les cultures et les langues de la préhistoire // J.M. HOMBERT (ed.), *Aux origines des langues et du langage*, Paris, Fayard, 118–61.

D'ERRICO F. & M. VANHAEREN. 2009. Earliest Personal Ornaments and Their Significance for the Origin of Language Debate // BOTHA & al. 2009.

D'ERRICO F., C.S. HENSHILWOOD, M. VANHAEREN & K. VAN NIEKERK. 2005. *Nassarius kraussianus* Shell Beads from Blombos Cave: Evidence for Symbolic Behaviour in the Middle Stone Age // *J. Hum. Evol.* 48, 3–24.

DAHL, Östen, and Maria KOPTJEVSKAJA-TAMM. 2001. Kinship in Grammar // *Dimensions of Possession*, edited by I. BARON, M. HERSLUND and F. SØRENSEN. Pp. 201–225. Amsterdam and Philadelphia: John Benjamins.

DAWKINS R. 1976 [2nd ed. 1989]. *The Selfish Gene.* Oxford, Oxford U. Pr.

DE WAAL F.B.M. 1982. *Chimpanzee Politics. Power and Sex among Apes.* Baltimore, Johns Hopkins Univ. Pr.

DELBRÜCK B. 1904. *Введение в изучение языка*. СПб.

DOLGOPOLSKY, Aharon. 1964. Gipoteza drevnejšego rodstva jazykovyx semej severnoj Evrazii s verojatnostnoj točki zrenija // *Vopr. jazykoznanija,* 2, 53–63. Eng. tr.: A Probabilistic Hypothesis Concerning the Oldest Relationships among Language Families of Northern Eurasia // V.V. SHEVOROSHKIN & T.L. MARKEY (eds), 1986, *Typology, Relationships and Time,* Ann Arbor, Karoma.

DOLGOPOLSKY, Aharon. 2008. *Nostratic Dictionary*. Cambridge: McDonald Institute for Archaeological Research.

DZIEBEL, German V. 2001. *Феномен родства: пролегомены к иденетической теории*. СПб., 2001.

DZIEBEL, German V. 2007. *The Genius of Kinship: The Phenomenon of Human Kinship and the Global Diversity of Kinship Terminologies*. Youngstown, NY: Cambria Press.

DZIEBEL, German V. 2008. Reconstructing 'our' kinship terminology: Comments on the Indo-European material in A. V. Dybo's and S. V. Kullanda's *The Nostratic terminology of kinship and affinity* // *Алгебра родства* 11. C. 42–92. St. Petersburg.

DZIEBEL, German V. 2009. Refining the points of an ongoing debate about Indo-European, Nostratic and kinship terminological reconstructions // *Алгебра родства* 12. Cc. 6–19. St. Petersburg.

EHRET C. 2007. *Proposals on Niger-Kordofanian Pronoun History*. Paper pres. at 18th Int. Conf. of Hist. Ling. (Montréal, August 5–11, 2007).

EVANS, Nicholas. 2000. Kinship Verbs // *Approaches to the Typology of Word Classes*, ed. by Petra M. VOGEL and Bernard COMRIE. Pp. 103–172. Berlin and New York: Mouton de Gruyter.

EVANS, Nicholas. 2003. Context, Culture, and Structuration in the Languages of Australia // *Annual Review of Anthropology* 32: 13–40.

Gardner R.A., B.T. Gardner & T.E. Van Cantfort (eds.). 1989. *Teaching Sign Language to Chimpanzee.* Albany, SUNY Pr.

Gell-Mann M., I. Peiros & G. Starostin 2009. Distant Language Relationship: The Current Perspective // *VJaR/JLR* 1, 13–30.

Greenberg J.H. 1987. *Language in the Americas.* Stanford, Stanford Univ. Pr.

Greenberg J.H. 1995. The Concept of Proof in Genetic Linguistics // *Mother Tongue* 1, 207-16.

Greenberg J.H. 2000–2002. *Indo-European and Its Closest Relatives: The Eurasiatic Language Family,* vol. I (2000): *Grammar*; vol. II (2002): *Lexicon.* Stanford, Stanford Univ. Pr.

Grégoire A. 1937. *L'Apprentissage du langage. Les deux premières années.* Paris, Félix Alcan.

Gruber, J. S. 1975. Plural predicates in ǂHòã // A. Traill (Ed.), *Bushman and Hottentot linguistic studies* (pp. 1–50). Johannesburg: University of the Witwatersrand, African Studies Institute.

Hamilton W.D. 1963. The Evolution of Altruistic Behavior // *Am. Nat.* 97, 354–6.

Harvey, Judy A. 1991. Kinship Terms in the Chadian Language of M'Baye // *Georgetown Journal of Languages and Linguistics* 2 (3–4): 243–250.

Helmbrecht J. 2004. *Personal Pronouns — Form, Function and Grammaticalization*. Erfurt.

Henshilwood C.S., F. d'Errico, C.W. Marean, R.G. Milo & R. Yates. 2001. An Early Bone Tool Industry from the Middle Stone Age at Blombos Cave, South Africa: Implications for the Origins of Modern Human Behaviour, Symbolism, and Language // *J. Hum. Evol.* 41, 631–78.

Hirschfeld, Lawrence A. 1989. Rethinking the Acquisition of Kinship Terms // *International Journal of Behavioral Development* 12 (4): 541–568.

Hirschfeld, Lawrence A. 2001. On a Folk Theory of Society: Children, Evolution, and Mental Representations of Social Groups // *Personality and Social Psychology Review* 5 (2): 107–117.

Hirt, Hermann. 1932. *Handbuch des Urgermanischen*, vol. II. Heidelberg: Winter.

Hoffmann, Karl & Forssman, Bernhard. 1996. *Avestische Laut- und Flexionslehre*. Innsbruck: IBS 84.

Jakobson R. 1960. Why Mama and Papa? // B. Kaplan & S. Wapner (eds.), *Perspectives in Psychological Theory. Essays in Honor of Heinz Werner*. New York, Internat. Universities Pr. [repr. in *Selected writings of Roman Jakobson. I. Phonological Studies,* 538–45. The Hague, Mouton].

Jakobson, R. 1971. Shifters, verbal categories and the Russian verb // *Selected writings of Roman Jakobson. II. Word and language* (pp. 130–147). Paris; The Hague: Mouton.

Jonsson, Niklas. 2001. Kin Terms in Grammar // *Language Typology and Language Universals: An International Handbook*, edited by Martin Haspelmath, Ekkehard König, Wulf Oesterreicher and Wolfgang Raible. Vol. 2. P. 1203–1214. Berlin and New York: Walter de Gruyter.

Kenneally C. 2007. *The First Word. The Search for the Origins of Language.* New York, Viking.

Leroi-Gourhan A. 1964. *Le Geste et la parole,* vol. I: *Technique et langage.* Paris, Albin Michel.

Littmann, Enno. 1902. *Neuarabische Volkspoesie*. Berlin: Weidmann. (Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen. Philologisch-Historische Klasse 5 (3).)

Marean C.W., M. Bar-Matthews, J. Bernatchez, E. Fisher, P. Goldberg, A.I.R. Herries, Z. Jacobs, A. Jerardino, P. Karkanas, T. Minichillo, P.J. Nilssen, E. Thompson, I. Watts & H.M. William. 2007. Early Human Use of Marine Resources and Pigment in South Africa during the Middle Pleistocene // *Nature* 449, 905–8.

Martinet A. 1960. *Eléments de linguistique générale.* Paris, Armand Colin.

Masica, Colin P. 1991. *The Indo-Aryan Languages*. Cambridge: University Press.

Matthey de l'Etang A. & P.J. Bancel. & M. Ruhlen. In press. Back to Proto-Sapiens (Part 1). The Inherited Kinship Terms *Papa, Mama* and *Kaka* // Milicic & Jones (eds).

Matthey de l'Etang A. & P.J. Bancel. 2002. Tracing the Ancestral Kinship System: The Global Etymon KAKA. Part II: An Anthropological Study // *Mother Tongue* 7, 245–58.

Matthey de l'Etang A. & P.J. Bancel. 2005. The Global Distribution of (P)APA and (T)ATA and Their Original Meaning // *Mother Tongue* 9, 133–69.

Matthey de l'Etang A. & P.J. Bancel. 2008. The Age of *Papa* and *Mama* // Bengtson (ed.).

Maynard Smith J. 1964. Group Selection and Kin Selection // *Nature* 201, 1145–7.

Milicic B. & D. Jones (eds). In press. *Kinship, Language, and Prehistory. Per Hage and the Renaissance in Kinship Studies* (Proceedings of the Per Hage Memorial Session of the American Anthropological Association 105th Annual Meeting, held at San Jose, Ca, 15–19 Nov. 2006). Salt Lake City, Univ. of Utah Pr.

MORGAN, Lewis H. 1871. Systems of consanguinity and affinity of the human family // *Smithsonian Contributions to Knowledge*, 17. Washington, DC: Smithsonian Institution.

MOYER, Colin C. 2004. *The Organisation of Lithic Technology in the Middle and Early Upper Palaeolithic Industries at the Haua Fteah, Libya*. Ph.D. dissertation. University of Cambridge.

MURDOCK G.P. 1957. World Ethnographic Sample // *Am. Anthr.* 59, 664–87.

MURDOCK G.P. 1959. Cross-Language Parallels in Parental Kin Terms // *Anthr. Ling.* 1–9, 1–5.

NEEL, J. V., and F. M. SALZANO. 1967. Further studies on the Xavante Indians. X. Some hypotheses-generalizations resulting from these studies // *American Journal of Human Genetics* 19 (4): 554–574.

NICHOLS J. 2008. *N-M* Pronouns // M. HASPELMATH, M.S. DRYER, D. GIL & B. COMRIE (eds.). *The World Atlas of Language Structures Online.* Munich: Max Planck Digit. Libr. (http://wals.info/index). Chap. 137 available online at http://wals.info/feature/137. Accessed on 9 Oct. 2009.

ŌNO, Hitomi. 1996. An Ethnosemantic Analysis of Gui Relationship Terminology // *African Study Monographs, Supplementary Issue* 22: 125–144. Kyoto: Kyoto University Center for African Area Studies.

PAGEL M. Maximum Likelihood Models for Glottochronology and for Reconstructing Linguistic Phylogenies // C. RENFREW, A. MCMAHON & L. TRASK (eds), 2000, *Time Depth in Historical Linguistics,* vol. I, 189–208. Cambridge, Eng., McDonald Inst. for Archaeol. Res.

PAGEL M., Q. ATKINSON & A. MEADE. 2007. Frequency of Word-use Predicts Rates of Lexical Evolution throughout Indo-European History // *Nature* 449, 717–20.

PATTERSON F. 1987. *Koko's Story.* New York, Scholastic.

PINKER S. 1999. *Words and Rules. The Ingredients of language.* London, Weidenfeld and Nicolson.

RASMUSSEN, Jens E. 1999. Indo-European personal pronouns // *Selected Papers on Indo-European Linguistics*. Copenhagen: Museum Tusculanum Press, 256–275; first published in *APILKU* 6, 1987, 89–112.

RUHLEN M. 1991. *A Guide to the World's Languages,* vol. 1: *Classification.* Stanford, Stanford U. Pr.

RUHLEN M. 1994*a. On the Origin of Languages. Studies in Linguistic Taxonomy.* Stanford, Stanford Univ. Press.

RUHLEN M. 1994*b. The Origin of Language. Tracing the Evolution of the Mother Tongue.* New York, John Wiley.

SAVAGE-RUMBAUGH S. & R. LEWIN. 1994. *Kanzi the Ape at the Brink of the Human Mind.* New York, John Wiley.

SINGER R. & J. WYMER, 1982. *The Middle Stone Age at Klasies River Mouth in South Africa.* Chicago, Univ. of Chicago Pr.

STILES, Patrick V. 1988. Gothic Nominative Singular *brōþar* 'Brother' and the Reflexes of the Indo-European Long Vowels in the Final Syllables of Germanic Polysyllables // *Transactions of the Philological Society* 86 (2): 115–143.

TITIEV, Mischa. 1967. The Hopi Use of Kinship Terms for Expressing Sociocultural Values // *Anthropological Linguistics* 9 (5): 44–49.

TURNER, Ralph L. 1966. *Comparative Dictionary of the Indo-Aryan Languages*. Oxford: University Press.

WEAVER, Timothy D., and CHARLES C. Roseman. 2005. Ancient DNA, Late Neandertal Survival, and Modern-Human Neandertal Genetic Admixture // *Current Anthropology* 46 (4): 677–683.

Статья продолжает серию работ авторов, основной идеей которых является происхождение личных местоимений древнейших праязыков человечества из терминов родства. В данной публикации предлагается математическая модель, подтверждающая, по мнению авторов, возможность подобной деривации.

Book reviews / Рецензии

Valentin Vydrin
St. Petersburg,
Museum of Anthropology and Ethnography

Eine historische Phonologie der Niger-Volta-Sprachen.
By Henning Schreiber.
Köln, Rüdiger Köppe Verlag, 2008. 331 pp.

In today's Mande language studies, where scholars' main efforts are concentrated on synchronic descriptions and typology, a publication dealing with comparative historical matters is a rare bird. In this context, Henning Schreiber's book is an important and very positive event: it is the very first serious and systematic attempt at reconstructing the proto-language of the Eastern Mande group[1] at the phonological and morphological levels[2]. It should be mentioned that the Eastern group had remained, till this publication, one of the few big loopholes in comparative Mande linguistics (along with the Samogho group, Bobo and Soninke-Bozo). Therefore, it brings us much closer to a step-by-step reconstruction of the Proto-Mande language, being based on the rich data accumulated during recent decades through extensive field research, some by Henning Schreiber himself, some by other Mandeists.

Below I will try to combine an overview of the content and of the most interesting findings of the book with a critical scrutiny of its weak points and topics which, to my mind, need further elaboration and substantiation.

The **first chapter**, "Foundations and prerequisites", introduces the principal languages of the Eastern Mande group ("Volta-Niger") dealt with in the book under review: San ("Samo"), Bisa, Boko-Busa, and their dialects. In fact, each of these is a "macro-language" whose varieties are often not mutually intelligible and can be regarded as different languages. Schreiber also mentions two endangered languages of NW Nigeria, Tyenga (Kyenga) and Shanga, but, in respect to these, he merely writes that they "can be used only as external evidence, as far as information on these languages is scarce" (p. 5).

There follows a discussion of ethnohistorical evidence available for the peoples speaking Eastern Mande languages[3]. The author concludes that in the

---

[1] Henning Schreiber follows Raimund Kastenholz's labels for the groupings within Mande family: Niger-Volta for the Eastern group, and Mani-Bandama for the Southern group. In the current review, I will use traditional names, "Eastern" (EM) and "Southern" (SM; and "South-Eastern", or SE Mande, for the grouping of the next level of the hierarchy, which includes both these groups, conforming with William Welmers' usage of this term).

When transcribing forms in Mande languages, I follow the principles of the African variant of IPA: *j* is used for the voiced palatal affricate, and *y* stands for the palatal glide.

[2] To be more precise, I am aware of one previous attempt at a comparative study of the Eastern Mande languages, in the context of a general reconstruction of the initial consonants of Proto-Mande; it is Konstantin Pozdniakov's dissertation [1978]. Unfortunately, this work, very innovative in method and groundbreaking in results for its time, exists only in Russian and has never been published, for which reason it has passed almost unnoticed for Mandeists outside Russia. In any case, Henning Schreiber's study is based on new data that was not available in the 1970s, and his results differ very much from those of Pozdniakov.

[3] Concerning this division, I have a minor objection to H. Schreiber's interpretation of R. Caillé's note concerning "Bambara inhabitants (who) did not understand the Mandingo language at all": "One should assume that the designation «Mandingo» could have been used in the historical sources not only for the speakers of Bambara and Malinke, but also as an ethnonyme" (p. 15). In my opinion, the situation described by René Caillé may have two explanations. First, in the 19th century, "Bambara" was not a true ethnonym, but rather a term for non-Muslims, whatever language they may have spoken. Even today, in the NE Côte-d'Ivoire and in the Sikasso area of Mali, "Bambara" are not Manding, but rather non-islamized Senufo, in opposition to the Jula, who are Muslim Manding people. (The vagueness of ethnic identities in West Africa in pre-colonial times was the subject of an extensive discussion in French cultural anthropology, see, for example, [Bazin 1985].) The other reason may be the fact that Caillé might have tried to speak Mandinka (the Senegambian variety) with Bamana speakers. In this case, his failure seems to me quite natural: although both languages belong to the Manding group, they are pretty distant

---

past, the East Mande speakers constituted, most probably, a unity that stretched from the Volta basin to the Northwestern Nigeria. Then they were split by Gur-speaking groups, and some of them (Bisa and Boko-Busa) were partly integrated into political organisms of Gur groups. Oral tradition data show that the Bisa and the San are aware of each other's existence and former unity, while Busa-Boko do not remember about their former unity with Bisa and San (and this ignorance is mutual), which testifies to the fact that the first split from East Mande was Busa-Boko (in agreement with glottochronological data, see [Vydrin 2009]).

In the **second chapter**, "Vorgliederung" ("Subdivision"?), H. Schreiber makes a short survey of the history of the external and internal classification of Mande, with extra consideration for the position of Eastern Mande, and deals with the theoretical problems of lexicostatistics and its practical application to SE Mande. He repeats the standard reproaches addressed to glottochonology: its principle of a constant rate of replacement in the basic vocabulary, resulting in wrong dating for historically well-known events, such as the division between English and German, and its acceptance of similar forms (rather than true cognates) for evaluation. Unfortunately, the author, while blaming the "classical" variant of glottochronology for its well-known shortcomings, seems to be unaware of more sophisticated versions of this method where these (and other) shortcomings have been taken care of (primarily, Sergei Starostin's "improved glottochronology", see [Starostin 2000/2007]). As an alternative, he suggests a statistical method of "Neighbour Joining" borrowed from biology and actively used in comparative linguistics as of late.

In relation to the application of lexicostatistics to the Mande languages, Schreiber mentions the following difficulties: (a) compound words; (b) great variability of quantifiers and qualifiers; (c) consonant alternation. He makes up his comparative list based on his own data for Bisa and San, on Ross Jones' data for Busa, Boko, Bokobaru, Kyenga and Shanga, on H.-C. Grégoire's 600-word list for Southern Mande. This list then undergoes a lexicostatistical handling according to the method of "Neighbour Joining". The resulting classification manifests certain serious divergences from those suggested by other authors (Grégoire & Halleux 1994; Vydrin 2009) in what concerns the arrangement of Southern Mande lan-

guages. The most visible innovation[4] is the position of Beng which, according to Schreiber, should either be included into the Eastern ("Niger-Volta") group or singled out into a separate group within the SE Mande. This point merits more careful consideration.

Apart from the statistical evidence, H. Schreiber mentions the innovations that Beng shares with Eastern Mande (presumably, not attested in other Southern Mande languages): the disappearance of the implosive ɓ; the lexical innovation /zɔ̃/ 'fish'; and probably also a couple of shared retentions (Beng and Eastern Mande), such as the words for "stomach" and "hand", as well as "flesh".

However, this evidence turns to be shaky. In the Southern Mande group, it is not only Beng who lost the distinction between plosive and implosive bilabial consonants; another such language is Gban. It is true that in Beng, the word for "fish" is zī̀ŋ which probably represents a reflex of the same root as Bisa-Lebir zɔ̀, Southern San zɔ̀, etc. However, there is another South Mande language, Gban, that has zɔ̃́ 'fish', which looks much closer to the Eastern Mande forms and invalidates the idea of shared lexical innovations in this point. The word for "hand and arm" in Beng is wɔ̃, which doubtlessly reflects the same root as Wan ɔ̃, Mwan kɔ́ɔ́, Mano kɔ̀, Tura kɔ̀ɔ̀, Dan kɔ̀, Gban kɔ̃́. The only instance that remains is the Beng form nɔ̃̌ 'stomach' (a reflex of a root very well attested in Western Mande as well), but this single common retention seems too slim as evidence for reclassifying this language with the Eastern group.

Unfortunately, Schreiber does not provide the full list used in his lexicostatistical study, which prevents me from verifying his calculations[5]. I can simply mention that, according to my own lexicostatistical study based on the standard 100-wordlist [Vydrin 2009], Beng turns out to be an obviously Southern, not Eastern (nor "intermediate") language. Here are the percentages of cognates it shares with other Southern and Eastern languages:

---

[4] More precisely, in this point Henning Schreiber follows Oswin Köhler, whose idea to include Beng into Eastern Mande, expressed in 1975 and based on one single isogloss (the word for "flesh"), passed unnoticed outside Germany.

[5] I can only guess that the divergence in our results may stem from the fact that Schreiber has used H.-Cl. Grégoire's data, notorious for their unreliability. In all honesty, I do not understand why he did not use in his study the more precise South Mande data, collected by members of the Russian research group and abundantly represented in [Vydrin 2005b] and [Vydrin 2007] (the former publication, by the way, is mentioned in Schreiber's list of references) and in the dictionaries available on the website MandeSud (now: Mandelang) since 2004.

from each other, and their mutual intelligibility is very limited (or even impossible at first contact).

	Southern Mande									
	Yaure	Tura	Guro	Dan-Blo	Dan-Gweeta	Dan-Kla	Mwan	Mano	Gban	Wan
Beng	67	66	65	63	61	61	61	60	58	55

	Eastern Mande						
	Kyenga	Bisa-Lebri	San (Southern)	Boko	Busa	Bokobaru	Shanga
Beng	49	48	46	43	42	41	41

The **third chapter**, "Description", represents a lengthy (almost 120 pages) synchronous analysis of the morphology (noun derivation and compounds, the problem of adjectives, verbal derivation, complex predication) and phonology (syllabic structures, nasality, tone, consonants, vowels, morphophonology) of each single language: Boko/Busa, Bisa, and San. The data are compared and summed up in the **fourth chapter**, where reconstructions of the corresponding segments of the Proto-Eastern system are represented. For the convenience of discussion, I will consider the reconstruction of each segment together with its synchronic overview (language by language).

First comes **noun morphology** (pp. 66–81 for the synchronic presentation, pp. 195–205 for the reconstruction). Schreiber singles out five derivative suffixes in Busa-Boko-Bokobaru (plus a zero suffix, a conversion marker), six to eight in Bisa (six in Barka, eight in Lebri), and six in Southern San (data for the Northern San varieties being not available). Certain suffixes are classified by Schreiber as innovations, and four are reconstructed for the Proto-Eastern-Mande (PEM) level: *-{/da/} 'female', *-{/sa/} 'male', *-{/de/} 'proprietor, master of', 'nominal', and *-{/nı/} 'diminutive'. In addition, he reconstructs two suffixes and one prefix that were unproductive already at the PEM level: *-{/si/} 'uncountable', *-{/ka/} 'body part', *{/N/}- 'inalienable'. Finally, he reconstructs one inflection, *-{/li/}, whose reflexes, according to Schreiber, are represented in modern EM languages as agent noun suffixes, although he prefers to interpret them as verbal noun markers[6].

Schreiber's analysis, however detailed and interesting, needs some comments and elaboration. Thus, he projects the PEM suffix *-*sa* 'male; small' onto the Proto-Mande level, referring to [Dwyer 1988]. However, David Dwyer's reconstructed Proto-Mande form is rather **gure-n* 'man'. Even worse, in Proto-South-Western Mande the suffix *-*sa* (well represented in all the languages of that group) means 'female', while the suffix for 'male' is **sina*.

The reconstruction of the suffix *-*li* as a definite marker (thus, an article; a definite/referential article, represented by a front vowel, is widely attested in the Western Mande branch), albeit highly original and challenging, needs further substantiation. First of all, Schreiber regards the agent noun suffixes -*ri* in Bisa, Busa, Bokobaru and -*li* in Southern San as reflexes of the *-*li* in question, and to make the semantic transition smoother, he argues that "the agentive meaning proceeds however … not from {/li/} itself, but from the verbal root contained by the compound word. Therefore, the function of {/li/} is a nominalization" (p. 79). This argument seems rather strained: if -*li* is really a nominalization (rather than agent noun) marker, I do not understand why Boko /màán-pīi-lí/ should mean 'buyer' rather than *'buying', or why Bisa-Barka /dumo-ba-l/ should mean 'weaver' rather than *'weaving'. It is true that in many Southern Mande languages and in Manding, verbal noun or gerund suffixes with the -*LI* structure are very common (Mandinka -*ri/-ndi*, Bamana and Maninka -*li/-ni*, Dan-Gwèètaa -*dẽ*, Guro -*lı*, Mwan -*lē*, etc.), but it seems to be not the case for Eastern Mande.

Another piece of counterevidence is the fact that definite/referential articles which look like reflexes of *-*li* or *-*ri* are rather uncommon for the Mande family[7]. In the numerous Western Mande languages where such an article is attested, it is most often a front vowel, sometimes preceded by a palatal resonant (-*yí*). Raimund Kastenholz [1986] derives it from the demonstrative **kE*; I provide arguments for an origin in another demonstrative, represented by Mende *jì*, Soso and Jalonke *yì* [Vydrin 2006: 209–210]. Therefore, Schreiber's reconstruction looks somewhat mismatching: in EM languages, *-*li* does not function as an article, but rather as an agent noun marker, and in Western Mande, there is no article that looks like -*LI*.

On the other hand, Schreiber's interpretation (p. 198) of vowel Ablaut in Bisa-Barka (front vowel in

---

[6] Oddly enough, Schreiber does not include the plural markers of EM languages into his consideration, which is a serious loophole.

[7] The only language that has such an article (with two allomorphs, *nì* after -*ŋ* and *lè* after any other vowel) seems to be Beng [Paperno 2006: 43], but it looks like an innovation in this language.

singular vs. back vowel in plural, e. g. *bir* 'goat' — *bu-rɔ* 'goats', *gyir* 'man' — *gɔ-rɔ* 'men') as the result of adjoining a singular marker (historically, an article) seems quite plausible, although even in these cases, I would reconstruct this marker as *-i* rather than *-li* (I think that the final *-r* in the singular forms belongs to the root, not to the suffix, cf. Proto-Southern Mande *kɔ̃lɛ̃́* 'man', *ɓólì* 'goat'; Proto-South-Western Mande *sìniŋ* 'husband', *ɓOli* 'goat'). However, in other cases Schreiber's eagerness to explain the oscillation between *i* and *u* among different languages through the influence of the definite article seems too hasty. Thus, the Southern Mande forms for "animal, meat" given on p. 197 (borrowed from Grégoire 1990) are partly wrong and partly misinterpreted[8].

Schreiber's reconstruction of an "unalienable noun prefix" **N-* seems to be shaky as well. His key evidence here is represented by the forms for "woman" and "wife" in Southern Mande languages ("Mani-Bandama"): he assumes that the forms for "wife" are derived (presumably, at the proto-language level) from those for "woman". Let us consider these forms[9]:

	Beng	Gban	Wan	Mwan	Guro	Tura	Mano	Dan-Ka
woman	*lēŋ̄*	*lè*	*lē*	*lē*	*lī*	*lé* 'wife'	*lēē*	*dē*
wife	*nā̃*	*nà* /lā̀/	*nɔ̀* /lɔ̀/	*nā* /lā̄/	*nā* /lā̄/	*nɔ̃́ŋ́* /lɔ̃́ŋ́/ 'wife; woman'	*nā* /lā̄/	*nɔ́ɔ̂* /dɔ́ɔ̂/

It seems to me quite evident that what we have here are reflexes of two separate roots which differ not only in the nasality of their vowels, but also in the vowel itself, and often in tone.

Another bit of counterevidence is that, in many Mande languages, words for "wife" and "(one's) child" (another key word in Schreiber's argumentation) are not "unalienable", but free nouns. Such is the case with the majority of Southern Mande languages, languages of the Mokole group, and others.

Let us consider the list of PEM forms where Schreiber reconstructs the "unalienable prefix" **N-: *(N-)bina* 'horn', **(N-)bo* 'a wound', **N-bo.i* 'pus', **(N-)gero* 'bone', **(N-)gɔN(-nē)* 'finger', **(N-)gɔN.kĩa* 'wing', **N-ya* 'to end', **N-yen.ta* 'to sit down', **(N-)yɛ-si* 'medicament', **(N-)yẽ* 'nose', **N-yiN* 'a fear', **N-yɔ́* 'breast', **N-yɔ́-yi* 'milk', **(N-)yɔN.ta* 'sand', **N-lɛ* 'tongue', **(N-)lĩnĩ* 'soul', **N-lo* 'wife', **masıe* (< *(*N-wo-se*) 'broom', **N-wuli* 'tail', **N-wĩ* 'head', **(N-)wa(-ru)* 'blood'. Almost half of these words represent items that are typically alienable in Mande ("wound", "pus", "medicament", "fear", "sand", "broom"); there is even a verbal root in the list ("to end"). On the other hand, many typically unalienable nouns are not on the list, such as **dari* 'father', **boN* 'intestine', **ga* 'cheek', etc.

Yet another strong argument against the reconstruction of this prefix is of a typological nature: the relation between possessor and alienable noun (in Welmers' terminology, "free noun") is semantically more complex than a rather natural relation between possessor and unalienable ("relational") noun. Languages which grammaticalize this opposition tend to explicitly mark the former, not the latter. In the Mande family, to the best of my knowledge, there is not a single example where a particular morpheme would mark the syntactic relation "possessor — relational noun", while markers of the relation "possessor — free noun" are omnipresent[10]. An unalienable noun prefix in a Mande language would be anomalous.

Not very convincing is Schreiber's reconstruction of the PEM suffix **-si* for uncountable nouns. In fact, among the four "core" stems enumerated on the p. 202 ("flour", "sand", "salt", "dust"), two of Schreiber's reconstructions do not contain the suffix in question ("dust": **bu.ta,* "sand": **N-yɔN.ta,* pp. 325–326). In Scheiber's list of reconstructions, I have found the following forms which presumably contain the suffix **-si: *(N-)yɛ-si* 'medicament', **yi/wi(-si)* 'millet', **yɔ.bi-si* 'animal', **yɔ̃(-si)* 'fat', **nɔN(-si)* 'bowels', **nɔ-si* 'pregnant', **wĩ-si* 'flour', **wu-si* 'salt'. Three out of eight words ("animal", "bowels", "pregnant"), to my mind, do not belong to the semantic group in question; on the other hand, there are many uncountable nouns which have no **-si* at all (**N-bo.i* 'pus', **(go.)boN* 'excrement', etc.). Therefore, the heuristic value of this suffix is not high. The same can be said about Schreiber's PEM "body part suffix" **-ka,* attested only in four

---

[8] E. g., the Tura form is not *wʉ,* but *wìì*; in Tura back unrounded vowels do not exist. As for the Dan form *wī* (in Grégoire's transcription, *wʉ*), its back unrounded vowel regularly corresponds to *i* in other South Mande languages. There is, in fact, a Kla-Dan form *wù,* but its back rounded vowel can be easily explained through the influence of the initial labial consonant. Therefore, the Proto-Southern Mande form is certainly **wī,* and it is unnecessary to perceive a fossilized definite article in it.

[9] The forms are from my Mande Etymological Dictionary (ms.). H. Schreiber provides forms from H.-Cl. Grégoire's dissertation, usually without tonal marks and often in an erroneous segmental transcription.

[10] With the exception of very few languages (such as Gban) that do not distinguish between these two groups of nouns.

stems: **gɔN(-ka)* 'arm + hand', **mɛ̃(-ka)* 'body', **sɔN(-ka)* 'tooth', **wu(-ka)* 'neck' (and not attested in many other names of body parts: **gwaN* 'foot', **toro* 'ear'…). In my opinion, even if we reconstruct **si* and **ka* as separate morphemes (and do not consider them as simply parts of the roots), they should be regarded not as derivational suffixes, but as roots that have broadened their lexical semantics and begun to easily enter into compound relations with other roots; such situations are very common in Mande languages[11].

In dealing with noun morphology, Schreiber pays special attention to compound formation. I can only approve of this approach: it is true that no lexical, morphological and even phonological reconstruction of Mande is possible if compounding is not taken into account. However, the example of the stem **(gɔ)boN* 'excrement' ("human's excrement") presented to illustrate this phenomenon (p. 200) does not look very appropriate, for several reasons. First of all, the author provides forms borrowed from old publications with imprecise transcription (H.-C. Grégoire 1990; Prost 1953) and disregards newer and more exact sources[12], even if they are mentioned in the bibliography accompanying the book (I have to mention, quite immodestly, my own publication), cf.:

	PSM	Beng	Tura	Wan	Guro	Dan	Mwan	Gban	Mano	Yaure
Schreiber 2008	–	ɣbo	ɣbo	ßu	ßo	ɣbo	ɣbo	ɣbe	gbọ	–
Vydrine 2004	*gbʋ̄	gbō	gbó	bù	bʋ̄	gbō	gbō	(gbè)[13]	gbō	pʋ̀

The correspondence "Guro *b* — Yaure *p* — Wan *b* — Beng, Tura, Mwan, Mano, Gban *gb*" before upper vowels is not at all "unsystematic"; it reflects, quite regularly, Proto-Southern **gb-,* and it is superfluous to interpret the labiovelars as the results of compounding followed by contraction. It should also be mentioned that the reflexes of the root for "excrement" have an initial labiovelar consonant in all South-Western Mande languages (**kpó*), in Soso (*gbíí*) and in Northern Jalonke (*gwíí*); as such, the explanation of the initial labiovelar in Busa, Boko and Proto-Southern Mande as the result of relatively recent compounding does not work[14].

On the other hand, words for "excrement" in SM are used not only for human feces, but also for animal droppings, and in many languages, for waste in general (I am sure that it was the same way on the Proto-Mande level as well). Therefore, if we follow Schreiber's reconstruction, in Dan-Gwèètaa, *gbɛ̂ gbɔ̃̀* 'dog's excrement' should be interpreted etymologically as "dog's human excrement", and *mlʋ̀ gbɔ̃̀* 'rice husk' as "human excrement of rice", which seems to me to run against common sense. Another argument: there are no reflexes of the root **gɔ* 'human' in Southern Mande. There is, however, a root **k̄ɔlɛ́* 'man, male' (Dan-Gwèètaa *gɔ̄ɔ̃̀* 'man', Tura *gɔ̃́ɔ̃́* 'man', etc.), but an interpretation of the form **gbʋ̄* 'excrement' as originating from **k̄ɔ̄-bʋ̄* 'male ecrement' (as opposed to "female excrement"?) looks even more bizarre.

A short section (p. 81–86) deals with **the category of adjectives**. It should be noted, however, that Schreiber's criterion for classifying words as adjectives (possibility of being used with a copula in the predicative function) does not seem to me convincing. Why should one take the predicative use, which is not prototypical for adjectives, as the diagnostic one? Studies of adjectives (and their correlation with stative/qualitative verbs) in Mande languages have a long history (see, for example and for references: Tröbs 2008a, 2008b), and it is clear that much more subtle criteria are necessary here.

In his analysis of the **verbal morphology** (pp. 86–93 for the synchronic presentation, pp. 205–207 for the reconstruction), Schreiber again, for some obscure reason, lets alone inflectional morphology; he contents himself with derivation and compounding. First of all, he singles out derivative means available in modern EM languages. In Boko/Busa, there are:

---

[11] For a very detailed and accurate analysis, both synchronic and diachronic, of a couple of such semi-grammaticalized roots, see [Erman 2005]. It is unfortunate that the author of the book under review is not acquainted with this publication.

[12] Even for Bamana, a very well documented language, Schreiber gives the form *ßo* originating from (Prost 1953); the correct form is *bò.* For Bobo, he provides the form *sanga* from the same source, although a more accurately transcribed form (*sàngàa,* pl. de *sùn*) could have been quoted from a more recent and widely known publication [Le Bris & Prost 1981]. Such negligent treatment of the data of languages outside the EM group is quite characteristic of the book under review, which seriously depreciates Schreiber's attempts at external comparison.

[13] The Gban form may represent a reflex of a different root; its original meaning seems to be 'remainder, rest'.

[14] In some other roots, however, it is possible to reconstruct a historical transformation **gu > gb*; see Series 31 of regular correspondences in Southern Mande in [Vydrine 2004]. See also a highly instructive article by Denis Creissels that deals specially with the establishment of labiovelars in Mande languages [Creissels 2004]; unfortunately, this paper, despite having been published in the only periodical specializing in Mande linguistics and available online, has not been taken into account by H. Schreiber.

— "altrilocal"/intransitive suffix -lɛ (on p. 86, the author contests its intensive interpretation; in the comparative section, however, he attributes to this suffix the meaning of repetition or duration). Schreiber finds certain fossilized verbal forms in Bisa and San that serve him as evidence for the reconstruction of the suffix *-lɛ in PEM[15];

— reciprocal suffix -aa;

— reduplication expressing iterative meaning.

In Bisa, Schreiber singles out:

— the "stative" suffix -ta (actually, his reasons for interpreting its semantics as "stative" seem unclear to me);

— reduplication which expresses the plural of the direct object or subject NP. Verbal reduplication is also found in San; its meaning can be interpreted either as intensive or that of verbal plural. Schreiber hesitates to make any statements about the semantics of reduplication in PEM and admits that the verbal plural meaning of reduplication may result from the influence of Moore[16];

— Ablaut, which is "no more productive in any modern language", but can be established "because of systematic similarity of semantically close verbal forms" (p. 206). This point, which, to me, seems highly controversial, needs more detailed analysis. The author mentions, in this relation, five couples of verbs in PEM and two in Bisa. Forms with unrounded vowels seem to be regarded by Schreiber as initial (this is not expressed directly, but follows from Schreiber's data), and forms with rounded root vowels presumably result from Ablaut:

1) *mini 'drink' — *muli 'swallow';

2) *da 'learn' — *dɔN 'know';

3) *yar 'break' — *wu 'break';

4) *ta 'go' — *to 'leave, abandon' — *tola 'put a foot on something';

5) *ka 'hold, take' — *ku 'pull toward oneself';

Bisa:

6) zè 'beat' — zo 'thrash millet';

7) bon 'draw to oneself' — wurun 'take to oneself completely'.

However, if we examine these pairs in a broader context and in more detail, the idea of Ablaut as a derivational means grows less attractive.

Pair 1: The first root is well represented in all the groups of Mande family (cf., in this relation: [Cl. Grégoire 1990]), while the second does not seem to be found outside South-Eastern Mande. Cf. their forms in Southern Mande languages (my own data):

	PSM	Dan-Blo	Dan-Gwèètaa	Kla-Dan	Tura	Mano	Guro	Yaure	Mwan	Wan	Gban	Beng
drink	*mīnì (?)	mūū	mūū	mā̋ŋ	mí	mī̄	míní	mìnì	mī̄	—	—	mī́n
swallow	*mani (?)	—	mā̋ŋ̋	mā̋ŋ	mā̋ī̋	mānī̄	míní	?	mìì	bɔ̀nī̄ (?)	mlɛ̋	mī̀ŋ̋

If we are to postulate a derivation through Ablaut, we should postulate it on the Proto-South-Eastern Mande level rather than for PEM; however, the SM data testifies against any *i — *u type Ablaut. It seems plausible that both roots may be related by some kind of derivation, but it can hardly involve Ablaut[17].

Pair 2: the Proto-Southern form for "learn, teach" looks like *dālā, and it is not attested outside the South-Eastern branch, while "know" is *dɔ̄, and reflexes of the latter are widely spread among the entire Mande family (with the exception of South-Western Mande and Soso-Jalonke). It would be strange to suppose that Proto-Mande *dɔ̄ 'to know' could have been derived from a Proto-South-Eastern root *dālā whose meaning ('to teach, to learn') is more complex semantically, and that the Ablaut had to be accompanied by a truncation of the second syllable.

Pair 3: In Proto-Southern Mande, the verbs of the second pair are reconstructed as *yEli and *wīí or *wīí correspondingly. It is evident to me that the rounded vowel in the PEM reconstruction *wu results from a progressive assimilation after a labial consonant (cf. Dan-Blo, Dan-Gwèètaa wúú, Kla-Dan wúú; in the latter variety the modification *i > u is regular in this context).

Pair 4: *ta 'go' and *to 'leave, abandon' are PEM reflexes of the roots well represented in many Mande groups; their Proto-Mande reconstructions will be, most probably, *táxa and *tó. Here again, if we are to accept the Ablaut hypothesis, we should postulate it

---

[15] Unfortunately, the author is unaware of Dmitry Idiatov's work [2003] where a detailed synchronic and diachronic analysis of the verbal derivational morpheme -LA in Tura and in other Southern and Eastern Mande languages is carried out.

[16] Cf. a similar function of verbal reduplication in the South-Western Mande language Loko spoken in Sierra-Leone [Vydrine 2004: 66–67].

[17] Another strong hypothesis is a borrowing of the root for "swallow" from Proto-Kru into Proto-South-Eastern Mande, cf. the following forms for "swallow" in different Kru languages: Krumen Tepo mná, Jrewe mná, Krahn mlà, Grebo mlá, Gere, Wobe mlà, Niabua mānā, Bete-Daloa mlà, Bete (g) mɔ̀nà, Neyo mlā, Koyo mlā, Godie mānī̄, Dida mnā, Aizi mra.

on the Proto-Mande (rather than PEM) level, and even there, it seems doubtful.

Pair 5: the root **kun* or **kũ* 'catch' (and similar meanings) is broadly attested in the Mande family (SM, Samogho, Bozo, Manding, Vai), while **ka* appears to be an EM root. The chronological priority of the form **kũ* over **ka* makes impossible the derivation of the former from the latter through Ablaut.

Pair 6: In Bisa-Lebri, the former root is reflected as *zέ* 'kill; beat; play (musical instrument), ply', etc.; the latter as *zɔ́* 'thrash (millet); strike (iron); pound' [Vanhoudt 1999]. It may seem that these two forms provide a good illustration of the proposed Ablaut, however, reflexes of both roots in SM and in some EM languages rule out this hypothesis:

	PSM	Dan-Blo	Dan-Gwèètaa	Kla-Dan	Tura	Mano	Guro	Yaure
beat, kill	*cɛ̄ (?)	zλ̃	zλ̃	zὲ	zέ	zɛ̄	jɛ̄	tὲ̃
beat, pound	*zɔ̄̃	zɔ̄̃	zɔ̄̃	zɔ̀̃	zɔ́̃	zɔ̄̃	zɔ̄̃	sɔ̀̃

	Mwan	Wan	Gban	Beng	Southern San
beat, kill	dɛ̄	tέ	zὲ	dɛ̄	dɛ
beat, pound	zɔ̄̃	—	zɔ̀̃	zɔ̄̃	zɔ̄ 'push, move'

Pair 7: the only argument in favor of a derivative relation between these two roots seems to be a remark in André Prost's Bisa-Barka Dictionary (Prost 1950: 186): "**wurun** … sert de F(orme) PL(urielle) à *bon*". Apart from this judgment, I do not see any reason to regard *wurun* as a form derived from *bon* and not as a different root.

This leaves us without a single reliable minimal pair, making the reconstruction of vocalic Ablaut in the PEM verbs groundless.

Schreiber devotes much attention to the "complex predicates", i. e. idiomatic combinations of noun stems with verbal stems. He proceeds from the idea that such combinations might have played an important role in the history of East Mande languages, and that they should be taken into account in the reconstruction. I fully agree with him; my experience with other Mande languages (especially Southern Mande) shows that contraction of such combinations is a considerable source of verbal roots in modern languages. For instance, in Dan-Gwèètaa *gbōō* 'to defecate' doubtlessly ascends to **gbō ɓō* (lit.: "excrement + get out"). What represents a difficult problem here is the morphosyntactic status of such combinations in the synchronic perspective (which also has diachronic implications)[18]. H. Schreiber follows Klaus Wedekind in his interpretation of "complex predicates" in Boko/Busa as instances of incorporation. However, it seems to me that what we have here is not incorporation in the usual technical sense of the term, which implies that

the complex in question functions morphologically as one word, and that noun stems that can be inserted into the incorporating complex make an open list. In Boko/Busa, neither of these requirements is fulfilled: there is no morphological marking of the "one-wordness" of the complex, and the nouns which can appear in that position, although quite numerous, make a closed list. What we have here are, rather, more or less lexicalized idiomatic word complexes.

Contrary to the established tradition, Schreiber precedes his analysis of the phonemic inventories with those of the **rhythmical structure: syllabic structures, nasality, tone.** Let us follow him in this arrangement.

In Boko/Busa, he singles out the following syllabic types: N, CV, CVV, CCV[19]. In Bisa, these are N, CV, CV:, CCV, CVC, CV:C, and in San, they are the same as in Boko/Busa. For PEM, Schreiber reconstructs the "syllabic types" *N, *CV, *CCV, *CV.C$_{lenis}$V. He discusses the phonological status of long vowels in Boko/Busa (pp. 99–100) and seems to share Ross Jones' opinion that we have here combinations of two short vowels rather than a single long vowel (the main criterion being the availability of different tones on CVV sequences, while modulated tones on CV are impossible[20]). However, Schreiber's position looks contradictory: if VV sequences are not long vowels (or diphthongs), but sequences of vowels, CVV should

---

[18] On this matter, see my article concerning the status of preverbs in Dan-Gwèètaa [Vydrin 2009b].

[19] Busa forms with the structure CVVV, e. g. *zũaa* 'loose', provided by Schreiber himself on the same page, remain unexplained.

[20] Unfortunately, Schreiber does not apply the criterion of morphemic boundary, which often provides more relevant results.

not be regarded as one syllable, but rather as a sequence of two syllables, CV-V, and in this case, we should establish a separate syllabic type V (otherwise, Schreiber should explain which definition of the syllable he is following)[21]. The same problem is valid for Bisa (pp. 121–123) and San (pp. 149–150).

Schreiber's problem is clear, and it is not new to Mande studies: he desperately needs a one-way unit, intermediary between a syllable and a phonetic word (not the same as a morpheme, which is a two-way unit). Meanwhile, such a unit in Mande languages, a metric foot, has been discussed in special literature for quite a long time. One can mention Joseph Le Saout's work of 1979 (this author preferred the term of "syllabème"), as well as more recent publications [Vydrine 2005a; Vydrine 2009c; Kuznetsova 2007]. It is unfortunate that this discussion has passed unnoticed by Schreiber; otherwise, he could have avoided some logical inconsistencies. Another mismatch of this kind is the reconstruction of a "syllable type" CV.C$_{lenis}$V in PEM (p. 209) which should rather be interpreted as a disyllabic foot. In other words, Schreiber's "syllable types" should better be regarded as "metric foot types".

An important place is allotted in Schreiber's interpretation of EM data to the "underspecified" nasal element N that appears in all EM languages in the position of coda. The author interprets it as a vowel, for the reason of consistency of syllabic structures: "Otherwise, in Boko/Busa … appears no closed syllable, and the nasal consonant would be the only consonant that could appear at the end of a syllable" (p. 104). Although I assume that a foot-final -N might be a vowel (like in the majority of Southern Mande languages), it should be noted that Schreiber's argument is insufficient. Languages where CVN is the only type of closed syllable are not rare at all; in the Mande family, they are represented by Mandinka, Xasonka and Vai (to mention but some of them). To display the vocalic status of a foot-final -N, for example, in Tura, one can adduce the case of perfect marking on the Subject whose final vowel is then doubled and acquires an extra-high tone: Nɛ́-ɛ́ lő 'A child has come'. If the word ends in -N (transcribed as -ŋ), we have: Nɔ́ŋ-ŋ́ lő 'A woman has come'[22]. However, no evidence of this kind is given by Schreiber for EM languages,

which leaves the reader unconvinced. In fact, I agree with him that reconstruction of a foot-final -N (quite probably, of a vocalic nature) is highly plausible, but his data favoring this conclusion are insufficient (and his reconstruction of an "inalienable marker" *N- is not helpful in this respect; see my discussion of this marker above).

There follows a section on **consonant alternation** in EM (p. 221–243), which, to me, seems among the most controversial passages in the book. H. Schreiber recognizes that, in modern EM languages, no consonant alternation exists, but he follows Kastenholz in his idea that it could have existed in Proto-Mande (resp. in PEM), and this phenomenon could have been responsible for all kinds of irregularities in consonant correspondences (my criticism of Kastenholz's approach was published in *Journal of African Languages and Linguistics*, vol. 21, No. 1, 2000, pp. 106–118). I shall try to formulate my objections to his reasoning.

1. In those Mande languages that do have incontestable consonant alternation on the phonemic (not merely phonetic) level, primarily in South-Western Mande (SWM), the mechanism of alternation is quite clear: the morphological elements (a definite/referential article for nouns, a direct object 3SG pronoun for verbs) that trigger this alternation can be easily reconstructed on the Proto-SWM level[23]. For PEM, however, it is not clear at all. Schreiber's explanation of the model of PEM and Proto-South-Eastern consonant alternation (p. 239–243) is rather confused. The only morphological means that is credited for triggerring this alternation seems to be the "inalienable noun prefix" *N-. However, as has already been shown above, its reconstruction is highly problematic, and even if we accept it, I do not understand how this prefix could be responsible for the alternation in such roots (adduced by Schreiber) as "to rot", "dew", "field" or "water".

It turns out that the consonant alternation here is not a precise tool, but, rather, a magic wand, produced each time the linguist encounters a difficulty in the reconstruction process. It is probably a philosophical question: Do we want a reconstruction "at any price", or is our priority to make it as sound and verifiable as possible? In the former case, we can be completely satisfied with a "magic wand"; in the latter, we are obliged to cast it away and to look for a ruler instead.

---

[21] In some instances, Schreiber tries to avoid this contradiction by treating -V in terms of "morae", but the introduction of morae implies that we deal with long vowels and diphthongs (otherwise, he is operating with a non-standard definition of the term "mora"), which seems not to represent the opinion of our author.

[22] According to personal communication from Dmitry Idiatov.

---

[23] It is true that the word-final nasal element in the majority of SWM languages also produces the same effect as the article and the 3SG pronoun, but, to my knowledge, there is not a single Mande language in which this element by itself, without being supported by a grammatical morpheme (like an article), would result in a *phonological* consonantal alternation.

2. Schreiber's interpretation of the mechanism of consonant alternation in SWM languages is erroneous: he postulates a transformation Fortis → Lenis in the intervocalic position, even though it actually took place anywhere in the absence of a preceding *–N*, even in the context #_V (see in detail: [Vydrin 2006: 99–108]). Understanding this relieves us from the obligation to postulate a hypothetic vocalic prefix in order to justify the lenition in SWM languages. Since Schreiber's "two steps of consonant change" in PEM (p. 239:

I. C_f → C_l\V_V; III. V → $\emptyset$\#_C) follow the SWM model, invalidation of step I of this model depreciates the suggested PEM model as well.

3. The next objection, probably the most serious one, concerns the entire book and deals with H. Schreiber's presentation and handling of his comparative data. In his comparative series, we often find forms that certainly do not stem from one and the same root, e. g. (example 94, p. 240–241):

	PEM	Busa [Jones 2004]	Bisa-Lebri [Prost 1953]	South San [Prost 1953]	North San [Prost 1953]	Bobo [Prost 1953]	Beng [Prost 1953]
tail	*N-wuri (<*guri)	vlã	mun	mui	muli	pègè	pinon

	Tura [Prost 1953]	Wan [Prost 1953]	Gban [Prost 1953]	Guro [Prost 1953]	Yaure [Prost 1953]	Dan [Prost 1953]	Mano [Prost 1953]
tail	wi	wé	wi	wori	weri	wén/wi	won

It is evident that Bobo and Beng forms do not reflect the same root of the proto-language as the forms in Bisa, San, Tura, Wan, Gban, Guro, Yaure, Dan, and Mano. A further third root might be represented by the Busa form. However, the author does not offer the slightest comment on this subject; the forms are given just like that! Such situations are encountered practically everywhere in the book; I could easily fill many pages by quoting and critically analyzing the examples. Sometimes it is easy to sort out the reflexes of different roots, but in many instances it remains unclear whether the author regards the forms as cognates or not.

It is, of course, normal that at the beginning stage of a comparative work, the linguist arranges forms from different languages according to their English (or German, or Russian…) equivalent, i. e. their basic semantics; it is the "ethic" stage of a study. However, one of the goals of the work is the creation of a root dictionary, in which forms are arranged according to the proto-language roots they reflect; this is the "emic" stage. My impression is that H. Schreiber essentially remains at the "ethic" stage; at least, I have not remarked any attempt at sorting the forms according to their proto-roots.

This attitude results in an "extra-lumping" approach: Schreiber tends to regard forms as cognates if there is even the slightest phonetic similarity, or if there exists, somewhere in the Mande family, another form which could be regarded as an intermediate link. For example, we find the following forms for "neck" (p. 228): Looma (SWM) *kɔŋ/wɔŋ*, Busa *waka*, South San

wi, Beng *lo*, Tura *vele*, Guro *ɓɔlɔ*. For me, these are reflexes of at least four (may be, even six) different proto-roots[24], but Schreiber puts them together and looks for plausible explanations (such as hypothetic consonant alternation in the proto-language…) for this great divergence in the forms.

In my opinion, this approach greatly depreciates Schreiber's entire work: the reader is, in fact, confronted with a pile of raw data and obliged to perform an analysis that normally should have been done by the author.

Returning to the issue of consonant alternation in PEM, I have to conclude that this idea does not seem to me sufficiently grounded. Before more substantiated arguments are given, it can hardly be taken seriously.

Unfortunately, this lack of rigorousness in both presentation and interpretation of the data has negatively affected the entire **reconstruction of the phonemic inventory** of PEM. And it is regretful, because the thorough analysis of the phonological systems of modern EM languages in Chapter 3 disposes the reader towards expecting a solid comparative proce-

---

[24] In some cases, this can be proven very effectively. Thus, in South San there is a form *bɔlɔ* 'throat' which doubtlessly reflects the same root as the Guro word; therefore, *wí* 'neck' in the same language stems from a different root. In the SWM languages Mende and Loko we have, respectively, *mbóló/bolo* and *mbóró/boro* for 'neck', but cf. in Mende *kɔ́ngá/gɔnga* 'back of the head', in Loko *kɔnga* 'neck' (Koelle's data), which proves that *kɔŋ/wɔŋ* in Looma cannot reflect the same root as *ɓɔlɔ* in Guro, etc.

dure. The abovementioned flaws (above all, the indiscriminate approach to the establishment of cognates and the "magic wand" of consonant alternation) are magnified by the selectivity in the illustration of phonemic correspondences: as a rule, each one is illustrated by one or two series of cognates, while all the others are represented only by proto-forms on the PEM level (without their reflexes in the modern languages) given in a list at the end of the book. As a result, the author's lexical reconstructions remain, to a great extent, unverifiable (unless the reader undertakes anew the entire job of the compilation of a comparative EM dictionary). In compiling a volume of 300-odd pages, it would hardly be a problem to include a full-fledged comparative lexicon for a small group of languages[25]; such a lexicon would raise the value of the book enormously, even in the eyes of those who disagree with interpretations advanced by the author.

It should also be mentioned that the book is riddled with inaccuracies, misprints and errors. I will mention only a few.

On page 67, the Boko and Busa forms for 'give' are presented as /gba/; in fact, in both languages, the form is *kpá*.

On page 77, /gún.ɲɛːm/ is given as the Bisa-Lebri word for 'chicken'; in fact, this word means 'kitten', whereas 'chicken' is *kʊ̀r* |kʊ̀rì| [Vanhoudt 1999].

On page 103, in the chart of resonant allophones in Dan, *ɲ* appears twice instead of *ɗ*, which is completely misleading for the understanding of the mechanism of the nasal/oral variants distribution in this language.

On page 141, it is said that "Hidden (1986) proceeds, for the dialect of Lebri, from five phonemic oral vowels…", while in the chart below, seven vowels are given.

In Table 34 (p. 155), the imperfective negative marker in Southern San is erroneously given as /tá/ (instead of /bā/).

On page 163, in the charts of distribution of voiceless labials in two San varieties (Tougan and Toma), each consonant is given two lines (with different sets of vocalic contexts in each case), without any explanation concerning their differences.

On page 167 it is said that in San, /n/, "just like /m/, does not go together with /e/ and /o/…", yet immediately following that paragraph, the author mentions the forms *nè* and *mè*, blatantly contradicting his own assertion (p. 168).

On page 83, when presenting his example 41 (Bisa: /záá nyíntáā-m/), the author simply forgets to switch to a different keyboard layout, and its German translation looks as follows: |ɗer ʍeg ist sandig.|… etc.

Any comparative linguist knows how difficult it is, when dealing with the data of many languages at the same time, to avoid errors and misprints, and, to some extent, they can be regarded as an unavoidable evil. But still, one wishes the author had invested more care into the editing of the work before publication.

* * *

My review of Henning Schreiber's book has turned out to be much more critical than I had intended at the beginning. Now, reading my text again, I see that this issue is rather natural: I disagree with Henning not only in certain particular interpretations, but over some basic methodological principles of comparative linguistics as well.

However sharp my criticism may be, I still think that the reviewed book is an important step forward in Mande comparative linguistics, and even those hypotheses and interpretations of Schreiber that, to me, seem erroneous, will play a positive role in triggering fruitful discussions among Mandeists. And I hope very much that my severe judgments will not discourage Henning Schreiber from the continuation of his research in the field of Mande comparative linguistics, but, rather, encourage him to undertake a new attempt of Eastern Mande reconstruction, for which I would be happy to write a laudatory review.

References

Bazin, Jean. A chacun son Bambara // J.-L. Amselle & E. M'Bokolo (éds.). *Au coeur de l'ethnie. Ethnies, tribalisme et l'état en Afrique.* Paris: La Découverte, 1985, pp. 87–129.

Creissels, Denis. L'occlusive vélaire sonore *g* et les labio-vélaires (*w, gw, kw, gb, kp*) en mandingue // *Mandenkan*, 39, 2004, pp. 1–22.

Dwyer, David J. Towards Proto-Mande morphology // *Mandenkan*, 14/15, 1988, pp. 139–152.

Erman, Anna. Le grammaticalisateur *-ga* en dan-blo // *Mandenkan*, 41, 2005, pp. 41–61.

Grégoire, Claire. Etude comparative des items pour «boire» et «manger» dans les langues mandé // *Mandenkan*, 19, 1990, pp. 39–46.

Grégoire, Claire & de Halleux, Bernard. Etude lexicostatistique de quarante-trois langues et dialectes mandé // *Africana Linguistica XI, Annales du Musée Royal de l'Afrique Centrale, Sciences Humaines*, vol. 142. Tervuren, 1994, pp. 53–71.

Grégoire, Henrie-Claude. *Tonétique et tonologie d'un groupe de langues mandé. Etude théorique et expérimentale.* Thèse pour le Doctorat d'Etat, Paris III, 1990, 465 p.

Idiatov, Dmitry. *Структура глагола в языке тура (Западная Африка, семья манде): Элемент |LÁ|.* Магистерская дис-

---

[25] Certainly, in this case, the comparative series should not be given in columns, which is a very uneconomical; in the current presentation, as a rule, two comparative series take an entire page.

сертация. СПбГУ, 2003. [*The Tura verb structure (Mande, West Africa): the element ǀLÁǀ*. MA Thesis, St. Petersburg State University, 2003.]

Kastenholz, Raimund. Specification and the morpheme -(Y)É in Central Mande (and beyond) // *Afrikanistische Arbeitspapiere*, 7, 1986, pp. 97–113.

Kuznetsova, Natalia. Le statut fonctionnel du pied phonologique en gouro // *Mandenkan*, 43, 2007, pp. 13–45.

Le Bris, Pierre & Prost, André. *Dictionnaire bobo-français, précédé d'une introduction grammaticale et suivi d'un Lexique français — bobo*. Paris, SELAF, 1981, 415 p.

Le Saout, Joseph. *Notes sur la phonologie du Gouro (zone de Zuénoula)*. Nice: C.E.P.L.A.N., 1979, 76 p.

Paperno, Denis. *Базовый синтаксис языка бен (южные манде)*. [*Basic syntax of the Beng language (Southern Mande)*. MA theses. Moscow State University, 2006, 110 p. http://mandelang. kunstkamera.ru/files/mandelang/dipl_beng.pdf

Pozdniakov, Konstantin. *Языки манде: Сравнительно-исторический анализ. Кандидатская диссертация*. М.: Ин-т языкознания АН СССР, 1978, 149+210 с. [*Mande Languages: A historical comparative analysis*. Moscow: Institute of Linguistics, 1978, 149 + 210 p.]

Prost, André. *La langue bisa: grammaire et dictionnaire*. Ouagadougou: Centre IFAN, 1950.

Prost, André. *Les langues mandé-sud du groupe mana-busa*. Mémoires de l'Institut français d'Afrique Noire, Dakar, 1953, 182 p.

Starostin, Sergei. Сравнительно-историческое языкознание и лексикостатистика // С. А. Старостин. *Труды по языкознанию*. М.: Языки славянских культур, 2007, с. 407–447. [English version: Comparative-historical linguistics and lexicostatistics // *Time Depth in Historical Linguistics, vol. 1*. Cambridge: The McDonald Institute for Archaeological Research, 2000, pp. 223–265.] http://www.nostratic.ru/books/(140)Starostin_Glotto1.pdf

Tröbs, Holger. Bambara // Tröbs, Holger, Eva Rothmaler & Kerstin Winkelmann (eds.). *La qualification dans les langues africaines. Qualification in African Languages*. Pp. 13–28. Köln: Köppe, 2008a.

Tröbs, Holger. Duun // Tröbs, Holger, Eva Rothmaler & Kerstin Winkelmann (eds.). *La qualification dans les langues africaines. Qualification in African Languages*. Pp. 71–86. Köln: Köppe, 2008b.

Vanhoudt, Bettie. Lexique bisa-français // *Mandenkan*, 34, 1999.

Vydrine, Valentin. La réduplication des adjectifs dans les langues mandé // *Mandenkan,* 39, 2004, p. 61–67.

Vydrin, Valentin. Quelques recommandations méthodologiques concernant la description des langues mandé-sud // *Mandenkan*, 41, 2005a, pp. 1–22.

Vydrin, Valentin. Reconstruction of initial consonants in Proto-South-Mande // Koen Bostoen & Jacky Maniacky (eds.). *Studies in African comparative linguistics*. Tervuren: Royal Museum for Central Africa, 2005b, pp. 43–88.

Vydrin, Valentin. К реконструкции фонологического типа и именной морфологии пра-манде // Труды Института лингвистических исследований. Т. 2, Ч. 2. СПб: Наука, 2006. С. 3–246. [Toward the reconstruction of the phonological type and the noun morphology of the Proto-Mande. *Acta Linguistica Petropolitana*. Vol. 2, Part 2. St. Petersburg: Nauka, pp. 3–246].

Vydrin, Valentin. South Mande reconstruction: Initial consonants // Аспекты компаративистики 2. Orientalia et classica XI: Труды Института восточных культур и античности. М.: Издательство РГГУ, 2007. С. 409–498. http://www.nostratic.ru/books/(353)Journal11.pdf

Vydrin, Valentin. On the problem of the Proto-Mande homeland // *Вопросы языкового родства — Journal of Language Relationship* 1, 2009a, pp. 107–142.

Vydrin, Valentin. Превербы в языке дан-гуэта // *Вопросы языкознания*, 2009b, №2. С. 75–84. [Preverbs in Dan-Gwèètaa // *Voprosy jazykoznanija*, 2, 2009, pp. 75–84.]

Vydrine, Valentin. Le pied métrique dans les langues mandé // *Essais de typologie et de linguistique générale. Mélanges offerts à Denis Creissels*. Lyon: ENS Éditions, 2009c, pp. 53–62.

О. А. Мудрак
РГГУ / Институт языкознания РАН

A Historical Dictionary of Yukaghir
by Irina Nikolaeva
Berlin — New York: Mouton de Gruyter, 2006. 500 pp.

Публикация «Исторического юкагирского словаря» наконец-то ввела в научный оборот обширные материалы, собранные и обработанные И. А. Николаевой; начинает заполняться лакуна сравнительно-исторического языкознания для палеоазиатской юкагирской семьи. Книга состоит из следующих частей: сначала идут Acknowledgements со списком благодарностей (с. V), Preface (с. VII), где вкратце обрисована лингвистическая ситуация и лексикография юкагирских языков, Table of contents (с. IX), с развернутыми названиями главок, но без указания собственно местоположения этого содержания, Abbreviations (с. XI), Introduction (с. 1–83), Notes (с. 84), References (с. 87), Dictionary (с. 95–464), 2659 вхождений (фактически 2623, что оговорено во Введении), расположенных по алфавитному порядку реконструкций, включая освоенные русские заимствования, Index of meanings (с. 465), Language index (с. 499), где даны отсылки к этимологиям современных и реконструированных языков отдельных семей. Сразу вызывает настороженность присутствие лишь 10 чукотских этимологий, 1 корякской, 1 эскимосской, полное отсутствие нивхского материала, 6 праалтайских этимологий при большом количестве уральских схождений (включая подгруппы и конкретные языки).

В списке сокращений, разработанных автором и привычных для нее, вызывает большое неудобство обилие двубуквенных сокращений (большими буквами) с неравнозначной интерпретацией. Например: при T — тундр. юкагирский, есть NT — северно-тунгусский, TU — тунгусский (при U — уральский и FU — финно-угорский), TD — тундр. юкагирский по рукописному словарю Йохельсона, TJ — тундр. юкагирский по Йохельсону (1900), TK — тундр. юкагирский по Крейновичу (1958 и 1982 — но эти источники стоило бы различать), и TR — транзитивный; при K — колымский юкагирский, закономерно есть KD — колымский юкагирский по рукописному словарю Йохельсона, KJ — колымский юкагирский по Йохельсону (1900), KK — колымский юкагирский по Крейновичу (на этот раз только 1982), KL — Кличка, колымские материалы XVIII в., MK — Миллер/ Линденау, ко-

лымские материалы XVIII в.; при M — колымские («чуванские») материалы Майделя XIX в., по Шифнеру (1871b), MC — чуванские материалы Матюшкина (по Врангелю 1841), ME — Мерк, колымские материалы XVIII в., MG — прамонгольский, MK — Миллер/ Линденау, колымские материалы XVIII в., MO — омокские материалы Матюшкина (по Врангелю 1841), MU — Миллер/ Линденау, Усть-Янские (~ тундренные) материалы XVIII в., B — Биллингс, колымские материалы XVIII в., BO — Бенсинг, материалы XVIII в. и т. д.

Во введении сперва рассматривается организация вхождений. Сначала идут колымские формы (от современных к более старым фиксациям), потом тундренные с такой же подачей, однако, судя по данным словаря, материалы Биллингса, омокские и чуванские записи идут после современных схождений, и получаются фактически включены в более старые тундренные фиксации. Например, на стр. 196, в этимологии *joŋq-* 'nose, cape, promontory' после T *joγul*, TJ *joγul-* … идут MC *(j)ewo*, MO *niongol'*, BO *júngol*, KL *iogul, jojunkul*, B *iongul*, ME *jinkol* (? — *О. М.*), MU *jong, jongla, jongd-*, MK *jónckchol*. После тундренного материала действительно можно было бы дать MU (Миллер/ Линденау усть-янск.) *jõgla, jõg, jõg-d-* [jongla, jong, jongd-], но потом уже должны были пойти старые колымские фиксации — B *jõgul, jõgŭ-d-* [iongul, iongund-Биллингса], ME *jõkol, jõkŭg-d-* [jonkul, jonkungd-Мерка], KL *ioγul', jöjŭkul'* [иогулъ, ею'нкулъ Клички], MK *jõkxol* [jónckchol Миллер/ Линденау] и отдельно чуванские материалы BO *jŭgol'* [ю'нголъ Бенсинга], MC *jevo* [ево], *jẽ-d-arče* 'ноздри' [ендарче Матюшкина], и, наконец, омокский *(n)iõgol* [ніонголь Матюшкина]. Сразу надо отметить, что Бенсинг имеет прямые глоссы именно с чуванским материалом Матюшкина, например, *xóįl'* [хоилъ Бенсинг], *xajla* [хайла Матюшкин], при омок. *x̄a, x̄aja-, x̄ao* [кха, кхая-, кхао Матюшкин] (тундр. *qojl*) 'бог'; *amuŋ'* [амунъ Бенсинг], *-amик', -amuŋ'* [-амукъ, -амунъ Матюшкин], при омок. *tamna, -mi* [тамна, -ми Матюшкин] (тундр., *amùn, -damun*) 'кость'; *tọt-lị* attr. [тóтлй Бенсинг], *tota* [тота Матюшкин], при омок. *ti-* [ти- Матюшкин] (тундр. *tət*) 'ты (2sg.)';

tŭdoḷę, tuṇdele, túdoĺ, túdoḷe, tuṇdo-ɣo [ту́долѣ, тундолѣ, туйделе, ту́долъ, ту́долѣ, туйдого Бенсинг], *tuδoĺ* [тудолъ Матюшкин], при омок. *tati* (тундр. *tudəl*) 'он (3sg.)'; *júxčoń* [юхчонь Бенсинг], *jökčeń* [экчень < "εкчень Матюшкин], при омок. *jekõči-* [екончи- Матюшкин] (тундр. *íuku-*, *juku-*) 'мало, маленький'; *tĩ-ɣi̧, tíŋgõ̂, tíńỗ* [ты́нги, тынгй, ты́нгіε, тынέ Бенсинг], *tije-δo-ɣo, tine* [тыіедого, тине Матюшкин], при омок. *tŭń* [туунь Матюшкин] (тундр. *tuŋ*) 'этот'; *i̧li̧, ele-* [йлй, эле- Бенсинг], *ilen̂, illi, ol* [иленъ, илли, ол Матюшкин], при омок. *alla-* [алла- Матюшкин] (тундр. *əl, əĺ*) 'не (отрицание)'; *jáĺɣon̂, -jalɣan̂* [я́л'тонъ, -ялганъ Бенсинг], *jalon̂* [ялонъ Матюшкин], при омок. *jalom̂* [яломъ Матюшкин] (тундр. *jalôń*) 'три' и др.

К сожалению, не дается русского перевода юкагирских значений, хотя почти все лексические записи были сделаны именно с русскими переводами, а только усредненный английский, дающий общее представление о семантическом поле. Сначала говорится, что группы слов, имеющих одинаковые переводы, следуют без него после первого разъяснения значений. В примере на стр. 2 K *köŋe:* переводится как 'to chop' — реально 'расколоть пополам' (СЮР 36), затем без перевода KK *koŋie-* 'разрезать; расколоть' (Крейнович, Материалы 123, 126), KJ *koŋie-* 'вспороть, разрезать живот' (Крейнович, ЮЯ 282), T *köŋie-* 'to undo, to unrip, to cut', реально 'вспороть, вспарывать; перен. прокладывать дорогу' (Курилов ЮРС 2001, 162). Там же K *köŋu:* 'split, crack' — реально *lebên köŋũ* 'ущелье' («земляное ~») (СЮР 42), KK *kuŋu:* + 'scratch' — реально 'трещина' (Крейнович, Материалы 82). В этом же примере неясно, почему с помощью черты «|» отделена тундр. форма *köŋiere-* 'to cut' (по-видимому, опять же 'распороть' (?)).

К общим принципам транслитерации современных записей нет никаких нареканий; для человека, знающего кириллицу, довольно легко сделать однозначный пересчет (разве что «щ» — это [*šč*], а не [*čš*], как в таблице пересчета на стр. 6). Но относительно транслитерации первых памятников следует сказать, что стоило бы помечать присутствие «ъ» ~ ера знаком обратного апострофа [́], а не просто откидывать — в конце концов, это маркер конца фонетического слова; не опускать знак палатализации [́] после мягкой аффрикаты, если выписаны «ь» или йотированные буквы, знак «ѣ» ять передавать как *ję/ -ʾę̣* — йотированный узкий гласный, тем самым отличая его от «е». В записях XVIII века «г» следовало транскрибировать скорее как [*ɣ*]; как показывают аналогичные записи других языков Сибири, именно таким было нормиро-

ванное произношение. Передача через простые звонкие спирантов характерна и *для латинизированных записей. Для получающегося таким образом* [*ɣ*] *имеется 33 лобовых соответствия с юкаг. -ɣ-. Несоответствия, когда при юкаг. -g- отмечается* [*ɣ*], *в памятниках разбиваются на группы (исключая дублет ст.-колым.* oɣ-, õg- ~ *колым.* oɣ-, og-ê, *тунд.* oɣ-, eg-ô- 'стоять').

Для юкагирской фонотактики характерно распределение -ɣ- в заднем ряду / -g- в переднем ряду, ср. ниже слова с задним рядом в ст.-колымском (9 основ):

ст.-колым. *puɣač, põɣũn-* [pugatsh, poogunn- Bil], *puɣač, puɣŭn-* [pugatsch, pugunn- Me], *púɣač* [púgatsch MLK] 'теплый, горячий', *puɣáče-, puɣéčẽ-d-* [pugátsche-, pugétschend- MLK] 'жара', *puɣáče-jēgíla kutãʒa* [pugátsche-jengíla kutãʒa MLK] 'очаг' ~ *puga-*, тунд. *pugi-* 'горячий';

ст.-колым. *aɣra* [агра Кл] ~ колым. *ejrə-j*, dur. *eguž-uj*, тунд. *ewrəj*, inch. *egur-êj* 'ходить';

ст.-колым. *nŭgân, nuɣán* [núngean, nugán Bil], *nũman* [nunman Me], *nuɣãga* [нуганга Кл] ~ колым. *nùgən*, тунд. *niŋin* 'рука';

ст.-колым. *õleɣa* [oolega Bil], *uleɣa* [ulega Me], *uĺéɣa* [uljéga MLK] ~ колым. *uĺəga*, тунд. *uĺəga* 'трава';

ст.-колым. *íuɣóveja* [ljugóweja MLK] ~ колым. *jubugõj* 'сытый';

ст.-колым. *čãgnui* [tshangnui Bil, tschangnui Me], *čeɣnuj* [чегнуй Кл] ~ колым. *seŋnij, soŋnə-* 'чихать';

ст.-колым. *čáɣin-mon* [tsháginmon Bil], *čaɣ̃n-mon* [tschagiinmon Me] ~ колым. *cegin-mõj* 'глубокий';

ст.-колым. *õlɣala* [oolgala Bil] ~ колым. *ulgelə* 'моча';

ст.-колым. *-ču[ɣ]a* [-tschuda Me], *čŭga* [tschúnga MLK] ~ колым. *sugö, sugə* 'дорога, след', тунд. *sug-ôń* 'быстрый';

основы-клитики, встречающиеся во второй части сложений:

ст.-колым. *-iɣia* [-igia Bil], *-ikeje* [-ikeie Me], *ńójx-d-iɣé* [-igée MLK] ~ колым. *igəjə*, тунд. *igija(ŋ)* 'ремень, веревка';

ст.-колым. *puɣelвi, -buɣúelвi* [pugelbi, -bugúelbi Bil], *pouxelву, -buɣelвê* [pouchelby, -bugelbie Me], *buɣélvije* [bugélbije MLK] ~ колым. *pugəlbê*, тунд. *puguca(ŋ)* 'шерсть, шкура';

ст.-колым. (*lerũ*)-*gužia* [lerungudshia Bil], (*lerũ*)-*gučai* [lerungutschai Me], (*lire*)-*ɣiδej*, (*lire*)-*ɣuδej* [лире-гидей, лирегудей Кл], (*lirẽ́*)-*gĩʒejè* [liréngindschejè MLK] ~ колым. (*jur*)-*guʒɛ̌ʒə* 'звезда';

ст.-колым. *-mṍɣa, -mõɣá* [-móoga, -moogá Bil], *-muka, -muɣa* [-muka, -muga Me], *muɣè, muɣà* [mugè, mugà MLK] ~ колым. (-)*mugö* '(выступающая) часть тела'.

Изоглоссы с колым. (3 основы):

ст.-колым. *liɣai* [ligai Bil, Me] ~ колым. *ligəl* (Крейнович, ЮЯ 274), *liɣəj*, тунд. *lugəj* 'старый';

ст.-колым. *moɣo* [mogo Bil] ~ колым. *moɣo*, тунд. *moɣo(ŋ)* 'шапка';

ст.-колым. *neɣir maɣil* [negir magil Me] ~ колым. *ńêr*, *ńerə* 'одежда'.

Остающиеся несоответствия: ст.-колым. *u̯ɣurči-ra* [agurtshira < "ugurtshira Bil] ~ колым. *ugurcə-rā* 'тополь', ст.-колым. *póɣul* [pógul MLK] ~ колым. *pogi͡l* 'хозяин'. Но они также могут трактоваться и как заднерядные основы.

Можно с большой степенью уверенности утверждать, что и дентальный «*d*» (не после носового) был ослаблен и имел неполную смычку, ср. ст.-колым. *pălaδ*, *poluδ* [pallad, polud Bil] при колым. *polut*, тунд. *pelur* ~ *pölur* 'муж, старик', ст.-колым. *čašãɡaδaj-k* imper. sing. [tshashangadike Bil] при колым. *šašaɣə-daj-m*, тунд. *sisaɣa-rəj-m* 'расколоть, разорвать', ст.-колым. *aδδi* [addi Bil], *aδδy̆* [addy̆ Me], *a[δ]ii* [азыи < "адыи Кл] при колым. *adi*, тунд. *war-i* (*warul*) 'крепко, крепкий', ст.-колым. *máeδ* [máëd MLK] 'камень' при тунд. *mêrī*, *mêriŋ* 'напильник', ст.-колым. *miδĩ̆ža* [midinzsha Bil], *miδĕ̆ša* [midénscha MLK?] при колым. *midə̀ʒə*, тунд. *muriʒə(ŋ)* 'иголка' и др.

Особый момент — это позиция после выписанного носового, где несомненно был представлен смычный вариант. Надо заметить, что через написание -ND- (как дублет может быть и -NT-; N — носовой, D — смычный звонкий, T — смычный глухой) в записях латиницей и кириллицей XVIII в. передаются как действительные сочетания согласных, отражаемые и в XIX–XX вв., так и простые звонкие, которые имеют чистый звонкий смычный вариант в языках потомках. Это похоже на фиксацию автоматической слабой назализации перед звонкими смычными и аффрикатами. Любопытно, что примерно в это же время европейцами фиксируется аналогичная назализация в японском произношении перед звонкими (которые не всегда восходят к сочетаниям с носовым по исторической фонетике). Возможно, это ареальная фонетическая изоглосса, позже исчезнувшая.

При подаче словарного материала (стр. 7–8) выделяются основы, но не даются их реальные сочетания с аффиксами, хотя бы в назывной (словарной форме), что является неким опрощением. Из важных непредсказуемых характеристик в современных языках не отмечается способ образования мн. ч. или магнификатива у имен существительных, а также способ образования первичного каузатива у глаголов, что является важной характери-

стикой, позволяющей разбить на классы производящие основы.

Очень информативна и интересна 2 часть Введения, называющаяся «Sources of the Yukaghir material». И. А. Николаева провела интенсивную работу в архивах, и ей удалось удачно сгруппировать памятники и проследить направления заимствования и вторичной переписки исходников. Учитывая, что эти списки не так многословны, хотелось бы иметь их свод, но именно в исходной записи, а не только в транслитерации с опусканием точек и других помет над буквами, ведь почти у каждого автора была своя орфография и способы обозначения важных для него оттенков произношения. К сожалению, в данном словаре лексический материал из этих списков присутствует лишь в словарных гнездах, нет также и указателей по старописьменным источникам.

В 3 части Введения «Basics of Kolyma Yukaghir phonology» с опорой на колымские данные даются основные данные по системе фонем колымского и тундренного юкагирских языков. Сначала разбирается вокализм. Говорится, что колымский и тундренный имеют идентичные системы вокализма. Приводится таблица:

	передние		задние	
	неог.	огуб.	неог.	огуб.
высок.	*i i:*			*u u:*
невысок.	*e e:*	*ö ö:*	*a a:*	*o o:*

Шва *ə*, графически записываемое как *-e-* в непервом слоге и иногда в первом слоге служебных слов (*mət* 'я', *tət* 'ты'), фонемой не считается.

Это существенное упрощение реальной ситуации. Во-первых, по данным информантов, в тундр. гласный *-u-* в подавляющем большинстве случаев первого слога сдвинут в средний ряд и приближается к произношению японского *-и-* или варианта русского *-ы-* огубленного; как задний отмечается в словах *qularqa(ŋ)* 'чайка', *culal / culon-* 'горностай'. Гласный *-e-* — средне-нижнего подъема с небольшим сдвигом в средне-передний ряд. В отличие от него, гласный *-ə-* имеет два варианта в зависимости от сингармонистического ряда слова — действительный *-ə-* средне-переднего ряда, среднего подъема и *-ъ-* среднего (~ средне-заднего) ряда среднего подъема. Гласный *-ö-* является гласным средне-переднего ряда среднего (~ средне-нижнего подъема) (как бы вариант русского «э» огубленного, в случае артикуляции на среднем подъеме иногда записывается как *-e-* (= [ə]). Гласный *-a-* — нижнего

подъема, а гласный -*o*- — средне-нижнего подъема средне-заднего ряда.

Кроме простых кратких гласных, представлены дифтонгоиды -*ê*-, -*ȏ*-, -*ô*-, -*â*-, записываемые как -*ie*-, -*иö*-, -*ио*-, -*ia*- соответственно в большинстве современных словарных источников (последний отмечен в тундр. в словах *kâjań*, *kâja-ʒəĺə* 'передний, первый', *ekâ* 'старшая сестра', *tânu-j* 'прыгать на одном месте', в колым. *polundâ* dim. 'старик' наряду с переднерядным *polundê* и некоторых др.). Есть, наоборот, ровные долгие гласные, сдвинутые в нижне-средний подъем, они отмечены в словах: тунд. *sūsēj-m* 'свалить, снять, бросить', *pilēj-m* 'вытереть', *cindilikēn* 'рябчик', *icē(n)* (pl. *icân-pə*) 'провидец', *cumurcē* 'холмистая местность', *kôlekē* 'пупок', *qôlēv-m* 'убить (эвфм.)', колым. *tētē* 'медведь' (если и редупликат, то не синхронный, ср. ительменское **mäm-č* 'медвежонок'), тунд. *ȏŋ* = колым. *ȏ* 'штаны', тунд. *ō-m* = колым. *ō-m* 'черпать', тунд. *ōrə-m* = кол. *ōro-m* 'указать, назначить', тунд. *ōriń́ə-j* 'плакать', *mōrqoń* 'один', *ȏrul* 'крик', колым. *ȏŋȏj* 'молодой, юный', *abōj* 'старший', *udōj* 'привычный', *ȏžɨ̄* 'вода', *jō* 'голова'. Полностью соглашаясь с автором с регулярностью морфонологических чередований по ряду типа -*ê*-/ -*ā*-, отмечу, что это историческое фонетическое чередование с переходом старого долгого **ē > -ie-*, а не реализация долгой фонемы -*ē*-.

Во-вторых, чередующийся с нулем гласный не обязательно является автоматическим. Это хорошо видно на чукотском материале (соседнем языке), отмечается и в славянских языках (ср. *земель* gen., где второе -*е*- автоматическое, и *пень*/ *пня*, где гласный -*е*-, наоборот, продолжает развитие старой фонемы). При постулировании юкагирского автоматизма не получат простого объяснения случаи тундр. *ɨ̄sɨ̄-ń́əj* 'с грудью', *ô-ń́əj* 'с ребенком', *kuku-ń́əj* 'с чертом', *sawa-ń́əj* 'со шкурой', *poɣo-ń́əj* 'с пеплом', *puŋ̣ə-ń́əj* 'с супом', *cit-nəj* 'длинный' при *cicəɣə-c* 'удлиниться', колым. *ȏžɨ̄-ń́əj* 'с водой', *tan-ń́əj* 'с долгом', *inmu-ń́əj* 'с рогом', *qoli-ń́əj* 'с шумом' и тунд. *oʒə-ń́i* 'с росой', *ciń́icə-ń́i* 'сумрачно', *aɣarə-ń́i* 'с душой', *jarqa-ń́i* 'со льдом', *monqa-ń́i ~ monqa-ń́ɨ́* 'с холмами', *môqa-ń́i* 'с рыбой чиром', *lejgumu-ń́i, nejgumu-ń́i* 'с сумерками', *nonɣa-ń́i* 'с табаком', *qawarqa-ń́i* 'с ямами', колым. *ń́āśə-ń́i* 'жадный (= единоличник)', *pömžilə-ń́i* 'вышитый кружками', *ejmə-ń́i* 'с ценой' и т. д. Здесь видно, что выпадение или невыпадение последнего «автоматического» гласного основы не предопределено видом последующего суффикса, а вид суффикса лишь частично определен видом конца основы. На выбор вида суффикса влияет не чистая фонотактика, а сопряженная с основой характеристика (тип основы, по Е. А. Крейновичу),

которой частично может быть нейтрализован фонотактически. Кроме того, постулирование суффикса из двух согласных с последующими разбивками представляется натянутым и излишним.

Не очень ясна аргументация И. А. Николаевой в споре о том, как записывать основы типа *mōj-m* 'держит', *qojl* 'бог', *law-m* 'пьет'. Языковое чутье информантов прекрасно решило эту проблему; здесь в ЮРС Курилова используются буквы -*й*- и -*в*-, причем видно нормальное различие — *лавм* (inch. *лав-аам*) противопоставлено *лавур* 'плавун, деревья на воде', а *мойм* противопоставлено *айим* 'стреляет'. Предложение автора о закрывающихся дифтонгах излишне; если же это сочетание двух гласных, то зачем тогда стоило избавляться от сочетаний двух гласных при обозначении [*ê*] (-ие-), [*ô*] (-уо-) (- -*е:*- и -*о:*- у И. А. Николаевой)? Особая фонотактика глайдов и других сонантов — вполне обычное явление. Постулирование дифтонгов с первым любым гласным (и даже «автоматическим (которого нет?)» -*ə*- шва, ср. примеры выше) совсем не обязательно.

В подразделе, посвященном консонантизму, все описано достаточно хорошо и понятно, неясен только выбор автором знака -*d'*- для обозначения палатальной звонкой аффрикаты — пары для -*č*-. В юкагирских языках это ярко выраженный палатальный согласный, активный орган — средняя часть языка. Если еще можно понять использование -*l'*-, -*n'*- для обозначения палатальных сонантов (ведь они, как правило, связаны со старым задним сингармонистическим рядом или ассимиляцией по -*j*- или палатальному), то звонкая палатальная аффриката встречается независимо (!) от ряда, как и -*č*-, и ее следовало бы обозначить через -*ǯ*-, принимая нотацию словаря. Наверное, и «гачек» здесь не до конца удобен. В современном тундренном юкагирском встречается мягкая аффриката -*ʒ́*-, образуемая кончиком языка, она противопоставлена палатальной аффрикате -*ʒ*-. Это вариант фонемы -*d*- перед -*i*-, что очень хорошо видно при словоизменении; орфографически он записывается как -*ди*-. Вот для такого случая подошел бы знак -*d'*-.

В главке «Нефонематические вариации» говорится о процессе спирантизации палатальной аффрикаты и ее переходе *č > ś*, который встречается в конечном положении и интервокале (так у автора). Таким образом, -*ś*(-) заменяется на -*č*- в орфографии словаря. Ср. случаи: колым. *pêdə-ś* perf./ sing. от *pêdə-j* 'гореть' ~ тундр. *pêdə-j*; колым. *poʒo-ɣo-ś* sing. от *poʒo-ɣo-j* 'сверкать, блестеть' ~ тундр. *poʒaɣa-c* (*poʒaɣajl*) sing.; колым. *puɡə-ś* (*puɡəj-*) 'горячий, теплеть' ~ тундр. *puɡə-c* (*puɡəl!*); колым. *porcaɣa-ś* sing.

'плеснуть, брызнуть' ~ тундр. *porcaγa-c* (*porcaγajl*) sing.; колым. *kebə-ś* (*kebəj-k* imper.) sing. 'уехать' ~ тундр. *kewə-c* (*kewəjl*) sing. 'уйти'; колым. *šašaγa-ś* sing. 'лопнуть, разорваться' ~ тундр. *sisaγa-c* (*sisaγajl*) sing.; колым. *uj-ś* (*uj-*) 'родиться' ~ тундр. *ô-ńəj* (*ôńəl*); колым. (*puddəgə*) *jêś* 'напасть' ~ тундр. *jew-gə-c* (*jewgəjl*) 'стукнуться'; колым. *joŋži-ś*, *uŋži-ś* perf./ sing. 'спать, уснуть' ~ тундр. *jandi-c* (*jandūl, janrūl*), (Кр.) *janri-c, jenri-ś*; колым. *jō-ś* 'болеть (о части тела)' ~ тундр. *jô-c* (*jawl*) 'заболеть болезнью'; колым. *cejli-ś* 'далекий' ~ тундр. *cejli-c* (*cejlul*); колым. *ceni-ś* 'интересный, забавный' ~ тундр. *ceńi-c* (*ceńūl*); колым. *pažili-ś* 'зудеть, чесаться, быть щекотным' ~ тундр. *parali-c* (*paralūl*) и др.

Надо заметить, что конечный -*č* (-*ś*) встречается лишь в глагольной парадигме и происходит из развития морфонологических сочетаний "-*j-j* и "-*v-j* (3sg. intr.), так что это не исконная аффриката. Неконечные случаи такого «ослабления артикуляции» отмечены в следующих основах: колым. *ńāśə* 'лицо' ~ тундр. *ńācə-ŋ*; колым. *eśê* 'отец'; колым. *qonro-śi* dur. 'дырявиться'; колым. *uldi-śiś* 'наполниться'; колым. *keśī-m* 'принести, привезти; привести', при *kêś*, (Кр.) *kośśi* 'прийти, приехать, явиться' ~ тундр. *keci-m* (*kecil*); колым. *lośil* 'огонь; очаг; дрова' ~ тундр. *lacil* 'огонь; дрова'; колым. *aŋśi-m* 'искать' ~ тундр. *waŋci-m*; колым. *śew* 'сугроб' ~ тундр. *cawə-ŋ*; колым. *ńańiś* (Кр.), *ńanʒə* (*pul*) 'грех' ~ тундр. *ńańic*; колым. *joŋśo* 'колокольчик, звонок' ~ тундр. *jöŋcə-ŋ*; колым. *u-śiś* 'шевельнуться'; колым. *jūśə* 'дыхание; дым'; колым. *ă-sī-m* 'подтянуть (подпругу), натянуть тетиву' ~ тундр. *wāj-ci-m* 'вытащить'; колым. *caśā* 'старший брат'; колым. *iśi* 'мужской член'; колым. (*pukəĺə*) *jōśo* 'снежинка' ~ тундр. *jossə-ŋ* 'лед (вечной мерзлоты)'; колым. *ekśiĺ* 'лодка (ветка)'; колым. *muśin* 'различный, всякий'; колым. *ńāśə-ńi* 'острый' ~ тундр. *ńācə-ńi*; колым. *siśkə* 'длинный (об одежде)' ~ тундр. *sis-nəj*; колым. *kuśê* 'комар, мошка́' ~ тундр. *kicê-ŋ*; колым. *ńāśə-də-ś* 'вернуться'; колым. *āśə* 'олень (домашний)'; колым. *ĭśə* 'острие' ~ тундр. *öjcə, ewcə-ŋ* 'вершина (горы)'; колым. *iśaγa-ś* 'упасть вниз головой'; колым. *lulśi* 'выть'; колым. *īśə* 'осетр'; колым. *ĭśə-d amin* 'локоть', *jā-d ĭśə-ńi* 'треугольный'; колым. *ńāśə* 'мыс (скалистый)'; колым. *iśkom* 'постоянно, все время'; колым. *eməśkə* 'вдруг, неожиданно'; колым. *siśkədê* 'конек (рыба)'; колым. *eśkərī-m* 'нападать'. Это, возможно, неполный, но показательный список с -*ś*- в колым. Однако уже на его материале видно, что распределение И. А. Николаевой не работает. Любопытно отметить, что несколько слов со срединным -*ś*- имеет в Сибири внешние параллели с -*j*- (ср. развитие форманта глаголов): колым. *ńāśə* 'мыс (ска-

листый)' ~ нивх. *ŋāju* 'ручей, овраг, балка, лощина', камчук. *ŋèj-ŋej* (~ -*ŋ-*) 'гора'; колым. *śew* 'сугроб' ~ нивх. *ʒïw* 'ледяной наст', камчук. *jiɣʷə̃j* 'шуга'; возможно, колым. *ńāśə-ńi* 'острый' ~ нивх. *ŋoj* 'сук, penis' (как англ. *prick*). Представляется правильным все же не убирать колымское различение палатальной аффрикаты.

Далее идет раздел о позиционных распределениях согласных. В нем даны две неудобочитаемые таблицы с плюсами и минусами и словесными обозначениями типов согласных, которые нагляднее представить так:

Комбинации на морфемной границе:

	T	D	R	j	#
T	TT	–	TR	–	T
D	–	–	–	–	–
R	RT	RD	RR	–	R
j	jT	jD	jR	–	j
#	T		R	j	–

Комбинации внутри основы:

	T	D	r	L	j
T	TT	–	–	–	–
D	–	–	–	–	–
r	rT	rD	–	rL	–
L	–	LD	Lr	LL	–
j	jT	jD	jr	jL	

Здесь через T обозначен глухой шумный, D — звонкий, R — любой сонант, L — любой сонант кроме *r*. Сразу видно, что таблица для сочетаемости внутри основ упрощена. Сочетания типа TL есть, кроме того, они имеют внешние схождения: колым. *ńaw-niklê* 'песец' (= «белый ~»), *ńetlə(ŋ)* 'лиса' ~ нивх. *ŋajq* 'щенок'; тундр. *ńa-tləbê* 'куропатка' (где первая часть соответствует омок. *ҏавео* [ӄабео < "ҏабео Матюшкин] 'куропатка' ~ нивх. *ńəvar* 'ласточка'), а вторая ~ тундр. *labuntə-ŋ* ~ нивх. *təliwə* 'куропатка', ител. ***tьlàwi-tьla(xʷ)*" 'кроншнеп, куличок'; колым. *ikĺ-oj* (caus. *ikĺə-žəm*), тундр. *ikĺāń* (caus. *ikĺə-rum*) 'твердый' ~ нивх. **hikila-* 'отличный, замечательный; сильно'; колым. *saqlə(ŋ)* 'полярная сова' ~ нивх. *orqəz-ak* (~ **orqəř̃-*) 'белоголовая синица' (значение 'синица' возникает под влиянием нивх. **əʒaq* 'синица', с которым это слово контаминирует), камчук. **'ə̣tqəl* 'сова' > прачук. **ətqəł* 'по-

лярная сова; торбей (вид птицы)', ител. *'*ə̀si̯ə̀l-i̯-* 'сова'; тундр. **moqlə-*'ставить (копну); бугриться (о сухожилии), (с)комкаться (о бинте)' ~ нивх. *ŋaqziv (~ -*r̃-*) 'сумочка (охотничья, для мелких вещей)', камчук. **ŋ̀aqlå* 'ворох, связка, постель' > прачук. *ŋaqlab, ител. **ŋak̀ʷə̀kå-čh̯i*; колым. *ńutnija*, тундр. *nutnə-ŋ* 'пуп', колым. *ńutńə-j* 'торчать' ~ камчук. *'*nətnə* 'рог, зуб' (*вырост); колым. *tuknə* 'гвоздь' ~ нивх. *däkəń 'ноготь, коготь', камчук. **tkə̀ŋěv-* 'разрывать, раздирать, царапать'; колым. *šoqno pögi* 'галоп (бежать галопом)' ~ прачук. **cə̀ŋqə-*, **cə̀ŋqə-* (~ -*k-*) 'прыгать, подпрыгивать; прыгать, делая большие прыжки'; ст.-колым. *харńä* [chapnjä Миллер/ Линденау] < *qapńə 'собака' ~ камчук. **qepe'ŋ̀ə* > прачук. **qeper(e)* 'росомаха'; колым. *pukŋ-ōj* 'густой (о шерсти)' ~ нивх. сложение *puγ-baqəŋa 'еж', где первая часть значит 'игла', т. е. «игольчатая шерсть».

Далее идет подробное рассмотрение юкагирского сингармонизма, само по себе интересное и хорошо написанное. В подразделе про удлинение (шва и переход его в долгие гласные «-*a:*-/ -*e:*-») представляется более уместным в случае глаголов трактовать появление -*ā-*/ -*ê-* не фонетическим законами, а морфологически — это показатель начинательности действия (ср. в ту же копилку колым. *eri-ś* (*erū-*) 'плохой, вредный' и *er-ê-ĺəl* 'ненавидеть' (= «так, что поплохело»). Случаи с именем не так однозначны, как это представляет И. А. Николаева, ср. тундр. *ködə* 'человек', dim. *ködə-dê*, *kön-dê* (Крейнович ЮЯ, с. 25) — выпадение шва даже с регулярным морфонологическим чередованием согласного(!), или *ńantə-p-tê-k* 'тальники' с dim. от pl. без всякого удлинения (там же), тунд. *cirəmə* 'птица', dim. *cirəmə-dê* (ЮРС 552) или колым. *toukā-dê* dim. 'собачка' от *touko* (*pul*) с полным гласным, *jelgölə-dê* dim. 'годовалый олененок', *cincə* и dim. *cincə-dê* 'мышца', *siśkə-dê* dim. (с -*ə-* перед dim.). Получается, что вопрос об образовании диминутивов требует дальнейшего исследования, и не исключено, что у аффикса есть два алломорфа (-*dê* и -*ādê* / -*êdê*) с пока неясным распределением, но уж точно нет удлинения шва и перехода его в долгий гласный полного образования.

В подразделе о чередованиях (звонких) шумных и (носовых) сонантов указывается четкое правило их дистрибуции («Obstruents occur before a vowel, while sonorants occur before a consonant or a pause» (с. 47), иными словами носовой гоморганный сонант на месте звонкого (смычного) в конце закрытого слога). Приводятся ряды *d* ~ *n*, *d'* ~ *n'* [*ʒ* ~ *ń*], *b* ~ *m*, *g/γ* ~ *ŋ* и колым. *ž*, тундр. *r* ~ *n*. Однако не со всеми этими рядами можно согласиться. Ряд *b* ~ *m* в юкагирских языках встречается только перед суф-

фиксом, начинающимся на носовой, что демонстрируют приведенные примеры тундр. *mub-egə-* 'to shorten' ~ *mum-nə-* 'short' (имеется в виду *mub-əgə-c* (*mubəgəjl*) 'оборваться, стать обрубком' и *mum-nəj* 'неполный, отрезанный; комолый' при колым. *mum-nəj* 'комолый, ущербный', причем не исключено, что прилагательное связано с тундр. *mimiĺ* 'панты; конец рога у оленей (мягкий)' и колым. *mimil* 'панты'), *sab-aγə-* 'to stretch' (широко развернуться, распрямиться) ~ *sam-nə* 'flat' (широкий и плоский).

Пример же тундр. *jaba:-* 'to die' ~ *jam-d'i:-* 'to be ill' некорректен. В юкагирском есть глагол 'заболеть болезнью 1, заболеть (о чем-л.) 2, рана 3, болезнь 4, боль 5, болеть, хворать 6': тундр. *jô-c* (*jawl*) 1, *jawl* 5, *jōj-m* (*jōjl*) dur. 6, (Крейн.) *jo(j)-*, *jaw-* 1, *jô-ʒə-ri* (*jôʒərul*) 6, *jô-qaj* 2, *jôĺə*, *joĺə* 3, *jô-ʒəŋ* 3, тундр. *jōś* (о части тела), *jō-ĺoj* (*jōĺo-ti*) [*jōĺ-ā-j* inch.] 1, *jow-ńəj* 'больной', *jóĺo* (*pul*) 4, *jō* 4, *jow* 5, *a-jō* 'ой, больно!', (Крейн.) *jô-* 1, *joj-qaj* 2, *jeju*, *jō-ʒə* 4, ст.-колым. *jòa-č* [yòatsh Bil], *joa-č* [joač Me], *ju-č* [юч Кл] 4, *jöju-leń* [єюлень Кл] 3, 4, чуван. *jú-lę* [ю́лɛ̀ Бен] 4. В нем представлен старый согласный *-*w-*. (Ср. нивх. **ʒovo-c-* 'бедствовать, ослабеть; бедный').

Кроме того, есть тундр. *jamʒi-j* (*jamʒil*) dur., (Крейн.) *jamʒi* (*jamʒil*) 'болеть, хворать; болезнь, хворь', который не совпадает с dur. от первого глагола *jōj-m* или *jôʒə-ri* (*jôʒə-rul*).

Наконец, есть глагол 'умереть 1, убить 2': тундр. *jaba-j* (*jabal*), (Крейн.) *jabə-j*, *jabaj* (*jabəl*) 1, ст.-тундр. *jõbòn* [jömbòn MLU] 1, колым. *jowloγī̇-m* 2, ст.-колым. *jôboj* [jómboi MLK] 1, чуван. -*jŭbotɨ* [-ю꙲боты Бен], *jewoj* [ебой МЧ] 1. Опираясь на колым. форму с -*w-* (*jowloγī̇-m* 'убить'), можно реконструировать *-*v-*, однако памятники указывают на простой смычный *-*b-*, а -*wl-* может быть из < "-*bl-*. Для тундр. глагола *jamʒi-* может быть лишь с натяжкой привлечен колым. глагол 'умереть 1, смерть 2, мертвый 3' *amdəj* 1, *amdəl* 2, *amdəjə* 3, (Крейн.) *amdəj*, *amdə-tum* fut.? 1, *amdə* 2, ст.-колым. *amda* [Me], *amdań*, -*amtań* [амдань, -амтань Кл] 1, 3 с нерегулярным соответствием аффрикаты дентальному (из-за контаминации с глагольными аффиксами), или ср. внешнюю параллель: нивх. *huńʒ 'болезнь (обычно глаз)'. Это — единственный случай такого типа чередования не перед носовым, так что можно довольно уверенно заявить, что его нет.

Нет также и примера на последнее соответствие. Дается колым. *ni:žə* 'to squeal, to squall' ~ *nen-γižə-* 'to growl, to snarl' и тундр. *nira-γə-* 'to grin' и *nin-γarə-* 'to growl'. Здесь смешаны в одну этимологию разные глаголы. С одной стороны— колым. *niŋžə-j* (*niŋžə-ti*) 'визжать', колым. (Крейн.) *nižəj* 'визжать',

nižəl 'визг' — записи разных авторов одного ономатопоэтического глагола с наличием или отсутствием *-ŋ-*. С другой стороны — тунд. глаголы *nira-ɣa-j* (*niraɣal*) intens. [*nira-ʒi-j* mult.] 'осклабиться (поднять верхнюю губу, показывая зубы)'; и соответствие колым. *nengi-ža-j* (*nengiža-ti*) (так в СЮР!) 'ворчать (о собаке)' ~ тундр. 'рявкнуть (о собаке) 1, скалить зубы 2, рычать 3' *ninɣa-c* (*ninɣajl*) sing. 1, *ninɣa-rəj* (*ninɣarəl*) [*ninɣanʒi-j* (*ninɣanʒil*) mult.] 3, (Крейн.) *ninɣa-rəj* (*ninɣa-rəl*) [*ninɣa-rəc* sing.] 2. То есть и такого соответствия нет. Вообще, если бы оно было реально, мы бы имели чередование колым. *d*, тунд. *r ~ n*, т. к. *ž/r*-соответствие является частным случаем *d/r* соответствия — палатализация перед старым **i*. Но ср. случай с **-δ-* : 'крепко 1, крепкий, прочный 2, стать крепким 3': тундр. *war-i* (*warul*) [*wari-tə-m* caus.] 2, *warinəŋ* 1, *warul-muj* inch., *warul-vəj* 3, (Крейн.) *wāruj* (*warul*) 2, *warʒə-* (!) 1, *wāra-təjm* 3, колым. *adi* 1, 2, *adul-bəj* (*adulbati*) [*adulə-dum* caus.] 3, (Крейн.) *adi* [*el-at* neg. (!!!)] 1, *adul-bəj* [*adulə-dum* caus.] 3, ст.-колым. *aδδi* [addi Bil], *aδδў* [addў Me], *a*[δ]*ii* [азыи < "адыи Кл] 1, 2. Здесь в конечнослоговых позициях нет и следа чередований. Также в большом количестве имен представлен ауслаут *-r* в тундр. и *-t* в колым.

Таким образом, остаются только чередования исторических звонких смычных (и аффрикаты) с конечнослоговым гомоорганным носовым (исключая, по-видимому, *-b-*).

Подразделы про ассимиляцию звонких по глухости, палатальную ассимиляцию, эпентезу в целом не вызывают серьезных нареканий. Надо отметить, что чередование начального *s-* (*š-* тундр.) основы с *-r-*, отмечаемое Е. А. Крейновичем, и трактуемое им как переход в позиции после гласного или сонанта (ЮЯ 17), здесь описывается более ограниченным правилом и трактуется как чередование после *-n-*, *-ŋ-*, носовые же предполагаются в части случаев выпавшими. Этому противоречат тундр. примеры *samqa-rāl* 'стол' = «плоская доска», *amacə-rukun* 'хорошая вещь', *Íuku-rukun* 'мелкое дело' (ЮЯ 31), где не предполагается никакого носового (первая основа — адъективная). Кроме того, есть примеры *cō-n-duskə* 'чашка', *cô-n-dawɣa* 'сковородка' (*sawɣa* 'корыто') и *cô-raskə* 'котел, кастрюля', *tenmə-ruskə* 'гортань, пищевод' (при *tenmə(ŋ)* 'горло'), *suska-rāl sāl* 'жердь над очагом, крюк над очагом' при *sāl* 'дерево, древесина' (ЮРС), а также предложения типа *ködə rusəj* (изолированно *susəj*) 'человек бросил', приведенного автором. Для тундр. языка это живое чередование.

Хотелось бы специально отметить, что и И. А. Николаевой, и Е. С. Масловой выделяются

распределенные по палатальному окружению «вставные» *u/i*. Но морфонологически это и есть особый беглый гласный **ŭ*, который отличается от кратких *-u-*, *-i-*, присутствующих в непервых слогах, и неподверженных этому распределению. Соответственно постулирование еще и беглого *-ə-* приводит к очень невероятным конструкциям. Значительно проще предполагать наличие алломорфов Тŭ / Тə для части аффиксов, распределенных по историческому типу основы (причем *-ə-* в глагольной словоизменительной парадигме не выпадает). Коротко дано развитие сочетаний согласных с (морфонологическим) *-j-*, но здесь стоило бы дать разобранную систему спряжения, например, в тундренном глаголе и показать чередования.

В этом языке есть два типа окончаний, оформляющих непереходные и переходные основы. Ниже рассмотрены системы аориста или настояще-прошедшего времени (3pl. не входит в парадигму и образуется по другой модели, но формы основы в этой форме также показательны, равно как и основы 3-го лица в отрицательном аспекте, где они отличны от окончаний положительного аспекта в непереходном спряжении).

Непереходный глагол

I тип: основы на *"-A-* — полный гласный (как краткий, так и морфонологически долгий):

1sg.	*A-jəŋ*	*"A-jəŋ*	1pl.	*A-jli*	*"A-jli*
2sg.	*A-jək*	*"A-jək*	2pl.	*A-jmut*	*"A-jmut*
3sg.	*A-j*	*"A-j*	3pl.	*A-ŋi*	*"A-ŋǐ-j*

neg.

3sg.	*A*	*"A-#*	3pl.	*A-ŋu*	*"A-ŋǔ*

Этот тип имеет довольно большое количество примеров и доминирует среди типов основ.

II тип: основы на гласный *"-ŭ-* (см. выше о гласном):

1sg.	*i-jəŋ*	*"ŭ-jəŋ*	1pl.	*i-jəli*	*"ŭ-jəli*
2sg.	*i-jək*	*"ŭ-jək*	2pl.	*i-jəmut*	*"ŭ-jəmut*
3sg.	*i*	*"ŭ-j*	3pl.	*u-ŋi*	*"ŭ-ŋǐ-j*

neg.

3sg.	*0*	*"ŭ-#*	3pl.	*0-ŋu*	*"ŭ-ŋǔ*

Этот тип немногочислен, включает в себя односложные основы, например, *pan-i* 3sg. (*panul* inf., *əl pan* 3sg. neg., *əl pan-ŋi* 3pl. neg.) 'стать, стоять', *war-i* 3sg. (*warul* inf.) 'быть крепким', *ńir-i* 3sg. (*ńirul* inf., *əl ńir* 3sg. neg., *əl ńir-ŋi* 3pl. neg.) 'рвать' и нек. др.

III тип: основы на согласный ″-*j*-:

1sg.	*A-cəŋ*	″*Aj-jəŋ*	1pl.	*A-cəli*	″*Aj-jəli*
2sg.	A-cək	″*Aj-jək*	2pl.	*A-cəmut*	″*Aj-jəmut*
3sg.	*A-c*	″*Aj-j*	3pl.	*Aj-ŋi*	″*Aj-ŋĭ-j*

neg.

3sg.	*A-j*	″*Aj-#*	3pl.	*Aj-ŋu*	″*Aj-ŋĭ*

Очень большая группа производных основ; есть также и непроизводные односложные основы, inf. имеет -*jl*.

IV тип: основы на ″-*ŭv*-:

1sg.	*i-cəŋ*	″*ŭv-jəŋ*	1pl.	*i-cəli*	″*ŭv-jəli*
2sg.	*i-cək*	″*ŭv-jək*	2pl.	*i-cəmut*	″*ŭv-jəmut*
3sg.	*i-c*	″*ŭv-j*	3pl.	*ŭ-ŋi*	″*ŭv-ŋĭ-j*

neg.

3sg.	*u*	″*ŭv-#*	3pl.	*ŭ-ŋu*	″*ŭv-ŋĭ*

Довольно большая группа производных основ.

V тип: основы на ″-*Al*-:

1sg.	*A-зəŋ*	″*Al-jəŋ*	1pl.	*A-зəli*	″*Al-jəli*
2sg.	A-зək	″*Al-jək*	2pl.	*A-зəmut*	″*Al-jəmut*
3sg.	*A-ń*	″*Al-j*	3pl.	*Al-ŋi*	″*Al-ŋĭ-j*

neg.

3sg.	*Al*	″*Al-#*	3pl.	*Al-ŋu*	″*Al-ŋĭ*

Большая группа производных качественных основ, где согласный стоит после гласного второго слога, инф. на -*Aləl*.

VI тип: основа на ″-*An*-:

1sg.	*A-зəŋ*	″*An-jəŋ*	1pl.	*A-зəli*	″*An-jəli*
2sg.	*A-зək*	″*An-jək*	2pl.	*A-зəmut*	″*An-jəmut*
3sg.	*Ańi, -n-(n)i*	″*An-(ŭ)j*	3pl.	*An-ŋi*	″*An-ŋĭ-j*

neg.

3sg.	*An*	″*An-#*	3pl.	*An-ŋu*	″*An-ŋĭ*

Основа глагола *mon-* (*monul* inf.) 'сказать'. В 3sg. ожидается форма ″*moń*, которая подвергается вторичным наращениям или выравниванию. В словаре Курилова (ЮРС 253) этот глагол вообще спрягается по спряжению на ″-*й* (*moni* 3sg.).

VII тип: основа на согласный ″-з-:

1sg.	*An-зəŋ*	″*Aз-jəŋ*	1pl.	*An-зəli*	″*Aз-jəli*
2sg.	*An-зək*	″*Aз-jək*	2pl.	*An-зəmut*	″*Aз-jəmut*
3sg.	*Aзi*	″*Aз-ŭj*	3pl.	*Ań-ŋi*	″*Aз-ŋĭ-j*

neg.

3sg.	*əl-eń*	″*Ań-#*	3pl.	?

Основа глагола *eзi* 3sg. (*eзil* inf.) 'жить'. Сейчас в этом значении по большей части употребляется другой глагол: *saɣanə-* 'сидеть, жить'. Отрицательная форма 3sg. сохранилась в виде *əl-eń* 'частица отрицания, употребляется только в этой форме и означает категорическое *нет, не*' (ЮРС 596).

Переходный глагол

Ниже даются типы переходных глагольных основ. Они имеют особое спряжение в положительном аспекте, а в отрицательном имеют окончания непереходных глаголов.

I тип: основы на ″-*A*- — полный гласный (как краткий, так и морфонологически долгий):

1sg.	*A-ŋ*	″*A-ŋ*	1pl.	*A-j*	″*A-j*
2sg.	*A-mək*	″*A-mək*	2pl.	*A-mk*	″*A-mk*
3sg.	*A-m*	″*A-m*	3pl.	*A-ŋa*	″*A-ŋa*

neg.

1sg.	*A-jəŋ*	″*A-jəŋ*	1pl.	*A-jəli*	″*A-jəli*
2sg.	*A-jək*	″*A-jək*	2pl.	*A-jəmut*	″*A-jəmut*
3sg.	*A*	″*A-#*	3pl.	*A-ŋu*	″*A-ŋĭ*

Тип с многочисленными примерами как производных многосложных, так и непроизводных односложных основ. В отрицательном аспекте полная аналогия непереходным основам такого

же типа, только в 1pl., 2pl. есть инновативный гласный -ə-.

II тип: основы на гласный ″-ŭ- (см. выше):

1sg.	-uŋ	″-ŭ-ŋ	1pl.	-uj	″-ŭ-j
2sg.	-mək	″-ŭ-mək	2pl.	-umk	″-ŭ-mk
3sg.	-um	″-ŭ-m	3pl.	-ŋa	″-ŭ-ŋa

neg.

1sg.	l-ʒəŋ	″lŭ-jəŋ	1pl.	l-ʒəli	″lŭ-jəli
2sg.	l-ʒək	″lŭ-jək	2pl.	l-ʒəmut	″lŭ-jəmut
3sg.	l	″lŭ-#	3pl.	l-ŋu	″lŭ-ŋŭ

Этот тип немногочислен, включает в себя односложные основы, например, *mol-um* 3sg. (*mol-ul* inf.) 'ночевать, провести время' и некоторые другие. Как можно заметить, ″-ŭ- выпадает перед суффиксом, имеющим собственный гласный. В отрицательном аспекте окончания как у I типа, однако нет ассимиляции йотом предыдущего плавного (!), что указывает на недавно выпавший гласный между этими согласными.

III тип: основы на согласный ″-j-:

1sg.	Aj-ŋ	″Aj-ŋ	1pl.	A₁j-c, A-c	″Aj-j
2sg.	Aj-mək	″Aj-mək	2pl.	A₁j-mk, A-mk	″Aj-ŭmk
3sg.	Aj-m	″Aj-m	3pl.	Aj-ŋa	″Aj-ŋa

neg.

1sg.	A-cəŋ	″Aj-jəŋ	1pl.	A-cəli	″Aj-jəli
2sg.	A-cək	″Aj-jək	2pl.	A-cəmut	″Aj-jəmut
3sg.	Aj	″Aj-#	3pl.	Aj-ŋu	″Aj-ŋŭ

Очень большая группа производных основ, есть также и непроизводные односложные основы (например, *paj-m* 3sg., *paj-c* 1pl. 'ударить', *wāj-m* 3sg., *wāj-c* 3pl. 'держать', *susəj-* 'бросать, сорить'), inf. имеет -jl. Непроизводные основы частично контаминируют со спряжением на ″-A-.

IV тип: основы на согласный ″-v-:

1sg.	Au-ŋ	″Av-ŭŋ	1pl.	Au-j	″Av-ŭj
2sg.	Au-mək	″Av-ŭmək	2pl.	Au-mk	″Av-ŭmk
3sg.	Au-m	″Av-ŭm	3pl.	Au-ŋa	″Av-ŭŋa

Одиночные непроизводные основы *lew-* 'съесть', *law-* 'выпить' и некоторые другие, inf. имеет -Aul (в словаре Курилова эти основы и их формы записываются как -Aвм, -Aвл). Отрицательные формы как в спряжении на ″-A-.

V типа основ на латеральный нет.

VI тип: основа на ″-Ań-:

1sg.	Ań-iŋ	″Ań-ŭŋ	1pl.	Ań-ij	″Ań-ŭj
2sg.	Ań-mək	″Ań-mək	2pl.	Ań-imk	″Ań-ŭmk
3sg.	Ań-im	″Ań-ŭm	3pl.	Ań-ŋa	″Ań-ŋa

neg.

1sg.	Ań-ʒəŋ	″Ań-jəŋ	1pl.	Ań-ʒəli	″Ań-jəli
2sg.	Ań-ʒək	″Ań-jək	2pl.	Ań-ʒəmut	″Ań-jəmut
3sg.	Ań	″Ań-#	3pl.	Ań-ŋu	″Ań-ŋŭ

Основа глагола *puń-im* (*puńil* inf.) 'забить, добыть'. Здесь хорошо заметно, что исторический ″-ń- не ассимилируется полностью последующим йотом, в отличие от ″-n- (см. VI тип непереходного спряжения).

VII тип: основа на согласный ″-ʒ-:

1sg.	A-ʒəŋ	″Aʒ-ŭŋ	1pl.	Aʒ-ij	″Aʒ-ŭj
2sg.	Ań-mək	″Aʒ-mək	2pl.	Aʒ-imk	″Aʒ-ŭmk
3sg.	Aʒ-im	″Aʒ-ŭm	3pl.	Ań-ŋa	″Aʒ-ŋa

neg.

1sg.	Ań-ʒəŋ	″Aʒ-jəŋ	1pl.	Ań-ʒəli	″Aʒ-jəli
2sg.	Ań-ʒək	″Aʒ-jək	2pl.	Ań-ʒəmut	″Aʒ-jəmut
3sg.	Ań	″Aʒ-#	3pl.	Ań-ŋu	″Aʒ-ŋŭ

Основа переходного глагола *teʒ-im* (*teʒil* inf.) 'взять, брать'. В словаре Курилова (ЮРС 275) и в работе И. А. Николаевой он дается с заглавной формой в виде *teń-*, но в этой основе -ń отмечается только в конце слога. Здесь видно, что в отрицательном аспекте на поверхностном уровне окончания совпадают с -ń- спряжением, но в силу морфонологических причин ″-ʒ- регулярно переходило в -ń в конце слога.

Рассматривая эти парадигмы, можно прийти к следующим выводам:

— твердые сонанты ″-l-, ″-n- полностью ассимилируются последующим ″-j- (> -ʒ-) на морфемном шве;

— палатальные согласные *"-ʒ-*, *"-ń-* не теряются перед последующим *"-j-* (> -ʒ-), а дают закономерный конечнослоговой *-ń* (орфографически *-ńʒ-* и *-nʒ-* равноценны);

— сочетания глайдов *"-j-* и *"-v-* с последующим *"-j-* развиваются в глухую аффрикату через ступени *"-jj-* > *"-ʒʒ-* > *"-c-* (в колым., кстати, *-ś-*, а не *-c-* на этом месте);

— старый *"-j-* после шумных реализовывался как аффриката (!).

Типы спряжений непереходного и переходного глагола в основном соответствуют друг другу: intr. I *"-A-*/ tr. I *"-A-*, intr. II *"-ŭ-*/ tr. II *"-ŭ-*, intr. III *"-j-*/ tr. III *"-j-*, intr. IV *"-ŭv-*/ tr. IV *"-v-*, intr. VI *"-n-*/ tr. VI *"-ń-*. Если учесть, что у непереходных глаголов типы V (за вторым гласным) и VII (после первого гласного основы) распределены, то можно считать, что основы intr. V *"-l-* ~ VII *-ʒ-* / tr. VII *-ʒ-* также образуют пару. Опираясь на аналогию с типами VI, здесь можно предполагать раннее состояние intr. *"-l-*/ tr. *"-ĺ-* с палатализацией латерального.

В целом в данной рецензии были рассмотрены вводная концептуальная часть словаря и описание важнейших характеристик юкагирских языков. Про проблемы юкагирской реконструкции, интерпретацию соответствий нужно говорить в другой работе, поскольку это уже является предметом научного спора. Конкретные этимологии, уточнения значений и перекомпоновку вхождений также нужно делать в другой по жанру работе — последовательно сведенном этимологическом словаре.

В любом случае, книга И. А. Николаевой представляет собой замечательный вклад в сибирскую этимологию; в ней впервые сводится современный материал тундренного и колымского юкагирских языков и материалы записей трехсотлетней давности. *Для исследователя, который захочет разобраться в языковой и лексической ситуации в Северо-Восточной Сибири, этот труд будет являться основополагающим.*

Литература

Крейнович Е. А., [Материалы] — *Исследования и материалы по юкагирскому языку*. Ленинград, «Наука» 1982.

Крейнович Е. А., [ЮЯ] — *Юкагирский язык*. Москва—Ленинград, Изд-во АН СССР 1958.

Курилов Г. Н., [ЮРС] — *Юкагирско-русский словарь*. Новосибирск, «Наука» 2001.

Николаева И.А. *Проблема уральско-юкагирских генетических связей* (диссертация на соискание ученой степени кандидата филологических наук). Москва 1988.

Николаева И. А., Шалугин В.Г., [СЮР] — *Словарь юкагирско-русский и русско-юкагирский (верхнеколымский диалект)*. СПб, «Дрофа» 2002.

Reports / Хроника

Reports / Хроника

Kirill Babaev
Russian State University for the Humanities

The World Congress of African Linguistics 6 (WOCAL–6)
Cologne, 17–21 August 2009

This year's WOCAL, marking the 15th anniversary of the Congress' first gathering in Swaziland in 1994, seems to have brought together a pretty vast collection of both scholars and topics for discussion. For five days, speakers and contributors from all over the linguistic world have fully occupied the Hörsaalgebaude of University of Cologne, where eight conference rooms were allocated for the non-stop chain of presentations in all aspects of African linguistic studies. The number of speakers has grown considerably, and it is nice to note that the share is steadily growing of those collaborators whose native language is indigenously African.

The opening keynote presentation of the first plenary session was made by Neville Alexander (Cape Town University, S. Africa) who spoke on the necessity of bringing the African languages more into the political and economic life in the independent states of the continent. The speaker emphasised the need for paying more attention to the issues of applied linguistics which can lead to establishing a higher social status of local languages in Africa and, therefore, promote their survival and development.

In the next days, plenary presentations were devoted to sign languages in Africa, African languages in Latin America as remnants of earlier slave populations on the continent, and challenges of documentary linguistics in African studies. Without questioning the importance of various spheres of African linguistics, we, however, would like to focus this brief report on a number of issues in comparative and historical linguistics and language contact discussed at the Congress, as would suit the general scope of this Journal.

Gerrit Dimmendaal (University of Cologne) presented a talk on the external classification of the remote Sudanese language Tima; he described a number of its features, recorded and analyzed by the members of the joint field expedition in the past few years. Tima was historically thought to belong to the Kordofanian branch of Niger-Congo [Greenberg 1963], more exactly, to its Katla subgroup. However, it was suggested by the speaker that Tima is as distant from Kordofanian as it is from the rest of Niger-Congo, and may therefore constitute a separate branch of the macrofamily, bearing interesting cognate traits with Bantu and other subfamilies of Niger-Congo.

Another important classification issue was raised by Bruce Connell (York University / University of Kent, UK) who challenged the well-known hypothesis of a close genetic relationship between Ijo and Defaka, the two languages of the Niger Delta region, within the Ijoid family of Niger-Congo [Jenewari 1989]. Since the first claims about Ijoid were made, a number of research papers on it were published, with even an attempt at reconstructing Proto-Ijoid [Williamson, unpublished]. Data gathered in recent years makes it possible to suggest that these two are actually not members of a single node but were subject to mutual influence, and Defaka is rather a language isolate.

The report by Oumarou Boukari (Côte d'Ivoire) examined the genetic position of the nearly-extinct Pre language in his country. This isolated language is poorly studied, with only a few mentions in literature, and its affiliation still seems murky. Contrary to previous scholars, however, the speaker attempts to show that Pre could be a remnant of Gur or Kru population of the area heavily influenced by Mande languages. To support this, a number of morphological and syntactic features were attracted, as well as lexical data.

The classification of Ometo, one of the groups of the Omotic language family, is revised by Hirut Woldemariam (Addis Ababa University, Ethiopia), who has studied in deep the speech communities previously treated as speakers of the Gamo language. The conclusion of the presentation, however, is unexpected: there is no Gamo language. Its varieties are in fact heterogeneous, and their speakers do not identify themselves as Gamo. So, a new subgroup of Ometo is called for, with some other members of Ometo also to

be placed under the same "umbrella" or into this new subbranch, based on their language features.

Roger Blench (Kay Williamson Educational Foundation, Cambridge, UK) presented yet another reclassification of the Bantoid family, this time tentatively including the Bendi subgroup, traditionally referred to as Cross River. Though the presentation was mostly devoted to severe criticism of all previous classifications, including those proposed by Blench himself, some appealing insights were also evident. One of the ideas that the speaker expressed overtly, and which is commonly discussed, is the need to create a solid and reliable genetic classification of Bantu to replace Guthrie's previous one [Guthrie 1967–1971], and to come out with a stricter separation of the terms 'Proto-Bantu' and 'Common Bantu' that are sill used chaotically.

Jonathan Allen Brindle (Norwegian University of Science and Technology) questioned the historical development of the numeral system in Southwestern-Grusi, a subgroup of Gur languages. By comparing data from six languages of the group, the speaker emphasises elements of three basic systems: vigesimal, decimal and quinary, all of which seem to have originated from the body part count. Still, since the lexemes for '20' are not cognate among the languages, the vigesimal system could only have spread across the area as a contact feature.

An interesting phonetic process of spirantization in Amazigh (also called Tamashek or Tamazight), a Berber language of the Sahara, was described from a typological and comparative standpoint by Yamina El Kirat (Mohammed V University, Morocco). He shed some light on the diachronic process of weakening which leads to further spirantization in the very same way as it took place in many other languages of the world, including Indo-European (Spanish, Greek), Uralic (Finnish), and Semitic (Aramaic, Hebrew). This trend seems to be supported by some concrete external conditions which generate spirantization in Amazigh.

Berber comparative studies were also the subject of another report, by Cécile Lux (University of Lyon). She focused on the aspectual system of Tetserret, a minor and undescribed language in Niger, which shows comparable similarity with that of Zénaga, another Berber tongue spoken in Mauritania. By analyzing it together with the aspect systems in neighbouring Tamashek and Tachelhit, it is possible to make conclusions on the linguistic pre-history and geographical diffusion of Berber languages.

The view on folk language studies as a powerful instrument for historical linguistics was expressed in a presentation by Gumma Ibrahim Gulfan (American University in Cairo, Egypt). The author revealed a number of common ancient traits in the folk songs of language communities belonging to the Kordofan Nubian cluster (Nilo-Saharan). Some of these traits, both morphological and lexical, observed in numerous contemporary dialects of the area, can be traced back to a common ancestor and may be used for reconstructing the proto-language of the group. Some chronological links of language features and processes can also be established by studying a number of historical facts reflected in the oral tradition.

Remote languages of another region, on the borderline between Nigeria and Cameroon, were analyzed by Roland Kießling (University of Hamburg) in his plenary presentation. He worked with the languages of the Grassfields subgroup of Bantoid, where a lot of evidence has been revealed in the recent years, shedding light on various aspects of Bantu historical morphology. The speaker pays much attention to the development of noun class marking systems in Grassfields, especially the transition from an old prefix-based system to a suffixal one, and the rise of nominal classificatory systems. These are indeed issues which contemporary Bantoid linguistics needs to study in more detail.

Another report shows how less-studied tongues of Sub-Saharan Africa can change our ideas of proto-language reconstruction. Ulrich Kleinewillinghöfer (University of Hamburg) has studied the specific *re* noun class in Longto (Adamaua family). Examining its functional scope in the language and making exciting comparative references to similar class marking in Gur, the speaker suggests that Longto has preserved an ancient noun class from Proto-North Volta-Congo, lost in other families of the stock.

Helma Pasch (University of Cologne) has built her report around the description and historical analysis of negation in Ubangian languages of Central Africa. It seems quite exciting that a link between the well known final negative particles of the Bantu sentence and Ubangian negation syntax can be established. The speaker suggested that the placement of the negative particle in the final position may be an areal feature, borrowed by Bantu languages of zones C and H. If so, this is yet another piece of evidence of close language contact between the two families which led to a considerable amount of Ubangian features in Bantu, including even personal pronouns in some languages, e. g. in Doko [Babaev 2008: 145].

Issues of convergence were further discussed in the report made by Henning Schreiber (Goethe-Universität, Frankfurt), devoted to similar sound changes taking place in neighbouring Gur and Mande languages of West Africa. It is claimed that the regular shift $\eta > h$ in Gurunsi languages (Gur) is not limited to that

group, but is also observed in Bissa (East Mande), being most probably an areal phonological feature. This example is definitely worth studying in more detail in order to understand the technology of phonetic change diffusion both in Africa and beyond.

A reclassification of Bantu itself, a subject widely discussed ever since Guthrie's referential classification, is the subject of Malin Petzell's (SOAS, London) report on the poorly explored languages of the Morogoro region in Tanzania. Speaking about common morphological traits of eight tongues of the area, the author concludes that there is evidence for regrouping the languages of zone G and, probably, beyond.

A notable comparative analysis was made by Dmitry Idiatov (University of Antwerp, Belgium) on the origins of the quotative verb *kó* in Western Mande languages. This syntactic element, more a copula than a full verb, is widespread in Mande with the function of citing indirect speech. It is typologically reasonable to suggest its origins in roots denoting speech, as in many other languages of the world (cf. English 'say'). The speaker shows that the feature can be traced back to the lexeme *gʊ́ʊ* 'sound, speech; say', reliably reconstructed for Proto-Mande.

Claude Rilly (CNRS, Paris) made a valuable contribution to the studies of the proto-language homeland for the Nubian family (part of Nilo-Saharan). Contrary to the traditional view that the cradle for the Nubian languages lay in the Middle Nile valley, the speaker presented both historical and linguistic data demonstrating his hypothesis of a Kordofan homeland for Nubians. The lexicon of these languages shows that some biological species, characteristic of the Nile valley, were not known to Proto-Nubians, and that words denoting them were only borrowed or innovated later. Moreover, there are some archaic features in Western Nubian that were lost in Nile Nubian prior to the most ancient Nubian texts of the 8th century. The author concludes that Proto-Nubian was in use in a rather restricted area of Kordofan for quite a long time before it started to disseminate. As for the traditional view on the Nubian homeland, it may have appeared due to legends of the medieval kingdom of Dongola, quite widespread among the ethnic groups of Sudan who claim themselves its descendants.

Quite an interesting observation on the typology of nominal classification in Africa was made by Viktor Vinogradov (Russian Academy of Science, Moscow) who devoted his presentation to two varieties of this morphological system, rather unusual for Africa. Though the majority of African languages either have a gender system of nominal classification or use noun classes, the system of 'classificative verbs' is used in Dogon, where

a covert class of a noun object is expressed by the form of the transitive verb of action, by modifying its stem. This structure is in use in some American Indian idioms. Another peculiar system is that of Ngyemboon, a Bantoid language of Cameroon which uses two systems simultaneously: usual noun classes and classifiers used with numerals. This latter one resembles amazingly the systems that we all know in East and South East Asian languages. Such independent development of similar morphologies on different continents should be interesting from a historical linguistics standpoint.

We would like to close this brief survey by mentioning a report by Tucker Childs (Portland State University), named "How to Pretend You Speak a Dying Language When You Don't Really Know How To". The speaker focused on methodological difficulties of recording dying and endangered languages of native communities, based on examples of fieldwork on three South Atlantic tongues of coastal Guinea and Sierra Leone. Keeping in mind all the hardships of getting through both unintentional and deliberate deception that informants may force on the researcher, those linguists who only have access to documented forms of languages should be extremely careful with the data, refraining from making long-range comparative conclusions based on just a single gloss or phonetic feature from a language that has not been properly studied. To raise fieldwork quality and, at the same time, to invoke more responsibility in making linguistic conclusions — those are the issues that were constantly discussed, with great concern, among the participants of WOCAL–6.

The Congress is now preparing a volume of Proceedings from WOCAL–6, to be ready in 2010, with a free accessible online version, hosted at the website of the Institute for African Studies (University of Cologne).

References

BABAEV, Kirill. 2008. Reconstructing Niger-Congo Personal Pronouns I: Proto-Bantoid. // *Journal of West African Linguistics*. Vol. XXXV: 1–2. Pp. 131–183. [http://www.nostratic.ru/benue-congo1.pdf]

GREENBERG, Joseph. 1963. *The Languages of Africa*. The Hague: Mouton; IJAL 29.1, Bloomington: Indiana University Press.

GUTHRIE, Malcolm. 1967–1971. *Comparative Bantu. An Introduction to the Comparative Linguistics and Prehistory of the Bantu Languages*. Vols 1–4. Farnborough: Gregg International Publishers.

JENEWARI, Charles. Ijoid // *The Niger-Congo languages* (ed. by John BENDOR-SAMUEL). Lanham-NY-London: University Press of America. Pp. 105–118.

WILLIAMSON, Kay. *Proto-Ijoid Reconstructions*. Unpublished.