Российский государственный гуманитарный университет
Russian State University for the Humanities

RSUH/RGGU BULLETIN

№ 2 (15)

Academic Journal

Series:
Philology. Journal of Language Relationship

Moscow 2017

ВЕСТНИК РГГУ

№ 2 (15)

Научный журнал

Серия
«Филология. Вопросы языкового родства»

Москва 2017

ISSN 2073-6320

УДК 800(05)
ББК 80/84я5

Вопросы языкового родства: Международный научный журнал / Рос. гос. гуманитар. ун-т; Рос. акад. наук. Ин-т языкознания; под ред. В. А. Дыбо. — М., 2017. — № 2(15). — ix + 78 с. — (Вестник РГГУ. Серия «Филология. Вопросы языкового родства»: Научный журнал).

Journal of Language Relationship: International Scientific Periodical / Russian State University for the Humanities; Russian Academy of Sciences. Institute of Linguistics; Ed. by V. A. Dybo. — Moscow, 2017. — No. 2(15). — ix + 78 p. — (RSUH/RGGU Bulletin. Series: Philology. Journal of Language Relationship: Academic Journal).

ISSN 2073-6320

http://www.jolr.ru/
gstarst@rinet.ru

Дополнительные знаки: С. Г. Болотов
Add-on symbols by S. G. Bolotov

Подписано в печать 10.06.2017. Формат 60×90/8.
Бум. офсетная.
Печать офсетная. Тираж 1050 экз.
Заказ №102

Издательский центр
Российского государственного гуманитарного университета
125993, Москва, Миусская пл., 6
www.rggu.ru
www.knigirggu.ru

Table of Contents / Содержание

Contributors

Alexander Bezlepkin — M.A. student, Institute of Eastern Cultures and Antiquity, RSUH (Moscow); a.s.bezlepkin@gmail.com

Fernando O. de Carvalho — adjunct professor, Federal University of Amapá, Amapá State, Brazil, fernaoorphao@gmail.com

Ilya Egorov — M.A. student, Institute of Eastern Cultures and Antiquity, RSUH (Moscow); i.m.jegorow@gmail.com

Luka Repanšek — assistant professor, Department of Comparative and General Linguistics, Faculty of Arts, University of Ljubljana; luka.repansek@ff.uni-lj.si

Mikhail Saenko — candidate of sciences (Philology), junior researcher, Institute of Slavic studies of the Russian Academy of Sciences, Moscow, veraetatis@yandex.ru

George Starostin — candidate of sciences (Philology), head of Department of the history and philology of the Far East, Institute of Eastern Cultures and Antiquity, RSUH (Moscow); head of the Laboratory of Oriental and Historical-Linguistic Studies, Russian Presidential Academy of National Economy and Public Administration, gstarst@rinet.ru

Mikhail Vasilyev — researcher, Institute of Slavic studies of the Russian Academy of Sciences, Moscow, mvhumanity@gmail.com

Сведения об авторах

Безлепкин, Александр Сергеевич — магистрант Института восточных культур и античности РГГУ (Москва), a.s.bezlepkin@gmail.com

Васильев, Михаил Евгеньевич — науч. сотрудник Института славяноведения РАН (Москва), mvhumanity@gmail.com

Егоров, Илья Михайлович — магистрант Института восточных культур и античности РГГУ (Москва), i.m.jegorow@gmail.com

Карвалью, Фернанду О. де — адъюнкт-профессор Федерального университета Амапы (Бразилия), fernaoorphao@gmail.com

Репаншек, Лука — доцент отделения сравнительного и общего языкознания факультета искусств Люблянского университета; luka.repansek@ff.uni-lj.si

Саенко, Михаил Николаевич — канд. филол. наук, мл. науч. сотрудник Института славяноведения РАН (Москва), veraetatis@yandex.ru

Старостин, Георгий Сергеевич — канд. филол. наук, зав. кафедрой истории и филологии дальнего востока ИВКА РГГУ, зав. Лабораторией востоковедения и сравнительно-исторического языкознания ШАГИ РАНХиГС (Москва), gstarst@rinet.ru

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 for Oriental and Classical Studies
Russian State University for the Humanities
125267 Moscow, Russia
Miusskaya Square, 6
E-mail: gstarst@rinet.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) and keywords.

For more detailed guidelines on article submission and editorial policies, please see our website at: http://www.jolr.ru or address the editorial staff directly at gstarst@rinet.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: gstarst@rinet.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) и список ключевых слов.

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

Fernando O. de Carvalho
Federal University of Amapá (UNIFAP), Amapá State, Brazil; fernaoorphao@gmail.com

On Terena (Arawakan) *-pâho* 'mouth': Etymology and Implications for Internal Classification*

This paper presents an etymological investigation of the Terena noun for 'mouth', *-pâho*, which, according to current comparative Arawakan linguistics, lacks known cognates. I show that cognates of this Terena noun exist in Mojeño, a close relative of Terena within the Bolivia-Paraná subgroup of the Arawakan family. In Ignaciano and Trinitario, the best known modern varieties of Mojeño, the cognates of *-pâho* are semantically-shifted nouns meaning 'door'. I propose an account of the semantic and formal relations between these forms via an etymological source **-paho* 'mouth' and a compound noun **paho-peti* 'door' (lit. 'mouth (of the) house'). This account relies on a more general pattern of noun formation in the Arawakan family, regular phonological correspondences and is consistent with modern views on the nature of diachronic metaphorical extensions in lexical semantic shift. The evidence presented and argumentation built to support this analysis adds Terena evidence to the correspondences supporting a vowel merger in the Ignaciano variety of Mojeño. Finally, I consider some implications for internal classification, advancing the hypothesis that Terena and Mojeño form a separate branch of the Arawakan family, the Achane branch, one that does not include Baure.

Keywords: Etymology; Arawakan languages; Terena; Semantic change.

1. Introduction

The Arawakan language family is routinely celebrated as the largest language family of the New World — both by its geographic spread and by the number of languages that belong in it (Kaufman 1990: 40; Aikhenvald 1999; Campbell 2012). In addition, linguists and other prehistorians attach significant importance to this language group as a potential source of privileged insight on the pre-history of South America (see for instance the collection of papers in Hill & Santos-Granero 2002 and Hornborg & Hill 2011). Nevertheless, our knowledge of the historical linguistics of the Arawakan language family remains arguably less advanced than is the case with other large groups of lowland South America, such as Cariban and Tupian (see e.g. Michael 2009; Campbell 2012). With this situation in mind, and seeking to complement more ambitious pioneering works such as Payne (1991), recent approaches to the historical-comparative linguistics of the Arawakan family have proceeded in a bottom-up manner, reconstructing from small sets of closely related languages, with a finer understanding of the phonology and the morphology of the relevant languages yet, at the same time, keeping an eye open to broader issues, such as that of internal classification (Michael 2011; Danielsen 2011; Lawrence 2014; Jolkesky 2016; Carvalho 2015, 2016a,b,c; forthcoming).

The present paper follows in the footsteps of this program. I will be concerned with providing a sensible etymological analysis of the form *-pâho* 'mouth' attested in Terena, a Southern Arawakan language of Brazil, within the broader background of an approach to the his-

* I am sincerely grateful to Andrey Nikulin and to Françoise Rose for enlightening comments and observations that helped improve this paper. The usual disclaimers apply.

torical linguistics of this language and the specific low-level subgroup where it presumably belongs. First, I will discuss how the historical-comparative linguistics of the Arawakan languages has so far failed to provide a sensible account for this item. After tracing this form to its earliest attested sources, I argue that it has semantically shifted cognates in the modern varieties of Mojeño, a language closely related to Terena. Evidence from early 18th century data on Old Mojeño plays a vital role in filling the gaps that relate these forms via lexical semantic shifts. In accounting for these shifts I will not only provide an account that is consistent with recent work on the directionality of semantic change — and on how reliably identified directionality trends can help in semantic reconstruction — but I will also discuss evidence for other developments in the history of Terena phonology and morphology, thus illustrating the inherent feedback between particular etymologies and a more general understanding of language's diachrony (Mailhammer 2014: 424–425).

Data for this study comes from the author's fieldwork on Terena (Cachoeirinha Reservation, Mato Grosso do Sul, Brazil), from published sources on the modern languages (Ekdahl & Butler 1969, 1979 on Terena; Ott & Ott 1983 on Mojeño Ignaciano; Gill 1957, 1970 on Mojeño Trinitario; Danielsen 2007 on Baure) and from a 18th century grammar and vocabulary of Old Mojeño (Marbán 1702). Additional sources on other languages discussed in the paper will be referred to accordingly.

The paper is organized as follows: Section **2** offers a brief overview of the Bolivia-Paraná Arawakan languages, a subgroup of the Arawakan language family including Terena, Mojeño, Baure and other less well-described languages. Section **3** reviews the treatment of the form for 'mouth' in Terena in the existing historical-comparative literature and thus sets the issues to be tackled in the remainder of the paper. Section **4** constitutes the core of the present contribution. Finally, section 5 considers some implications for internal classification of the findings discussed here, advancing the hypothesis that Mojeño is the closest relative of Terena within the family and that both, perhaps with the inclusion of Paunaka as well, form a separate subgroup, the Achane branch.

2. The Bolivia-Paraná Arawakan Languages: Brief Overview

A branch of the Arawakan language family composed of Terena (including under this label other geographically or chronologically-defined varieties such as Kinikinau and Guaná; see Carvalho 2016a), spoken in southwestern Brazil, the Mojeño varieties (such as Ignaciano and Trinitario) and Baure, all spoken in Bolivia, is usually identified in classifications of the Arawakan languages (see Matteson 1972: 186-192, who does not include Mojeño; Kaufman 1994; Payne 1991: 489; Aikhenvald 1999: 67; Walker & Ribeiro 2010: 3; Campbell 2012: 75). Other, less well-attested languages such as Paunaka and Paikoneka are also assigned to this subgroup by other researchers (Danielsen 2011; Jolkesky 2016). The label *Bolivia-Paraná* was proposed for this group by Payne (1991) and I will retain this use in the present paper.

As is the case elsewhere in the Arawakan family, the evidence for recognizing a Bolivia-Paraná subgroup has been less than compelling, however. Areal-geographic factors have had a major influence in promoting the reality of this subgroup, as implied by its label (see specially Aikhenvald 1999). Strictly linguistic evidence has been presented by Payne (1991), but this amounts to figures for shared lexical retentions and loosely defined phonological outcomes, such as a 'weakening' of Proto-Arawakan (henceforth PA) dorsal stops (Payne 1991: 440), which, even if defined in more precise terms, are known to have occurred independently in many Arawakan subgroups. Other studies, both preceding (e.g. Matteson 1972)

and following (e.g. Walker & Ribeiro 2010) Payne's work have applied similar methods, relying on shared proportions of lexical cognates to recognize this as a coherent subgroup of the family.

As to the internal organization of the Bolivia-Paraná subgroup, there seems to be a widespread, recent opinion — again, based on geography, shared lexical retentions or other assessments of 'structural similarities' — that Baure and Mojeño are closer to each other than any of these is to Terena (Walker & Ribeiro 2010: 3; Danielsen, Dunn & Muysken 2011: 185). This claim should be taken with caution, however, not only due to the different methodologies and datasets employed in these studies, but also because no investigation has so far established this by identifying shared innovations and because other authors refrain from proposing any internal structure to this branch (see e.g. Aikhenvald 1999: 67; Danielsen 2011). The latest and most detailed study of these languages, Jolkesky (2016), does not consider Terena data, taking it as a premise that Mojeño, Paunaka, Paikoneka and Baure are more closely related to each other than any of these is to Terena. This generalized but not consensual view is expressed in the arrangement below:

(1) Prevalent Internal Classification of the Bolivia-Paraná Subgroup:
Bolivia-Paraná Subgroup
Terena
Mamoré-Guaporé branch
Baure-Paikoneka
Baure
Paikoneka
Mojeño-Paunaka
Paunaka
Mojeño (Ignaciano, Trinitario)

I will assume the structure above as a kind of null hypothesis enshrining some common ground reflecting a certain degree of agreement that has been reached among researchers, even if strong, compelling evidence for it (especially for the classification of Terena) is yet to appear in print. My own opinion is that the existence of a Bolivia-Paraná subgroup is very plausible and perhaps even obvious; yet, unless this impressionistic and intuitive assessment is moved from the level of a gut feeling to a detailed, methodologically sound understanding of the historical development of these languages, hardly any progress can be achieved on this specific issue and on Arawakan historical linguistics more generally. Moreover, unless a detailed understanding of the diachrony of these languages is offered, there is little hope that the more puzzling question of the internal classification of the Bolivia-Paraná languages can be properly addressed. I turn to this issue in section **5**.

3. On Terena *-pâho* 'mouth': Treatments so far and statement of the problem

Payne (1991: 413) presents three separate cognate sets for the meaning 'mouth', each ascribed to a different Proto-Arawakan (PA) etymon. Interestingly, however, not a single member of the Bolivia-Paraná subgroup features in these cognate sets. Since Payne's (1991) study remains to this day the most extensive historical-comparative investigation of the Arawakan languages, it is perhaps surprising that this gap has not drawn any attention in the comparative literature so far. This has additional significance as the lack of a cognate in these languages —

otherwise thought to constitute a distinct subgroup in itself — might indicate the existence of lexical or morphological innovations.

Earlier, Matteson (1972: 186-192) proposed a reconstruction of 'Proto-Shani', the common ancestor of Terena, Kinikinau and Baure (she did not include any Mojeño data). In the case of her reconstructed etymon for 'mouth', Matteson (1972: 191) postulates a form **báaho* based on Terena and Kinikinau reflexes alone. Setting aside the fact that Terena and Kinikinau are close dialectal variants of the same language, so that cognate matches involving only these two speech varieties could hardly justify reconstructing an etymon at the deeper level of her Proto-Shani, there are many shortcomings in Matteson's data, particularly with respect to Kinikinau. The differentiation between Kinikinau and Terena has been mistakenly overestimated as an artifact of poor morphophonological analysis: the word-initial voiced stop in **báaho*, for instance, results from an incorrect analysis of Kinikinau as having phonemic voiced stops (see Carvalho 2016a for details on these and other points and Payne 1991: 368-371 for a general evaluation of Matteson's 1972 study). Be as that may, the form **báaho* reconstructed by Matteson does not appear in her Proto-Arawakanan cognate sets, thus underscoring the isolated status of the Terena noun *-pâho*.

Jolkesky (2016: 13) notes the existence of non-cognate material for the meaning slot 'mouth' in a comparison of Mojeño and Baure varieties: the former has a root *-haka* while the latter has *-nuki*/*-noki*. Nevertheless, based on these Baure forms and on partial cognates in Mojeño and in Paunaka compounds meaning 'beard, mustache', as in Mojeño Trinitario *-hii-nuku* (Gill 1970: 7),[1] Jolkesky (2016: 19) reconstructs an etymon **-nukɨ* 'mouth' for his Proto-Mamoré-Guaporé (PMGU), the postulated common ancestor of Mojeño, Baure, Paunaka and Paikoneka (in section **5** I note potential cognates of this etymon in other branches of the Arawakan family).

In synthesis then, nothing certain can be said about the Terena noun *-pâho* 'mouth', which seems to remain historically unaccounted for and comparatively isolated within the family. First, it does not appear in any of the cognate sets on which Payne's (1991) comparative study of the Arawakan family (and reconstruction of the Proto-Arawakan family) is based. Second, it is related only to the forms attested in Kinikinau by Matteson (1972) and by Walker & Ribeiro (2010), not a surprising or illuminating finding since Terena and Kinikinau are very close co-dialects. Finally, Terena *-pâho* 'mouth' bears no suggestive formal resemblance to the etymon reconstructed by Jolkesky (2016: 19), **-nukɨ* 'mouth', to the common ancestor of the Mojeño, Baure, Paunaka and Paikoneka. All of this could suggest that Terena *-pâho* 'mouth', despite its status as a basic vocabulary item (see e.g. Tadmor *et al*. 2010: 239) could be a loanword, perhaps, from a non-Arawakan language. As I show in the next sections, however, this form has a plausible Arawakan etymology linking it, at least, to its closest relatives, the Mojeño varieties.

4. Terena *pâho* 'mouth' and its cognates

The Terena noun *-pâho* [paːhɔ] 'mouth', is attested in virtually the same form from the earliest available documentation on the language: <*baho*> (Castelnau 1845, *apud* Martius 1867); <*bahó*> (Taunay 1868); <*pahotí*> (Schmidt 1903: 332). As stop consonant voicing is not distinctive in the language,[2] documented forms with initial <*b*> are either too phonetic or reveal other

[1] As Jolkesky (2016) does not discuss Terena, I note that the closest match to this form I could find in Terena is *-inúku* 'forehead', though the semantics in this case suggests that the similarity is merely accidental.

[2] The Terena phonological segmental inventory is as follows, symbols having the standard IPA interpretations unless noted otherwise: *p, t, k, s, ʃ, m, n, ɲ, w, j, r, h, a, e, i, o, u*. The mid vowels *e* and *o* are most frequently re-

inadequacies in the analysis. As discussed by Carvalho (2016a), some of the earliest records of Terena (Guaná) inalienable (dependent) nouns, such as body-part terms, actually present 1Psg possessive forms, a common form in elicitation, not the absolute or non-possessed forms implied by the accompanying glosses or translations.[3] 1Psg is realized in Terena by a floating nasal feature that induces stop consonant voicing, in addition to the formation of a nasal-oral contour at the left edge of the leftmost obstruent consonant. Thus, given *∅-pâho* 'his/her mouth', with ∅ coding of 3P, one has *ᵐbâho* 'my mouth' (see Eastlack 1968: 4). The circumflex diacritic <^> used in written Terena indicates a falling pitch contour and a lengthened vowel. Finally, the form <*pahotí*>, recorded by Max Schmidt, includes the suffix *-ti* which indicates a generic or non-specific possessor (e.g. *∅-hêwe* 'his/her foot', *hewêti* 'somebody's foot'; see e.g. Ekdahl & Butler 1979: 66 and next paragraph).

I have used a dash to indicate that *-pâho* like other inalienable nouns is a root. Hence, most body-part terms, many kinship terms and a few nouns denoting man-made objects or parts of wholes (e.g. 'root (of a tree)'), the semantic domains usually represented in the class of inalienable nouns, will always occur with some sort of morphological elaboration: in case they lack a prefix indicating the person-number features of a possessor (the prefix may be ∅-, as in the case of the Terena 3P),[4] these items show up either with a suffix indicating their unpossessed status or in some other construction such as a nominal compound or incorporated within a verb. Suffixes signaling the unpossessed status of inalienable nouns are known as Absolute suffixes in the Arawakanist literature. In Terena the Absolute suffix has the form *-ti*, and is commonly glossed as expressing a generic, non-specific possessor. Thus: *-pâho* 'mouth' (a root, not a free-standing word), *∅-pâho* 'his/her mouth', *ᵐbâho* 'my mouth' but *pahóti* 'someone's mouth'. All other nouns, that is, alienable nouns such as *pirítaw* 'knife' or *wáteke* 'canoe', ordinarily occur as self-standing free forms in the absence of an explicit possessor. If marked for possession, however, additional morphology in the form of suffixes indicates their 'marked', possessed status: *∅-piritáw-na* 'his/knife', *wãⁿdéke-na* 'my canoe'. These suffixes are, in turn, called Genitive or Possessive suffixes in the Arawakanist tradition (see e.g. Payne 1990: 80–83; 1991: 378; Carvalho 2015).

A comparison of the candidates for cognate status in the meaning slot 'mouth' for the Bolivia-Paraná Arawakan languages reveals an obvious (that is, semantically-matched) cognate of the Terena form *-pâho* in Old Mojeño <*nupahò*> (Marbán 1702: 160). Also attested in Old Mojeño is an apparently competing form, <*nuhacà*>, which has a clear match in the form *-haka* 'mouth' attested in both modern Mojeño varieties, Ignaciano and Trinitario (see table 1). The latter is the Mojeño form presented in Jolkesky (2016: 13), no mention being made of Old Mojeño <*nupahò*>. The Old Mojeño forms contain the 1Psg possessive prefix <*nu-*> and can

alized as open *ɛ* and *ɔ*, respectively, though I will adhere to the customary practice of using *e* and *o* in writing forms of the language. Note also that *w* is frequently realized as [v] or [ß], though phonological patterns show it patterns like a sonorant, and that *ɲ* is a marginally contrastive segment only.

[3] Glossing conventions used in this paper are as follows: REFL: Reflexive; CONT: Continuative; TH: Thematic consonant, ACT: Actual Mood; AUX: Auxiliary verb; 3P: third person; 2P: second person; 1Psg: first person singular; 1Ppl: first person plural.

[4] Though the person-number markers indicating possession in nouns or a subject in verbs are arguably prefixes (see e.g. Danielsen 2011: 508), some comment is necessary due to the non-concatenative exponence of some of these markers (Eastlack 1968: 4). As noted above, 1Psg is realized by a 'floating' nasal feature that spreads from the left edge of a word until blocked by an obstruent consonant, which becomes voiced. Thus, *éno* 'mother', *ẽnõ* 'my mother' but *ówoku* 'house', *õwõŋgu* 'my house'. The prefix *j-* indicates 2P in vowel-initial roots (cf. *j-ówoku* 'your house'), but fronting of the leftmost vowel in the word marks 2P elsewhere: *∅-hánaʔi-ti* 'he/she is big, tall', *hénaʔi-ti* 'you are big, tall'. The status of these elements as prefixes is indicated by their left-aligned pattern of realization.

thus be reduced to the roots <*-pahò*> and <*-hacà*>. See the contrast with<*topahò cahacurè*> 'mouth of the river', where <*-pahò*> occurs instead with the 3Psg non-human possessive <*to*> (Marbán 1702: 12, 160; cf. Rose 2015: 244).

Table 1. Forms for Mouth in the Bolivia-Paraná Arawakan Language

Language	Form	Source
Terena	*-pâho*	Author's field data
Old Mojeño	<*nu-pahó*>, <*nu-hacà*>	Marbán 1702: 160
Ignaciano	*-haka*	Ott & Ott 1983: 493
Trinitario	*-haka*	Gill 1970: 7
	-nuku	Françoise Rose, p.c.
Baure	*-noki*	Danielsen 2007: 469
Paunaka	*-niki*	Danielsen & Terhart 2014: 253

Both the Baure and Paunaka forms, in turn, show no obvious formal correspondence to Terena *-pâho* and to Mojeño *-haka*. As mentioned before, Jolkesky (2016: 19) considers the Baure and Paunaka forms cognates of a bound root *-nuku* found only in Mojeño forms such as *hii-nuku* 'mustache', all traced back to a PMGU etymon **-nuki* 'mouth'. Recent data on the Trinitario variety shows, however, that *-nuku* does occur as a non-derived root in Mojeño, meaning either 'mouth (of a person)' or 'neck of a bottle' (Françoise Rose, personal communication).

The comparative data in table 1 suggests that Old Mojeño had a cognate of Terena *-pâho*, but that no such form was retained in the modern Mojeño varieties. Additional comparative data in table 2 below shows, however, that the relevant comparative patterns are more interesting than a simple case of vocabulary obsolescence in Ignaciano and Trinitario.

Table 2. Bolivia-Paraná forms for 'door'

Language	Form	Source
Terena	*paha-péti*	Author's field data
Mojeño Ignaciano	*ta-paha*	Ott & Ott 1983: 344
Mojeño Trinitario	*ta-paho, -pahra*	Gill 1970: 33
Old Mojeño	<*topaho*>	Marbán 1702: 317
Baure	*haki-*	Danielsen 2007: 41
Paunaka	*nuineki*	Lena Terhart (p.c.)

The Terena form for 'door' is *pahapéti*, a nominal compound meaning 'mouth of the house' (cf. *péti* 'house'; see e.g. Ekdahl & Butler 1979: 182). The structure of this form is in agreement with the general pattern for endocentric nominal compounds expressing part-whole relations or material provenance in Terena, which are regularly head-initial, as in *naũwaka* 'beef' (lit. meat-cow; *-naũ* 'meat, flesh'), *hêwetapi'i* 'hen's foot' (foot-hen; *-hêwe* 'foot, leg'). The earliest, published attestation of this form is <*pahapetiˇ*> in Schmidt (1903: 566), who seems, however, to have overlooked the relation between this compound and the root for mouth (see Schmidt 1903: 593).

In relation to the other forms in table 2, note that the formative *ta-* in the Mojeño forms for 'door' is the 3P non-human possessor prefix (Ott & Ott 1983: 36; Rose 2015) also described as an 'impersonal possessive' (Gill 1970: 6). While Old Mojeño 'mouth' appears in Marbán (1702) with the 1Psg possessive prefix <*nu-*>, the noun for 'door' has the same root preceded by the

3Psg non-human possessive <*to*> (see Rose 2015: 244). This is the same marker used with the other use of <*-paho*> denoting a non-human, metaphorically shifted sense, that of 'river mouth' noted above. In Trinitario one has a root *-paho* which, if possessed, shows the suffixation of the Possessive marker *-ra* and the effects of a recurrent process of vowel syncope (*paho-ra* > *pah-ra*; cf. Rose 2015: 253–254). The Paunaka form may be etymologizable to **nui-nɨkɨ*, containing the root *-nɨki ~ -nɨkɨ* 'mouth', but this is not clear, and the remaining formative, *nui-*, is unattested elsewhere (Lena Terhart, personal communication).

Two facts in need of discussion and explanation are revealed by the data in table 2. First, that Terena *-pâho* 'mouth' has an allomorph *-paha* appearing, at least, in the compound *paha-péti* 'door'. Second, that Old Mojeño <*-pahò*> was polysemous, meaning either 'door' or 'mouth'. Note that this differs crucially from Terena *-pâho* 'mouth' and *pahapéti* 'door', two separate lexemes, even if clearly etymologically related. Accounting for the polysemy in the Old Mojeño form will be of vital importance for the diachronic developments postulated here but, first, I will deal briefly with the formal variation internal to Terena, that is, the *-pâho ~ -paha* allomorphy.

4.1. Terena **paho* > *paha*

Of the two forms for 'mouth' attested in Terena, *-pâho* and *-paha*, the former is the older, inherited (conservative) one, while *-paha* is innovative. Though the precise nature of the developments behind the emergence of the allomorph *-paha*, as well as their chronology, will remain an object for future investigation, pending a more thorough understanding of Terena historical phonology and morphology, the postulation of a single pre-Terena allomorph **-paho* 'mouth' seems to be plausible in view of the considerations bellow.

Formally, it is possible to derive *-paha* from *-pâho*, though not the reverse, by invoking a contextual factor such as vowel harmony or assimilation. The restriction of this process to forms such as the compound *pahapéti* 'door' may be explained on the basis of prosodic properties such as 'strength asymmetries' within the Foot or the Prosodic Word. In compounds such as *pahapéti*, from *pâho* 'mouth' and *péti* 'door', the first element usually loses its stress to the rightmost one (Bendor-Samuel 1961: 35), a rough indication of prosodic weakness. According to the evidence in Ekdahl & Butler (1979: 185), incorporated[5] *-pahô* does retain its round vowel *o* where it is stressed, as a comparison of (2a) and (2b) clearly shows (the syllable bearing main stress is indicated in bold):[6]

(2) *Evidence for the role of stress placement in conditioning harmony*

(a) ∅ *-**tí**maru -paha -ʃ -o -wo*
3P -lick -mouth -TH -ACT -REFL
'He/she licked his own mouth'

[5] Certain verbs in Terena, such as *kipó-* 'to wash', allow the incorporation of objects. Thus, given a root such as *-nône* 'face', one has *kipónonewoti* 'he/she washes his/her own face'. An alternative description, which seems to be favored in the literature, postulates the existence of a finite set of bound forms, called either 'qualifiers' (Ekdahl & Butler 1979: 185) or 'verbal classifiers' (Passer 2016) that can appear within the verbal word and may bear only an etymological relation to the independent noun roots.

[6] As in other Arawakan languages, the morphemes glossed 'thematic' in Terena are affixes with little semantic content but which function as a kind of 'stem-closure' formative (Wise 1990: 90). In Terena these thematic suffixes are either *-ʃ* or *-k*. So-called athematic verb stems lack any thematic ending. Structures involving verbs and incorporated nouns function as stems which are either thematic or athematic, just like underived/simplex verb stems, hence the difference between the athematic stem in (2b) and the *ʃ*-thematic stem in (2a).

(b) ∅ *-paru -pa**hó*** *kójee*
3P -open -mouth AUX
'he/she is agape' (lit. 'he/she is with the mouth open')

Moreover, patterns that could be described as instances of *a* ~ *o* ablaut, often though not always associated with vowel harmony, are recurrent throughout the nominal and verbal morphology of Terena. In all cases, the overarching generalization is that *o* is basic and forms with *a* are derived. As noted in section **4**, alienable nouns in Terena, as in other Arawakan languages, differ from inalienable nouns in bearing some additional morphological elaboration in possessive constructions, usually in the form of reflexes of one of the Possessive or Genitive suffixes reconstructed for the PA language. A subclass of these alienable nouns shows, when possessed, a change of every *o* to *a* (see Ekdahl & Butler 1979: 72, 182; a dash indicates a possessed form):

(3) Change of *o* to *a* in Terena alienable nouns

sopôro	*-sápara*	'maize'
wôso	*-wása*	'line, thread'
tôroro	*-tárara*	'gourd'
kohôʃu	*-kaháʃa*	'oven'
wojôre	*-wájara*	'yam'

As the Possessive affixes in Arawakan languages are suffixal in nature, it is probable that the derivation of the possessed forms in (3) show the effects of vowel harmony targeting *o*. Elsewhere, change of *o* to *a* in what looks like a vowel harmony process is attested in Terena verbal morphology, where suffixation of the Irrealis (or Potential Mood) suffix *-a* triggers a change of every *o* in the verb stem to *a* (Ekdahl & Grimes 1964: 263; Ekdahl & Butler 1979: 46–47).

Whether the correct account of the emergence of the allomorph *-paha* will rely for the most part on morphological or prosodic considerations is unclear for now. However, the phonological and morphological patterns addressed above suggest that *-pâho* is the most conservative allomorph, *-paha* being derived by a recurrent process attested elsewhere in the language but conditioned by still unknown conditions in this specific case. This is enough for our present purposes.

4.2. Reconstruction of two etyma **paho* 'mouth' and **paho-peti* 'door': form and meaning

Based on the evidence displayed in tables 1 and 2, plus additional assumptions and facts that will be spelled out in detail below, I propose that a form essentially identical to **-paho* can be reconstructed to an earlier, common stage of development shared by Terena and Mojeño, for the meaning 'mouth' (see section **5** for discussion of what this common stage amounts to *vis-à-vis* the classification in (1)).[7] A compound **-paho-peti* 'door' was also derived at this stage (see Ignaciano and Trinitario *peti* 'house'; Ott & Ott 1983: 300; Gill 1970: 34). It is plausible that after the Mojeño varieties innovated a separate lexeme to express the meaning 'mouth', *-haka*, also attested in Old Mojeño as a competing form along with a reflex of **-paho*, it was no longer

[7] I say 'essentially identical' because the falling pitch contour and vowel length characteristic of Terena *-pâho* 'mouth', and of many other nouns and verbs in this language, has no accepted diachronic explanation at the moment. As these are clearly innovations of the language, I will deal only with the segmental content of the form, the one which is formally comparable to the cognates found in Mojeño.

necessary to employ the modifier *peti* to express the derived (or 'target') meaning 'door', hence the restricted meaning of *paho* as 'door' in the modern Mojeño varieties (*paha* in Ignaciano). As very few linguists and readers can be assumed to be familiar with the languages under discussion, and because there is little systematic historical investigation of the relations between Terena and Mojeño, it is perhaps fitting to present evidence that the reconstruction of **-paho* and **paho-peti* is supported by regular segmental correspondences, even if most of these turn out to be trivial identity correspondences (OM = Old Mojeño):

(4) *Regular segmental correspondences supporting* *-paho *and* *paho-peti.

(a) ***p*** > Ter. ***p*** : Ign. ***p*** : Trin. ***p*** : OM ***p***
FINGERNAIL Ter. *-hî**p**o* : Ign. *-hi**p**aɲa* : Trin. *-hi**p**ɲo* : OM <*nuhi**p**oñó*>; BONE Ter. *-ô**p**e* : Ign. *ʔia**p**e* : Trin. *ó**p**e-ra* : OM <*nuo**p**è*>; DUCK Ter. <***p**ôhi*>, Trin. ***p**ohi* : OM <***p**ohi*>; TO WASH Ter. *-ki**p**o-* : Ign *-si**p**aka* : Trin *-si**p**ko* : OM <*nusi**p**ocò*>; ROOT Ter. ***p**ôe-hewe* : Ign. *ta-**p**are* : Trin. *-**p**ore* : OM <*to**p**orè*>.

(b) ***a*** > Ter. ***a*** : Ign. ***a*** : Trin. ***a*** : OM ***a***
HUSBAND Ter. *-îm**a*** : Ign. *-im**a*** : Trin. *-im**a*** : OM <*niym**a***>; GRANDSON Ter. *-**á**mori* : Ign. *-**á**mari* : Trin. *-**a**mri* : OM <*nu**a**mori*>; NAME Ter. *-îh**a*** : Ign. *-íh**a**re* : Trin. *-íh**a**re* : OM <*nih**à***>; TO HEAR Ter. *-k**â**mo* : Ign. *-s**a**ma* : Trin. *-s**a**mo* : OM <*nus**a**momoroicò*>; PERSON Ter. *ʃ**â**ne* : Ign *atʃ**a**ne* : Trin. *ʔtʃ**a**ne* : OM <*ach**a**nè*>.

(c) ***h*** > Ter. ***h*** : Ign. ***h*** : Trin. ***h*** : OM ***h***
NAME Ter. *-î**h**a* : Ign. *-í**h**are* : Trin. *-í**h**are* : OM <*ni**h**à*>; FINGERNAIL Ter. *-hî**p**o* : Ign. *-hi**p**aɲa* : Trin. *-hi**p**ɲo* : OM <*nuhi**p**oñó*>; TOUCAN Ter. ***h**onôʔe* : Ign. ***h**anare* : Trin. ***h**nore* ; TAIL Ter. *-î**h**i* : Ign *tá-i**h**i-ki* : Trin. *-i**h**gi* : OM <*ta**h**iqui*>; MOON Ter. *ko**h**êʔe* : Ign. *ka**h**e* : Trin. *kó**h**e-ra* : OM <*co**h**è*>.

(d) ***o*** > Ter. ***o*** : Ign ***a*** : Trin. ***o*** : OM ***o***
BONE Ter. *-**ô**pe* : Ign. *ʔia**p**e* : Trin. *ó**p**e-ra* : OM <*nu**o**pè*>; HOUSE (POSS.) Ter. *-pên**o*** : Ign. *-pen**a*** : Trin. *-pen**o*** : OM <*nupen**o***>; BROTHER (ELDER) Ter. *-éʃ**o**wi* : Ign. *-étʃ**a**wi* : Trin. *-étʃ**o**vi* : OM <*nech**o**bi*>; MOON Ter. *k**o**hêʔe* : Ign. *k**a**he* : Trin. *k**ó**he-ra* : OM <*c**o**hè*>; TOUCAN Ter. *h**o**nôʔe* : Ign. *h**a**n**a**re* : Trin. *hn**o**re* ; TO HEAR Ter. *-kâm**o*** : Ign. *-sam**a*** : Trin. *-sam**o*** : OM <*nusam**o**moroicò*>; EARTH, MUD Ter. *m**ó**te* : Ign. *m**a**te-hi* : Trin. *m**o**te-hi* : OM <*m**o**tehi*>; TO STEAL Ter. *-**o**méʃ**o*** : Ign *-**á**metʃ**a*** : Trin. *-**ó**metʃ**o*** ; ROOT Ter. *p**ô**e-hewe* : Ign. *ta-p**a**re* : Trin. *-p**o**re* : OM <*top**o**rè*>.

(e) ***e*** > Ter. ***e*** : Ign. ***e*** : Trin. ***e*** : OM ***e***
Breast Ter. *-ʃ**ê**n**e*** : Ign *-tʃen**e*** : Trin *-tʃen**e*** : OM <*nuch**e**n**e***>; BONE Ter. *-ôp**e*** : Ign. *ʔiap**e*** : Trin. *óp**e**-ra* : OM <*nuop**è***>; BROTHER (ELDER) Ter. *-éʃowi* : Ign. *-étʃawi* : Trin. *-étʃovi* : OM <*nechobi*>; TOUCAN Ter. *honôʔ**e*** : Ign. *hanar**e*** : Trin. *hnor**e*** ; MOON Ter. *kohêʔ**e*** : Ign. *kah**e*** : Trin. *kóh**e**-ra* : OM <*coh**è***>; MOTHER-IN-LAW Ter. *-imós**e*** : Ign. *-ímas**e*** : Trin. *-ims**e*** : OM <*mos**e***>; TO STEAL Ter. *-oméʃo* : Ign *-ámetʃa* : Trin. *-ómetʃo* ; ROOT Ter. *pôe-hewe* : Ign. *ta-par**e*** : Trin. *-por**e*** : OM <*topor**è***>.

(f) ***t*** > Ter. ***t*** : Ign. ***t*** : Trin. ***t*** : OM ***t***
HEAD Ter. *-tû**t**i* : Ign. *-tʃu**t**i* : Trin. *-tʃu**t**i* : OM <*nuchu**t**i*>; BLOOD Ter. *i**t**i* : Ign. *i**t**i* : Trin. *i**t**i* : OM <*i**t**i*>; TERMITE Ter. *mo**t**ôu* : Ign. *ma**t**a-ru* : Trin. *m**t**o-ru* : OM <*mo**t**orù*>[8]; BROTHER (YOUNGER) Ter. *-â**t**i*, Trin. *-a**t**i*, OM <*nua**t**i*>; NIGHT Ter. *jó**t**i* : Ign. *ja**t**i* : Trin. *jo**t**i* : OM <*ya**t**i*>; BAT Ter.[9] *wi**t**ete* : Ign. *wi**t**e* : Trin. *vi**t**e* : OM <*vi**t**ê*>.

[8] The gloss in Marbán (1702: 254) is 'hormigas, que comen la yuca recien plantada'.

[9] Reduplication is a frequent property of animal names in Terena, as in *wáhaha* 'spider', *wétekeke* 'cayman', *firifiri* 'hummingbird' and *howôwo* 'frog'.

(g) ***i*** > Ter. ***i*** : Ign. ***i*** : Trin. ***i*** : OM ***i***
TAIL Ter. *-**î**h**i*** : Ign *tá-**i**h**i**-ki* : Trin. *-**i**hgi* : OM <*tah**i**qui*>; HEAD Ter. *-tût**i*** : Ign. *-ʧut**i*** : Trin. *-ʧut**i*** : OM <*nuchut**i***>; BLOOD Ter. ***i**t**i*** : Ign. ***i**t**i*** : Trin. ***i**t**i*** : OM <***i**t**i***>; BROTHER (YOUNGER) Ter. *-**â**t**i***, Trin. *-at**i***, OM <*nuat**i***>; NAME Ter. *-**î**ha* : Ign. *-**í**hare* : Trin. *-**í**hare* : OM <*n**i**hà*>; FINGERNAIL Ter. *-h**î**po* : Ign. *-h**i**paŋa* : Trin. *-h**i**pŋo* : OM <*nuh**i**poñó*>; TO WASH Ter. *-k**i**po-* : Ign *-s**i**paka* : Trin *-s**i**pko* : OM <*nus**i**pocò*>; JAGUAR Ter. *s**î**n**i*** : Ign. *iʧ**i**n**i*** : Trin. *ʔʧ**i**n**i*** : OM <*ich**i**n**i***>; GRANDSON Ter. *-ámor**i*** : Ign. *-ámar**i*** : Trin. *-amr**i*** : OM <*nuamor**i***>; BROTHER (ELDER) Ter. *-éʃow**i*** : Ign. *-étʃaw**i*** : Trin. *-étʃov**i*** : OM <*nechob**i***>; NIGHT Ter. *jót**i*** : Ign. *jat**i*** : Trin. *jot**i*** : OM <*yat**i***>.

Though the reflex of **-paho* was lost in modern Mojeño varieties as the expression of the meaning 'mouth', it was retained in the form for 'door' as a (slightly) obscure cognate, whose existence often points to the occurrence of semantic or functional shifts.

Correspondence (4d) above, directly relevant for the etymology Terena *-pâho*, Old Mojeño <*-pahò*>, Trinitario *-paho*, Ignaciano *-paha*, requires further discussion. Given the lack of a phoneme *o* in Ignaciano (Ott & Ott 1959: 7–8; 1983: 5–7), this correspondence suggests that a merger **o, *a* > *a* took place in this Mojeño variety. The operation of this merger was suggested by Rose (2015: 245, fn.3) and is explicitly advanced in Jolkesky's (2016: 17) comparative work. Below I subject the relevant correspondence sets to scrutiny, bringing in the data from Terena which so far has not been included in the discussion.

In tables 3 and 4 I show cognate sets for two correspondences, one matching Terena *o* to Ignaciano *a* and Trinitario *o*, and the other having *a* in all three languages (corresponding vowels appear in bold).

Table 3. Correspondence set Ter *o* : Ign *a* : Trin *o*

	Terena	Ignaciano	Trinitario
Night	*j**ó**ti*	*j**a**ti*	*j**o**ti*
Rain, cloud	*û**k**o*	*uk**a***	*uk**o***
Earth, mud	*m**ó**te*	*m**a**te*	*m**o**te*
Grandmother	*-**ô**se*	*-**a**tse*	*-**o**tse*
Wife	*jê**n**o*	*jen**a***	*jen**o***
Shoulder, arm	*-p**ô**w**o***	*-p**a**w**a***	*-p**o**w**o***
Tooth	*-**ô**e*	*-**a**ʔe*	*-**o**ʔe*

Table 4. Correspondence set Ter *a* : Ign *a* : Trin *a*

	Terena	Ignaciano	Trinitario
Sky	*w**a**núke*	***a**nu-ma*	***a**nu-mo*
Stone, stony floor	*m**a**rîpa*	*m**a**ri*	*m**a**ri*
Sun	*k**á**tʃe*	*s**a**tʃe*	*s**a**tʃe*
Person	*ʃ**â**ne*	*atʃ**a**ne*	*ʔtʃ**a**ne*
Son-in-law	*sîŋ**a***	*tʃin**a***	*tʃin**a***
Name	*-î**h**a*	*-ih**a***	*-ih**a***
To hear	*-k**â**mo*	*-s**a**ma*	*-s**a**mo*

The distribution of the two correspondences — Ter *o* : Ign *a* : Trin *o* and Ter *a* : Ign *a* : Trin *a* — does not suggest any contextual factor that could point to a split in Terena and in Trinitario. Since the latter identity correspondence is non-controversially accounted by reconstructing **a*, a sensible assumption is to assign the former, non-identity correspondence to **o*, implying a merger of the two phonemes in Ignaciano.

Some complexities involving these correspondences should be noted. On the one hand, Old Mojeño data is equivocal, as it patterns with Trinitario in showing *o* matching Ignaciano *a* in most cases (see (4d)), but in many instances *a* is found instead in the Old Mojeño cognates of forms showing *o* in Trinitario and *a* in Ignaciano (e.g. OM <*tih**a**pú*> 'white', Trin. *-h**o**pu*, Ign. *-h**a**pu*). My own intuition in this respect is that the Old Mojeño documents of Marbán are dialectally heterogeneous, not an implausible thesis in view of the multi-ethnic environment of the Christian Missions in 17th and early 18th century Llanos de Mojos. Sorting this problem out will demand a detailed investigation of Mojeño phonological diversification, a task beyond the immediate concerns of this paper. On the other hand, there are additional issues arising from attempts at establishing correspondences with languages elsewhere in Arawakan family and with the PA forms reconstructed by Payne (1991). Some of these issues were identified by Payne (1991: 472) himself. My own position, sketched in section **1** is that after Payne's (1991) ambitious attempt at dealing with 24 distinct Arawakan languages in a single stroke, the time has come for comparative investigation of this family to proceed in a bottom-up manner, reconstructing from less inclusive intermediate subgroups. The correspondences amassed above, with the inclusion of data from Terena (arguably the closest relative of Mojeño within the family; see section **5** for discussion) robustly support the inference of a merger **o*, **a* > *a* in Ignaciano, a conclusion which, in my view, was already justified by comparing the known Mojeño varieties, notwithstanding the ambiguous testimony of Old Mojeño. I agree with one of the reviewers of this paper that the issue is not definitely closed, though I would add that bringing Terena data to the discussion not only tilted the balance in favor of the hypothesized merger of **a* and **o* in Ignaciano, but helped constitute a so far unacknowledged set of comparative patterns that must be successfully addressed by any competing explanation.

Before turning to questions of meaning, one should note that the postulation of a compound **paho-peti* 'door' (lit. 'mouth (of the) house') is consistent not only with the structure of endocentric, part-whole compounds in Terena, but also matches the structure of similar compounds in Mojeño. Compounds with modification structures, often involving two inalienable (hence, bound) lexemes follow the same order in this language. For the Ignaciano variety, for instance, Olza Zubiri *et al*. (2004: 219) note that for a root such as *-híja* 'hair' it is possible to derive *-hijatupa* 'chest hair' (cf. *-tupa* 'chest'), *-hijasumu* 'mustache' (cf. *-sumu* 'upper lip'), *-hijamama* 'beard' (cf. *-mama* 'jaw, chin'), *-hijaʔa* 'body hair' (cf. *-ʔa* classifier for the body of humans or large animals).[10]

Reconstruction of the meaning 'mouth' for **-paho*, as opposed to the meaning 'door' attested in the modern Mojeño varieties, seems plausible on the grounds that (1) both Terena and Old Mojeño agree in this respect and (2) 'mouth' seems to be a more basic or salient mean-

[10] The following comments are in order: I assume here a very simple notion of headedness for compounds, one based on meaning. As *paho-peti* 'door' is a 'kind of opening', *-paho* is taken to be the head of the construction. The same reasoning applies to Mojeño forms such as 'mustache' or 'beard', as all are distinct kinds of 'hair'. Note also that the status of Mojeño *-hijaʔa* 'body hair' as a compound is debatable; one could treat classifiers synchronically along with more grammatical or functional markers and claim that *hijaʔa* is a suffixed noun. This is hardly problematic, however, as such bound classifiers are plausibly related, at least diachronically, to independent nominal lexemes.

ing than ‘door’ and, accordingly, metaphorical extensions relating these meanings are expected to operate in the direction ‘mouth’ > ‘door’, rather than in the reverse direction (with ‘mouth’ as the base and ‘door’ as the target). Similar semantic relations, such as ‘mouth’ > ‘estuary, mouth of river’ show a similar pattern where ‘mouth’ is the basic member of the relation (see Urban 2011: 12). In addition, the basic character of ‘mouth’ in relation to ‘door’ is reflected in the overt marking (in the sense of Urban 2011: 6) attested in Terena, where a nominal compound whose head is *paha* (< *paho*) is modified by the noun *peti* to express the meaning ‘door’ (literally “mouth of the house”), a pattern also attested elsewhere in the Arawak family (I return to this below).

Going beyond the mere classification of the semantic change from ‘mouth’ to ‘door’ as a metaphorical extension, standard assumptions make it likely that an intermediate stage of polysemy was involved in this shift (see e.g. Job 1982; Wilkins 1996; Urban 2011). On the model of Wilkins’ (1996: 269) graphic presentation of what he calls the ‘polysemous’ view of semantic change, the proposed relation between the Terena and Mojeño reflexes of the etymon **-paho* can be depicted as follows:

(5) Stages:	(I)	(II)	(III)
Form:	**-paho*	*-paho*	*-paho*
Meaning:	‘mouth’	‘mouth’ & ‘door’	‘door’

Moment (I) has a single form for ‘mouth’, even though, in a compound, it can be modified by the noun **peti* to express the notion ‘door’. This situation is what is attested for Terena, where both *-pâho* ‘mouth’ and *pahapéti* ‘door’ co-exist. Stage (II) is characterized by the existence of polysemy, that is, the two related meanings ‘mouth’ and ‘door’ are associated with the sign *-paho*. This is the pattern attested in Old Mojeño: as seen in section **4**, *-paho* can mean either ‘mouth’ or ‘door’, though the morphosyntactic context establishes one reading over the other. The Old Mojeño pattern provides the core link in a change relating ‘mouth’ and ‘door’ as it shows the existence of synchronic polysemy in a language that arguably represents an early, documented stage of languages that now have *-paho*, the form subject to change, only in the target meaning, in this case, ‘door’ (see Wilkins 1996: 269-270). Finally, stage (III) is that attested in the modern Mojeño varieties Ignaciano and Trinitario. The form *-paho* (*-paha* in the Ignaciano variety) is associated only with the meaning ‘door’, the meaning ‘mouth’ being now associated with a different lexeme *-haka*.

A reviewer suggests an alternative semantic reconstruction according to which a single polysemous form **-paho* ‘mouth/door’ is posited. While Terena would have resolved the inherent ambiguity of the form by means of a compound *paho-peti* ‘door’, in Mojeño the meaning ‘mouth’ was taken over by the innovative form *-haka*, *-paho* being retained only for the meaning ‘door’. This scenario would be preferable for its greater simplicity, as it avoids the postulation of the compound **paho-peti* for any stage of the development of Mojeño, where this compound is, differently from Terena, unattested. Though I find the reviewer’s zeal against postulating this unattested compound structure for Mojeño history highly commendable, I will nevertheless stick to the view sketched above for two reasons: First, the polysemy associated with the Old Mojeño reflex of **-paho* is still something in need of an explanation and, following Urban (2011: 24-29), it is plausible to think that the metaphorical extension behind the use of a form for ‘mouth’ as also meaning ‘door’ was formally mediated, in this case, by the compound **paho-peti*. That is, following Evans (2010) and Urban (2011) I take polysemy to provide a snapshot of semantic change in course, one that is mediated by the kind of formal structure such as nominal compounds.

Second, though it is true that **paho-peti* 'door' is not attested in Mojeño, external evidence in the form of recurrent compounds for 'door' derived from 'house' and 'mouth' in other languages/ branches of the family make the postulation of **paho-peti* not entirely far-fetched. Examples include Yucuna *-numa* 'mouth', *numana* 'door' (Schauer *et al*. 2005: 231), Resígaro *pó?kónoomú* 'door', (Allin 1979: 442; Payne 1991: 408),[11] Bahuana *-numada* (Ramirez 1992: 121), Wapixana *panii-nom* 'door' (WLP 2000: 115), in these cases all having reflexes of a Proto-Arawakan root for 'mouth', reconstructed as **-numa* by Payne (1991: 413) (see also Ramirez 2001: 643). Moreover, the use of such compound expressions is independent of which Proto-Arawakan etymon for 'mouth' happens to be preserved in a language — remembering that Payne (1991) reconstructed three etyma for the meaning 'mouth' at the PA level. Languages like Paresi, which show reflexes of a different form, Payne's (1991: 413) PA etymon **kʰanakɨ* 'mouth', employ the same mechanism, with cognates of pan-Arawakan roots for 'house', as in *hati-kanatse* 'house-mouth', 'door' (see Brandão 2014: 248; Paresi *hati* 'house' is a cognate of Terena *péti*, **p> h* being a regular unconditioned development in Paresi).[12] It is therefore plausible to think that overt marking for the less basic meaning, in this case, 'door', was an intermediate stage in bringing about the polysemy seen in Mojeño (see Urban 2011: 25–29 for more general considerations).

In this section I have offered reasons to support the hypothesis that Terena *-pâho* 'mouth' has cognates in semantically-shifted modern Mojeño (Ignaciano and Trinitario) nouns for 'door', all being reflexes of earlier **-paho* 'mouth'. I have argued that this lexical semantic shift was formally mediated by a compound structure, **paho-peti*, expressing the meaning 'door', a pattern found throughout the Arawak language family. This is in agreement with the more general model of Urban (2011), relating overt marking with preferred directionality trends (as in 'mouth' > 'door') in diachronic semantics. Properties of the intermediate stages, including the existence of polysemy, are retained in the attested material on Old Mojeño, while Terena preserves the reconstructed overt-marking strategy for deriving the meaning 'door' as a compound involving the root for 'mouth' as the head element.

5. On the internal classification of Terena: The Achane branch hypothesis

In (1) I presented a working hypothesis on the internal classification of the Bolivia-Paraná languages, one that places Mojeño and Baure (along with other, less well-known languages such as Paunaka and Paikoneka) in one branch (the 'Bolivia' subgroup, presumably) and Terena as an independent, coordinate branch. As noted in section **2**, this classification probably owes a lot to geographic factors and to non-conclusive assessments of 'relative linguistic proximity' (Walker & Ribeiro 2010: 3; Danielsen, Dunn & Muysken 2011: 185) — such as shared lexical retentions and structural similarities — that are consistent with but not indicative of subgroups. This classification has seemingly attained the status of orthodoxy, to the point that the best recent work on the historical-comparative linguistics of the Bolivia branch of the Bolivia-Paraná subgroup (Jolkesky 2016; labelled 'Mamoré-Guaporé') assumes the more distant position of Terena as a premise not worth discussing.

[11] Resígaro *-noomú* 'mouth' is a straightforward cognate of the other cited forms for 'mouth'. Resígaro has a single back rounded vowel, *o*, corresponding regularly to *u* in the other Northeastern languages such as Yucuna. Word-finally, **a* > *u* in Resígaro (Payne 1991: 473), where <*u*> stands for unrounded *ɯ*.

[12] Though the use of 'mouth' as a base to express the target meaning 'door' is widespread among Arawakan languages, it is not a self-evident fact that it can be reconstructed at the Proto-Arawakan level. Some daughter languages, Wayuunaiki and Baniva de Maroa being two examples, use the root for 'eye' instead: in Wayuunaiki, given *-oʔu* 'eye' and *piiʧi* 'house', one has *piiʧoʔu* 'door' (see Captain & Captain 2005: 36). In Baniva de Maroa, *paniʂipuli* 'door' is a compound of *panîʂi* 'house' and *-puli* 'eye' (Mosonyi 2000: 504).

In the preceding sections of this paper I have relied heavily on Mojeño data to elucidate the etymology of a Terena 'basic vocabulary' noun, while Baure, the other (relatively) well-described member of the Bolivia-Paraná subgroup has contributed nothing. I was unable to find in this language any cognate of the Terena and Mojeño forms for 'mouth' or 'door' studied here, and this seems to agree with the comparative vocabulary amassed by Jolkesky (2016) in his study of the Bolivian languages.

I advance here the hypothesis (see (6) below) that Terena and Mojeño are more closely related to each other than any of these is to Baure, as an alternative to the scheme in (1).

(6) Achane branch and the internal classification of Bolivia-Parana Arawakan
- Bolivia-Paraná subgroup
  - Baure
  - Paikoneka
  - Achane branch
    - Terena
    - Mojeño
    - Paunaka

To the best of my knowledge, this is the first time an internal classification of the Bolivia-Parana languages in which a branch composed of Terena, Mojeño and Paunaka, but excluding Baure and Paikoneka, is proposed (I will, for the moment, rely on Jolkesky 2016 for the assumption of a rather close relation between Baure and Paikoneka). The forms *-*paho* 'mouth' and **paho-peti* 'door' can be taken as reconstructions for the common ancestor of this branch, or Proto-Achane (after the Old Mojeño noun for 'person'). Moreover, the fact that Terena *-pâho* and its previously undetected cognates in Mojeño have not been included in any compilation of Arawakan cognates (see section **3**) is at least suggestive of its status as an innovation. Note that, in this respect, the etymon *-*paho* differs from forms such as Baure *-noki* and Mojeño *-nuku*, given that the latter two, on the contrary, fit clearly within Payne's (1991: 413) 'Mouth3' etymology, having cognates in Wayuunaiki *-aanɨkɨ* and Waurá *-kanatɨ*, both meaning 'mouth' as well, and occurring in languages far apart from each other within the family. Therefore, an etymon close to Jolkesky's (2016) **-nukɨ* 'mouth' can be plausibly assumed for Proto-Bolivia-Paraná, while *-*paho* 'mouth' would constitute a shared innovation of the Achane languages. The best candidates I am aware of for the status of cognates of *-*paho* 'mouth' outside of the Bolivia-Paraná subgroup are the forms attested in the languages of the Campa branch, such as Nanti and Matsigenka *-bagante* and Asháninka *-paante* (see Michael 2011 and Heitzman 1973: 37). However, though Michael (2011) has successfully reconstructed the segmental phonology of the Proto-Campa language he offers no reconstructed etyma, there being no published reconstruction of the lexicon of Proto-Campa that supersedes the deeply flawed reconstruction of Matteson (1972). The present author is currently working on a lexical and morphological reconstruction of Proto-Campa and, if it turns out that the Proto-Campa etymon for 'mouth' is indeed a cognate of Proto-Achane *-*paho*, this would invalidate its status as a Proto-Achane innovation.

Nevertheless, interesting additional evidence from the lexical and morphological domains furnish strong candidates for the status of shared innovations pointing to a stage of development common to Terena and Mojeño but not to Baure. The Baure root *-poʔe* 'head' (Danielsen 2007: 120) is used both as a syntactically independent noun (with appropriate morphology) or incorporated into a verb stem (Danielsen 2007: 126). Baure *-poʔe* is plausibly a cognate of Mojeño Ignaciano *-puʔi*, a classifier indicating round or spherical objects (Olza Zubiri *et al*. 2004: 286–288) and of Terena *-puʔi*, a classifier for 'head-like' objects (Ekdahl & Butler 1979:

167, 185). Moreover, this set probably reflects a much older etymon, as shown by apparent cognate forms attested in widely separate languages of the family, such as Baniva de Maroa *-bu* 'head' (Mosonyi 2000: 511), Yucuna *-pula* 'forehead' (Schauer & Schauer 2005: 205) and Garífuna *ábu-lugu* 'head' (Sabio & Ordoñez 2006: 7). Terena and Mojeño agree, however, in employing their cognates of Baure *-poʔe* only as incorporated or bound classifiers; Terena *-tûti* and Mojeño *-ʧuti* are used instead for 'head' when this is expressed as a syntactically independent expression. Terena *-tûti* and Mojeño *-ʧuti* are, like Proto-Achane **-paho*, reasonably good candidates for being innovations. Payne (1991: 405) included *-tûti* and *-ʧuti* in his cognate set for PA 'forehead' though this is, for diverse reasons, a questionable etymology. Note, first, that the final syllables of both forms are arbitrarily excised from the comparison; there are, however, no clear grounds for analyzing these as *-tu-ti* and *-ʧu-ti*, respectively, at any level. Second, Payne proposes that Ashéninka and Matsigenka reflexes of Proto-Campa **gi-to* 'head' (Matteson 1972: 213) present, in the final syllable *-to*, a cognate of the *-tu-/-ʧu-* formative he identifies in the Terena and Mojeño forms. Recent and more extensive documentation of Ashéninka varieties reveals, however, that Payne's (1991: 405) analysis of *-to* in **gi-to* as a classifier meaning 'head-shaped, round' is incorrect; *-to* is, indeed, a classifier, but its meaning is 'hollow, long, rigid' (see Mihas 2015: 414), where the meaning 'rigid' is probably the one relevant for 'head' (see that the remaining morpheme **-gi-* is a straightforward reflex of the PA etymon **kiwɨ* 'head' reconstructed by Payne 1991: 407).

Certain specific morphological patterns attested in both Terena and Mojeño, but not in Baure, also imply a rather close structural similarity between the putative members of the Achane branch, and preliminary inspection of comparative data suggest that these could be shared innovations. A 1Ppl verbal suffix is usually reconstructed as **-wa* or **-w(a)* at the PA level or at another intermediate level (see Aikhenvald 1999: 88; Danielsen 2011: 514–515, the latter for 'Proto-Southern-Arawakan'). In Mojeño, this suffix has not only a final *i* vowel that seems characteristic of some southern Arawakan languages (see Danielsen 2011: 215) but differs as well in having a -VCV structure, appearing as *-avi* (Rose 2015: 244). Interestingly, this initial vowel of the Mojeño 1Ppl suffix, unattested anywhere else in the family, furnishes an explanation for a morphophonological quirk of Terena: in this language, the 1Ppl has the form *-wi* but it is unique among all person-marking suffixes in the language in that it triggers the lengthening of a preceding vowel (see Eastlack 1968: 5; Ekdahl & Butler 1979: 35). Thus, contrasting with *peréʃa-**nu** úne* '(you) give me water!', with the 1Psg object suffix *-nu*, one has *peréʃaa-**wi** úne* '(you) give us water!', with the 1Ppl object suffix *-wi* triggering lengthening of the final vowel of the verb stem (data from Ekdahl & Butler 1979: 33). This lengthening effect on a preceding vowel can be easily explained as the result of *sandhi* processes (compensatory lengthening) involving absorption of the vowel present in the Mojeño cognate suffix, a suffix-initial vowel not reconstructed for the PA language.

There is, I submit, enough reasons to consider the existence of a branch I label Achane, including Terena and Mojeño but excluding Baure, as a credible alternative to the geographically-based scheme in (1). At this point, however, the existence of this Achane branch is no more than a hypothesis worth investigating; it goes without saying that further investigation of potential shared innovations in lexicon, morphology and phonology is necessary before any definite conclusions can be attained.

6. Conclusion and final remarks

This paper demonstrated, with material from a language that is highly understudied from a diachronic standpoint, how etymological analysis, if properly conceived and conducted, can

considerably increase our understanding of the historical development of a language and its relatives.

I have shown here that Terena *-pâho* 'mouth 'has cognates in the different speech varieties forming the Mojeño language. Accounting for this etymology calls, however, for an explicit hypothesis relating *-pâho* 'mouth' to forms in Ignaciano and Trinitario whose meaning is 'door', thus implying the action of a lexical semantic shift. Evidence from Old Mojeño was crucial in that it shows a stage in which a reflex of **-paho* is associated with synchronic polysemy, which is predicted to exist given our general understanding of how semantic change proceeds (see Wilkins 1996: 269–270; Urban 2011). The whole account proposed involves the postulation of two proto-forms, a root **-paho* 'mouth' and a nominal compound **paho-peti* 'door', derived with the use of **peti* 'house', as a modifier of the noun **-paho*. It is plausible that after the Mojeño varieties innovated a separate lexeme to express the meaning 'mouth', *-haka*, also attested in Old Mojeño as a competing form along with a reflex of **-paho*, it was no longer necessary to employ the modifier *peti* to express the derived (or 'target') meaning 'door', and, consequently, reflexes of **-paho* came to mean 'door' exclusively in these speech varieties. The etymological account proposed is consistent not only with the usual formal, that is, phonological and morphological constraints on compelling etymologies, but is also consistent with constraints on semantic reconstruction. Reference to the more general and widespread character of formations for 'door' as a compound involving nouns for 'mouth' and 'house', in particular to the ubiquity of this pattern in the Arawakan language family, also meets one of the demands usually placed on credible semantic developments (see Job 1982). In the end, a rather close relationship between Terena and Mojeño is suggested by the argumentation presented here, advancing the hypothesis of a branch composed of the most recent and exclusive common ancestor of these two languages as a viable proposal for internal classification.

References

Aikhenvald, Alexandra. 1999. The Arawak Language Family. In: R. M. W. Dixon, Alexandra Aikhenvald (eds.) *The Amazonian Languages*. Cambridge University Press: 65–106.

Brandão, Ana Paula. 2014. A Reference Grammar of Paresi-Haliti (Arawakan). Ph.D. Dissertation, The University of Texas at Austin.

Campbell, Lyle. 2012. Classification of the Indigenous Languages of South America. In: Lyle Campbell, Verónica Grondona (eds.) *The Indigenous Languages of South America: A Comprehensive Guide*. De Gruyter Mouton: 59–166.

Captain, David, Linda Captain. 2005. *Diccionário Básico Ilustrado Wayuunaiki- Español, Español-Wayuunaiki*. Bogotá: Editorial Buena Semilla.

Carvalho, Fernando O. de. 2015. On the Realization of Nominal Possession in Mehinaku: A Diachronic Account. *International Journal of American Linguistics* 81 (1): 119–132.

Carvalho, Fernando O. de. 2016a. Terena, Chané, Guaná and Kinikinau are one and the same language: Setting the record straight on Southern Arawakan linguistic diversity. *LIAMES* 16 (1): 39–57.

Carvalho, Fernando O. de. 2016b. Internal and Comparative Reconstruction in Yawalapiti: Palatalization and Rule Telescoping. *International Journal of American Linguistics* 82 (3): 285–316.

Carvalho, Fernando O. de. 2016c. The Diachrony of Person-Number Marking in the Lokono-Wayuunaiki Subgroup of the Arawakan Language Family: Reconstruction, Sound Change and Analogy. *Language Sciences* 55: 1–15.

Carvalho, Fernando O. de. Forthcoming. Fricative Debuccalization and Primary Split in Terena (Arawakan) Historical Phonology. *International Journal of American Linguistics* (july/2017 issue).

Danielsen, Swintha. 2007. *Baure: An Arawakan Language of Bolivia*. Indigenous Languages of Latin America (ILLA). University of Leiden.

Danielsen, Swintha. 2011. The Personal Paradigms in Baure and other Southern Arawakan Languages. *International Journal of American Linguistics* 77 (4): 495–520.

Danielsen, Swintha, Lena Terhart. 2014. Paunaka. In: Mily Crevels, Pieter Muysken (eds.) *Lenguas de Bolivia, Vol. III: Oriente*. La Paz: Plural Editores: 221–258.

Danielsen, Swintha, Michael Dunn, Pieter Muysken. 2011. The Spread of the Arawakanan Languages: A View from Structural Phylogenetics. In: Alf Hornborg, Jonathan D. Hill (eds.) *Ethnicity in Ancient Amazonia*. University Press of Colorado: 173–195.

Eastlack, Charles. 1968. Terena (Arawakanan) Pronouns. *International Journal of American Linguistics* 34 (1): 1–8.

Ekdahl, Elizabeth, Nancy Butler. 1969. Terêna Dictionary. Arquivo Lingüístico 095. Brasília: Summer Institute of Linguistics.

Ekdahl, Elizabeth, Nancy Butler. 1979. *Aprenda Terêna*. Brasília: Summer Institute of Linguistics (SIL).

Evans, Nicholas. 2010. Semantic Typology. In: Jae Jong Song (ed.) *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press: 504–533.

Gill, Wayne. 1957. *Conversación y Gramática Trinitaria*. San Lorenzo de Mojos: Misión Evangélica Nuevas Tribus.

Gill, Wayne. 1970. *Diccionario Trinitario-Castellano y Castellano-Trinitario*. Misión Evangélica Nuevas Tribus.

Heitzman, Allene. 1973. Reflexes of Some Proto-Campa Consonants in Modern Campan Languages and Dialects. MA Thesis, University of Kansas.

Hill, Jonathan D., Fernando Santos-Granero. 2002. *Comparative Arawakanan Histories*. Chicago: University of Illinois Press.

Hornborg, Alf, Jonathan D. Hill. 2011. *Ethnicity in Ancient Amazonia*. Boulder: University Press of Colorado.

Job, D. M. 1982. Semantic Change and Etymologies. In: Anders Ahlqvist (ed.) *Papers from the 5th International Conference on Historical Linguistics*. Amsterdam: John Benjamins: 163–170.

Jolkesky, Marcelo. 2016. Uma Reconstrução do Proto-Mamoré-Guaporé (Família Arawakan). *LIAMES* 16 (1): 7–37.

Kaufman, Terrence. 1990. Language History in South America: What We Know and How to Know More. In: Doris Payne (ed.) *Amazonian Linguistics: Studies in Lowland South American Languages*. Austin: University of Texas Press: 13–67.

Kaufman, Terrence. 1994. The Native Languages of South America. In: Christopher Moseley, R. E. Ascher (eds.) *Atlas of the World's Languages*. Routledge: 46–76.

Lawrence, Aimee. 2014. Reconstruction of Proto-Kampa Verbal Morphology. *Proceedings of the Annual Meetings of the Berkeley Linguistics Society 2014*: 256–271.

Mailhammer, Robert. 2014. Etymology. In: Claire Bowern, Bethwyn Evans (eds.) *The Routledge Handbook of Historical Linguistics.* London and New York: Routledge: 423–441.

Marbán, Pedro. 1702. *Arte de la Lengua Moxa*. Lima: Imprenta Real de Joseph de Contreras.

Martius. Karl Friedrich von. 1867. *Beiträge zur Ethnographie und Sprachenkunde Amerikas zumal Brasiliens*. Leipzig: Friedrich Fleischer.

Matteson, Esther. 1972. Proto-Arawakanan. In: Esther Matteson (ed.) *Comparative Studies in Amerindian Languages.* The Hague: Mouton.

Michael, Lev. 2009. Review of Ramirez 2001. *International Journal of American Linguistics* 75 (3): 447–450.

Michael, Lev. 2011. *La Reconstrucción y la Clasificación de la Rama Kampa de la Família Arawakan*. Presented at CILLA V, 2011.

Mihas, Elena. 2015. *A Grammar of Alto Perené (Arawak)*. Berlin: De Gruyter Mouton.

Mosonyi, Esteban E. 2000. Introducción al Análisis del Idioma Baniva. In: María S. González de Pérez, María L. Rodríguez de Montes (eds.) *Lenguas Indígenas de Colombia: Una Visión Descriptiva*. Santafé de Bogotá: Instituto Caro y Cuervo: 499–513.

Olza Zubiri, Jesús, Conchita Nuni de Chapi, Juan Tube. 2004. *Gramática Moja Ignaciana (Morfosintaxis)*. San Cristóbal: Universidad Católica de Táchira.

Ott, Willis, Rebecca B. de Ott. 1959. Fonemas de la Lengua Ignaciana. *Notas Lingüísticas de Bolivia, 1.* La Paz: Instituto Lingüístico de Verano (ILV): 1-13.

Ott Willis, Rebecca B. de Ott. 1983. Diccionario Ignaciano y Castellano Con Apuntes Gramaticales. Cochabamba: Summer Institute of Linguistics (SIL).

Passer, Matthias B. 2016. (What) Do Verbal Classifiers Classify? *Lingua* 174: 16–44.

Payne, David. 1990. Some Widespread Grammatical Forms in South American Languages. In: Doris Payne (ed.) *Amazonian Linguistics: Studies in Lowland South American Languages*. Austin: University of Texas Press: 75–87.

Payne, David. 1991. A Classification of Maipuran (Arawakanan) Languages Based on Shared Lexical Retentions. In: Desmond Derbyshire, Geoffrey K. Pullum (orgs.) *Handbook of Amazonian Languages*. Vol. 3. Mouton de Gruyter: 355–499.

Ramirez, Henri. 1992. *Le Bahuana: Une Nouvelle Langue de la Famille Arawak*. *Amerindia* 17.

Ramirez, Henri. 2001. *Línguas Arawakan da Amazônia Setentrional*. Manaus: Editora da Universidade do Amazonas.

Rose, Françoise. 2015. Innovative Complexity in the Pronominal Paradigm of Mojeño: A Result of Contact? F. Gardani *et al.* (eds.) *Borrowed Morphology*. Berlin: The Gruyter Mouton: 239–265.

Sabio, Fernando, Celia K. Ordoñez. 2006. *Hererun Wagüchagu*. La Ceiba: Asociación Misionera Garífuna de Honduras.

Schmidt, Max. 1903. Guaná. *Zeitschrift für Ethnologie* 35 (2/3): 324–336.

Tadmor, Uri, Martin Haspelmath, Bradley Taylor. 2010. Borrowability and the Notion of Basic Vocabulary. *Diachronica* 27 (2): 226–246.

Taunay, Alfredo. 1868. *Scenas de Viagem: Exploração entre os Rios Taquary e Aquidauna no Distrito de Miranda*. Rio de Janeiro: Typographia Americana.

Urban, Matthias. 2011. Asymmetries in Overt Marking and Directionality in Semantic Change. *Journal of Historical Linguistics* 1 (1): 3–47.

Walker, Robert, Lincoln A. Ribeiro. 2010. Bayesian Phylogeography of the Arawakan Expansion in Lowland South America. *Proc. R. Soc. B* 278: 2562–2567.

Wilkins, David. 1996. Natural Tendencies of Semantic Change and the Search for Cognates. In: Mark Durie, Malcolm Ross (eds.) *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. Oxford University Press: 264–304.

Wise, Mary Ruth. 1990. Valence-Changing Affixes in Maipuran Arawakan Languages. In: Doris Payne (ed.) *Amazonian Linguistics: Studies in Lowland South American Languages*. Austin: University of Texas Press: 89–116.

WLP (Wapixana Language Project). 2000. *Scholar's Dictionary and Grammar of the Wapishana Language*. Lethem: Wapishana Language Project and SIL International.

Фернанду де Карвалью. Этимология терена (аравакской) основы 'рот' и ее значение для внутренней классификации аравакских языков

В статье рассматривается этимология существительного *-pâho* 'рот' в языке терена. В современном аравакском языкознании принято считать, что в других языках этой семьи у нее нет когнатов; автор, однако, показывает, что такие когнаты все же есть в языке мохеньо, близкородственном терена (оба входят в боливийско-паранскую подгруппу аравакской семьи). В игнасиано и тринитарио (два наиболее хорошо изученных диалекта мохеньо) эти когнаты подверглись семантическому сдвигу и стали означать 'дверь'. Автор описывает семантические и формальные связи между этими формами через реконструкцию простой основы **-paho* 'рот' и композита **paho-peti* 'дверь' (букв. 'рот /дома/'). Данная этимологизация имеет значимость и с точки зрения внутренней классификации аравакских языков: автор предполагает, что терена и мохеньо составляют отдельную ветвь аравакской семьи — подгруппу ачане (в которую не входит язык бауре).

Ключевые слова: аравакские языки, язык терена, семантические сдвиги

George Starostin
Russian State University for the Humanities / Russian Presidential Academy; gstarst@rinet.ru

Lexicostatistical Studies in East Sudanic I: On the genetic unity of Nubian-Nara-Tama

In this paper, I present a detailed lexicostatistical survey of the reconstructed 50-item word-lists (the "more stable" half of the classic Swadesh list) for three language groups of North-east Africa — Nubian, Nara, and Tama, commonly ascribed to the East Sudanic family and often described in related literature as forming a specifically tight-knit node within that taxon. The survey shows that both the number and the nature of direct lexicostatistical matches between these three groups is plausibly interpretable as decisive evidence for genetic relationship, adding one more formal confirmation to the evidence previously assembled by J. Greenberg, M. L. Bender, Claude Rilly and other scholars. Glottochronological interpretation of the evidence, however, indicates that Nubian-Nara-Tama should be dated to at least the 5th millennium BC, which makes it older than Indo-European and presumably very hard to reconstruct in sufficient detail. The paper itself is the first in a series of planned publications that will explore the East Sudanic hypothesis from a combined lexicostatistical and etymological perspective.

Keywords: Nilo-Saharan languages, East Sudanic languages, Nubian languages, Tama languages, African historical linguistics.

General introduction

Of the three macrofamilies that Joseph Greenberg had delineated in his seminal works on African language classification (most importantly Greenberg 1966[1]), the "Nilo-Saharan" taxon has always shared the most vague outlines. While Greenberg's "Niger-Kordofanian" languages are informally understood as "the ones with the complex noun class systems" (subsequently, the few subgroups that violate this feature, such as Mande, are sometimes viewed with suspicion even by supporters of the Niger-Kordofanian hypothesis[2]), and Greenberg's "Khoisan" is just as informally understood as "the click family", there are no such definitive features to characterize all, or even the majority of the language groups that, according to Greenberg, constitute the Nilo-Saharan macrofamily: the hypothesis is based on numerous, if not properly systematized, lexical and grammatical resemblances rather than any structural homologies.

This fact in itself is not necessarily problematic for historical linguists, since it is commonly accepted, and has frequently been pointed out by Greenberg himself, that genetic relationship is not to be established based on typological features of languages, easily open to areal influence (cf. the spread of "Khoisan" click phonemes to neighboring Southern Bantu languages), but should always be defined primarily by the presence of important homologies

[1] The fourth macrofamily — Afro-Asiatic, formerly known as Hamito-Semitic — was already more or less securely recognized as a genetic unity long before Greenberg's works, and may be kept out of any general discussion on the overall quality of Greenberg's methods and arguments.

[2] Blench (2011) presents a seemingly strong case for the innovative nature of nominal class markers in the bulk of NK, but this view has not yet gained extensive support from specialists.

in the phonetic structures of lexical and grammatical morphemes bearing identical or similar meanings. To that end, Greenberg's argumentation in favor of his macrofamilies always consists of comparative lists and tables of such morphemes. Nevertheless, typological considerations still continue to play an important part in the general acceptance of macrofamily hypotheses — if anything, they offer intuitive support in situations where form-based arguments are either too complex or too dubious for us to quickly assimilate and evaluate. Since the primary methodology behind Greenberg's macrofamilies has been that of "mass comparison", commonly criticized by linguists as a procedure that is unable to properly separate genuine traces of genetic relationship from either areal contacts or chance similarities, it is not surprising that his comparative lists of words and morphemes do not seriously impress modern specialists, whereas such features as the presence of click phonemes or noun class markers do — at the very least, such structural homologies cannot be easily explained away as accidental resemblances.

In this type of situation, linguists who properly dedicate themselves to the construction of an optimal scenario of genetic relationship in a particular linguistic area should find it of essential importance to define specific sets of "genetic markers" (a term that seems quite naturally borrowable from molecular biology) that concisely characterize the postulated taxon and distinguish it from its neighbors. Roughly speaking, such markers should:

(a) constitute either grammatical morphemes or lexical roots that belong to the basic (i.e. generally more resistant to diachronic change) layer of language;

(b) be reconstructible for *all* or *most* of the proposed subbranches of the taxon (at the very least, be reliably reconstructible in its most distant branches, to assure their protolanguage status);

(c) respect the general laws of phonetic change, suggested for the taxon, or, if the taxon is a high-level one, at least yield reflexes in daughter branches that could be deemed "phonetically compatible", i.e. explainable through typologically and historically realistic scenarios of phonetic change[3];

(d) demonstrate either the *exact same* meaning in all or most of the daughter branches, or display minimal semantic variety, confined to diachronically and synchronically frequent types of semantic change or polysemy found in the world's languages (such as 'eye : see', 'black : dark', 'know : hear', etc.)[4];

(e) preferably, at least some of them should be *exclusively* representative of the suggested taxon, in that it could be at least approximately demonstrated that they are reconstructible in that particular form and meaning for the proto-language of that particular taxon and no other.

For linguistic taxa that have diverged within the last five or six thousand years and whose linguistic history has been reasonably well studied, due to an abundance of both primary data and analytical research, the presence of such genetic markers is an obvious fact — a lexical root such as, e.g., Proto-Indo-European *ok^{w}- 'eye' satisfies all of the listed conditions. For speculative linguistic "macrofamilies" whose hypothetical age goes far beyond the specified chronological range, producing such markers is a highly complex challenge, since the prob-

[3] A detailed explanation of the idea of "phonetic compatibility" and its difference from both the weaker criterion of "phonetic similarity" and the stronger criterion of "phonetic correspondence" may be found in Starostin 2013: 57–64.

[4] Although, as of now, there is still no single definitive list of such polysemies that would be both sufficiently comprehensive and obtained through a formal methodology, progress is slowly being made with such works as Youn et al. 2016. As far as basic lexicon is concerned, careful fixation of attested polysemies is conducted by contributors to the Global Lexicostatistical Database project, which allows to perform rough statistical estimates of what may count for a "trivial" polysemy or semantic shift.

ability of their successful recovery decreases with each added millennium. Nevertheless, even a highly limited set may be convincing if it can be shown to have been arrived at without any distortions of available evidence or violations of known tendencies of language change through idiosyncratic assumptions.

In the case of Nilo-Saharan, the proper search for such "genetic markers" was originally launched by M. Lionel Bender, whose sets of "excellent", "good", and "fair" isoglosses (Bender 1997: 77–105), assembled in favor of the hypothesis, satisfy some of the above-listed criteria. However, even some of his "excellent" isoglosses play quite loosely with semantics (e. g. such connections as 'elbow/claw/foot' or 'horn/bone/rib' are quite suspicious) and remain uninterpretable in terms of reasonable historical scenarios of semantic change; numerous phonetic deviations are recorded without any attempts at constructive explanations; and, perhaps most importantly, a huge number of comparanda are not shown to be reconstructible for the required intermediate levels of comparison, which means that they have been too quickly transferred to a deeper level of comparison without proper completion of the preceding stage of analysis — and, consequently, without a reliable "safety net" against accidental resemblances.

The late Lionel Bender himself may have been well aware of these limitations of his own research; in any case, it is somewhat instructive that, instead of expanding his relatively short overview monograph on Nilo-Saharan (Bender 1997) to the size of an etymological dictionary (such as the huge, but ultimately unconvincing volume by Christopher Ehret (2001)), he preferred to follow it up with an equally short comparative treatise on East Sudanic (Bender 2005) — a pioneering study, focusing on one of the largest sub-taxa originally defined by Greenberg within Nilo-Saharan.

The natural implication behind Bender's East Sudanic book is that, without a proper understanding of what exactly is "East Sudanic", we cannot gain any understanding of what exactly could be "Nilo-Saharan". Ironically, in his introduction to the book, Bender mentions having been unable to establish an "East Sudanic Working Group", since "the main problem seems to be that no one is willing to go beyond a narrower focus on sub-families" (p. vi). Indeed, genealogical nodes like East Sudanic find themselves in double trouble: the proverbial "splitters" (or simply specialists with a narrow focus) are not interested in working on them because the explored genetic connections are seen as too deep and complicated to recover, whereas the proverbial "lumpers" (linguists with a pronounced interest in macro-comparative studies) view them, at best, as quick stepping stones, postulated mainly for the sake of classificatory convenience, then more or less forgotten as the interest rapidly shifts to highest-level taxa.

The only work other than Bender's all-too-brief monograph that actually tries to tackle East Sudanic on a serious basis seems to be Rilly 2009, which includes a very thorough comparative analysis of the phonological systems and lexica of those branches that, according to the author, constitute the "Northern" division of this family, including Nubian, Tama, Nara, and Nyimang. However, even in Rilly's book, the arguments in favor of East Sudanic are not really assigned any stand-alone value; rather, they are considered significant inasmuch as they help determine the genetic affiliation of the Meroitic language, which, based on scarce evidence of often dubious quality, Rilly seeks to relate to "Northeast Sudanic" (including Nubian, which seems to have the strongest links with Meroitic, although it still remains unclear whether most of them are of a genetic or areal nature). Furthermore, dealing with but one branch of East Sudanic is certainly not the same thing as trying to evaluate the validity of the entire family.

It was mostly these considerations that eventually led to a general lexicostatistics-based survey of possible genetic connections between the various groups of languages that constitute Greenberg's "Nilo-Saharan", in which the East Sudanic hypothesis was tested first — without taking into account any higher level connections. The test, carried out as part of a

large ongoing project on the general classification of African languages, followed a standardized methodology that had already been tried out on the so-called "Khoisan" languages, yielding results that seem to be largely consistent with current mainstream views on their classification (Starostin 2013). The main stages of this procedure may be briefly summarized as follows.

1. Define the primary constituents of the analysis. These are identified as relatively small language groupings whose genetic reality is beyond reasonable doubt and commonly accepted by all specialists — e.g., Nubian, Tama, Daju, Kuliak, etc.; all the languages within each such group share numerous cognates easily linked together with sound laws, as well as robust grammatical isoglosses, indicating a relatively recent split from a common ancestor (not to exceed 2,000–3,000 years based on any available historical, archaeological, and lexicostatistical estimates).

2. Assemble and check complete 100-item Swadesh lists for as many languages of these small groupings as possible, based on the most recent and accurate sources available. The compilation procedure closely follows the guidelines that were laid down in earlier methodological publications (Starostin 2010; Kassian et al. 2010).

3. Carry out a lexicostatistical analysis of the data in order to determine the internal classification of the groupings (most importantly, the primary splits within each of them; these results will have a direct bearing on the efficiency of point 4).

4. Reconstruct the proto-wordlist for each such grouping, based on regular etymological analysis and a complex set of criteria used to determine the "optimal" candidate for the expression of each particular Swadesh meaning in the protolanguage. Unlike wordlists for attested languages, reconstructed proto-wordlists are limited to 50 of the most generally stable Swadesh items (out of 100), since reconstruction of the second, less stable, half usually turns out to be cost-ineffective for purposes of high-level comparison and classification[5]. As a rule, this is the most complicated, time-consuming, and text-heavy part of the entire procedure (unless the group in question consists of several very closely related dialects that do not require detailed historical analysis).

5. Subject the reconstructed proto-wordlists to several additional stages of lexicostatistic analysis, which include running a completely automatic procedure of finding "pseudo-cognates" between reconstructions, based on the "Dolgopolsky consonantal classes" method of phonetic comparison (general description of the method and an example of its application may be found in Kassian, Zhivlov, Starostin 2015). After that, the results undergo a procedure of "manual correction" which takes into account the locally specific phonetic features of compared (proto-)languages, not recognized in the universally applicable method.

6. Compare the lexicostatistical matrices and classificatory trees generated by the "fully automated" and "manually corrected" methods and select one as the optimal choice for a working model (in most cases, this turns out to be the tree/matrix based on the "manually corrected" list of hypothetical cognates, although there may be occasional exceptions).

The current results of this procedure[6] are summarized in the following lexicostatistical matrix (Table 1) and phylogenetic tree (Fig. 1), both of them reflecting the "manual correction"

[5] See Starostin 2010 for additional information on how the average "stability index" for various Swadesh items was calculated and on other technical factors that have influenced the final compilation of the universally applicable 50-item list. The procedure of proto-wordlist reconstruction, illustrated by specific examples, is described in detail in Starostin 2016.

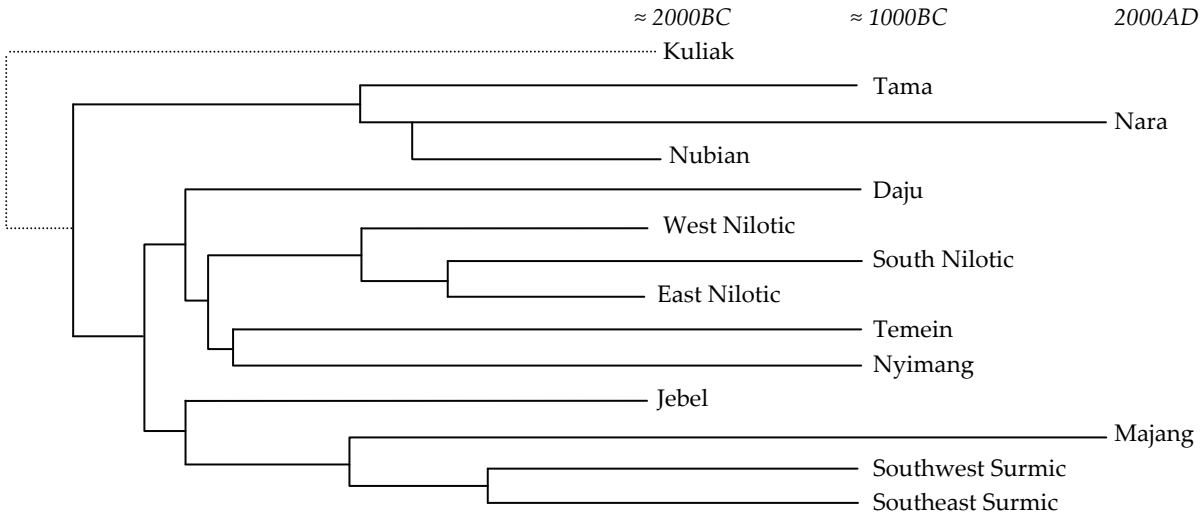
[6] These results differ slightly, but not crucially, from the results published earlier in Starostin 2014: 677 — an inevitable development that is due to corrections of previously produced reconstructions in the light of newly available data or occasional spotted mistakes in previous analysis. It goes without saying that these results as well are liable to future amendments, since new sources of data that allow for deeper insights become available to researchers on a steady basis.

model (which is not very different from the fully automatic model, except for the relative position of the Daju branch on the tree; this is due to certain rare types of phonetic change that took place on the way from the Proto-East Sudanic stage to Proto-Daju, some of which are quite evident even on the limited data of the 100-item wordlists).

Table 1. Lexicostatistical matrix for Greenberg's "Eastern Sudanic" (50-item wordlists).

	Nara	Tama	SWS	SES	Maj.	WNil	ENil	SNil	Nyi.	Tem.	Jebel	Daju	Kul.
Nubian	26%	20%	14%	12%	4%	18%	16%	20%	22%	12%	12%	4%	8%
Nara		20%	10%	10%	8%	12%	10%	12%	12%	12%	12%	6%	4%
Tama			6%	10%	6%	8%	12%	16%	12%	6%	4%	6%	2%
Southwest Surmic				40%	22%	16%	14%	20%	14%	14%	18%	8%	6%
Southeast Surmic					14%	20%	12%	18%	12%	14%	16%	10%	4%
Majang						12%	10%	10%	10%	10%	14%	12%	2%
West Nilotic							35%	18%	14%	18%	18%	16%	4%
East Nilotic								40%	12%	15%	20%	18%	4%
South Nilotic									20%	17%	14%	12%	8%
Nyimang										18%	14%	12%	2%
Temein											20%	16%	6%
Jebel												12%	4%
Daju													6%

Figure 1. Phylogenetic interpretation of the matrix in Fig. 1[7]



Both the matrix and the tree diagram suggest that, in general, Greenberg's "East Sudanic" is a viable proposition. In the majority of cases, pairwise percentages exceed 10% and sometimes rise as high as 20–25% — for a procedure that relies exclusively on phonetic similarity and inevitably omits a share of true historical cognates, this is a significant number that is very

[7] The tree diagram has been generated by means of the distance-based neighbor-joining method used in the StarLing software, with a glottochronological component (needed as a comparison basis for reconstructed protolanguages of varying time depths); see S. Starostin 2000 on details of the glottochronological method and Kassian 2015 for a more detailed description of the tree-building procedure. Glottochronological dates on the tree in question are only given up to the approximate time depths of all the intermediate reconstructions involved in the comparison; due to the "automated" cognate-finding procedure forming the core of the present analysis, chronological figures beyond the threshold of 3–4 thousand years will most likely be incorrect.

rarely reached under the same conditions by unrelated pairs of languages. Additionally, the results are in agreement with Bender's and Rilly's idea of a primary split into two branches (Bender's "Ek" and "En" and Rilly's "Northeast" and "Southeast" ones, respectively), with Nubian, Nara, and Tama constituting the bulk of the former; only Nyimang, which both researchers decidedly place in the "Ek/Northeast" branch, is grouped closer to Temein on the resulting tree, but this may be a phylogenetic error caused by some unrecognized convergence processes between Temein and Nyimang, an issue to be investigated later on a more thorough etymological basis.

The only glaring candidate for potential exclusion from the East Sudanic inventory is the Kuliak group: these languages consistently show around 4% to 6% resemblances with other East Sudanic branches on the 50-item wordlist — a figure that makes Kuliak as "East Sudanic" in nature as, say, the Hadza isolate (with which Kuliak languages also share 6% of superficial matches), most of which are monoconsonantal and either reflect chance similarities or, perhaps, occasional traces of much deeper relationships that are, at the present stage of analysis, indistinguishable from the former[8].

Nevertheless, in order to be properly convincing, any "working model" constructed by means of preliminary lexicostatistics has to undergo further scrutiny. Even a situation where two or more languages show 20–25% of similarities on the 50-item list may theoretically be interpreted as the result of tense linguistic contact, perhaps multiplied by a few accidental resemblances. From the regular historical-comparative point of view, pure statistics is not enough: the observed and quantified similarities must satisfy our general expectations for a situation of language relationship. In particular, similarities must be organised into patterns of recurrent correspondences — a task that is often impossible to perform based on the limited material of 100, let alone 50 items, so additional material must be considered — and, if possible, additional argumentation must be presented as to why these similarities are more conveniently explained as the results of vertical rather than horizontal transmission, since regular correspondence patterns can be observed between donor and recipient languages just as frequently as between the descendants of a single protolanguage.

The chief goal of the current paper is to investigate one particular node of the preliminary lexicostatistical tree — the hypothetical ancestor of the Nubian, Tama, and Nara languages. Among supporters of the East Sudanic and the broader Nilo-Saharan hypothesis, close relationship of these groups seems to be a given: it is supported by Lionel Bender (2005: 1), who groups these three taxa together into the "Ek" subbranch of East Sudanic (with the further addition of Nyimang), Christopher Ehret (2001: 88–89), who calls this tripartite taxon "Astaboran", and Claude Rilly (2009: 44), who agrees with Bender's classification, renaming his "Ek" subbranch "Northeast Sudanic" (as opposed to "Southeast Sudanic", comprising Surmic, Nilotic, and several other branches). However, a formal demonstration of this relationship based on a general, universally applicable methodology is still lacking, to the extent that some "conservative" encyclopaedic sources do not acknowledge the genetic link between these language groups as established beyond reasonable doubt[9].

[8] Occasional biconsonantal matches can be found as well, but these are almost always scattered and confined to pairwise rather than mutil-lateral matches — cf., for instance, a curious match between Temein and Ik in the word for 'star': Ik *ɗɔ́ɬɛ̀át* = Temein *ɖúlì-t*, pl. *kʊ̄=ɖúl-àʔ* id. Considering that lexical contacts between speakers of Temein, who dwell in the Nuba mountains, and Ugandan Ik people are hardly likely, this phonetic similarity is currently best explained as an accidental resemblance.

[9] Cf.: "No conclusive, methodologically sound basis for assigning Nubian to East Sudanic or to an alleged full or partial Nilo-Saharan has been presented" (Hammarström et al. 2017: http://glottolog.org/resource/languoid/id/nubi1251).

The perfect way to demonstrate this relationship would have been a thorough, methodologically rigorous reconstruction of the phonological inventory of Proto-Nubian-Nara-Tama, supported by a large etymological corpus and based on recurrent phonetic correspondences, along with comparative grammatical evidence. However, even such a demonstration, in order to be easily appreciated by non-specialists in these languages, would still have to distinguish between "core" and "peripheral" layers of evidence, where only the "core" would serve the primary purpose of proving the relationship, whereas the "peripheral" layer (e.g. comparanda drawn from the cultural lexicon, featuring phonetic irregularities or questionable semantic shifts, etc.) would rather serve the purpose of multiplying our alleged knowledge on the already proven common ancestor of Nubian, Nara, and Tama.

Therefore, our intention here is to concentrate on the "core" evidence, extracting it by means of a formal lexicostatistical procedure. The procedure involves:

— demonstrating that a statistically significant number of phonetic homologies is detected between the compared protoforms for Proto-Nubian, Proto-Tama, and Nara equivalents for Swadesh meanings on the 50-item wordlist;

— interpreting these homologies in terms of regular phonetic correspondences, bringing in additional lexical data where necessary or possible;

— detecting additional potential cognates on the same wordlist that have not been identified automatically due to general limitations of the "consonantal class" method, and also interpreting them in terms of regular correspondences, if possible;

— detecting even more additional potential cognates between the compared taxa that involve typologically frequent, "trivial" semantic shifts from a basic Swadesh meaning to a semantically adjacent meaning;

— justifying a *genetic* rather than *areal* interpretation of the attested homologies/regularities by analyzing their distribution across various subdivisions of the 50-item wordlist, from terms that are "more stable on the average" to those that are "less stable on the average".

The data

Complete 100-item Swadesh wordlists have been compiled and annotated for all the languages from the three taxa in question where officially published or archival data were available in sufficient quantity; semantic selection of the optimal equivalents was performed based on the guidelines laid down in Kassian et al. 2010. Reconstruction of the optimal wordlists for Proto-Nubian and Proto-Tama (Nara, having no close relatives of its own, does not require a separate reconstruction, although one might occasionally resort to elements of internal reconstruction) was carried out for the 50-item subdivision of the complete 100-item wordlist; since a very detailed explanation for each of the items has already been published in Starostin 2013, only the least trivial and most significant decisions will be outlined in this paper.

Below we list all the principal data sources and briefly comment on the internal taxonomy of the respective language groups, as well as on previous and current research on the phonological reconstruction of their ancestral states.

A. *Nubian*. Wordlists were compiled for 10 languages belonging to the Nubian group:

(a) Nobiin; primary source — Werner 1987, with Bell 1970 used as an additional control source and Lepsius 1880 consulted for historical purposes. Unfortunately, the large dictionary Khalil 1996 may not be used for lexicostatistical purposes, since it intentionally omits all Arabic borrowings and mixes together data from a variety of old and new sources on different dialects of the language.

(b) Kenuzi-Dongolawi. These two closely related languages (or dialects of a single macrolanguage) are respectively represented by the data in Hofmann 1986 (Kenuzi) and Armbruster 1965 (Dongolawi), with Massenbach 1962 used as a control source for both.

(c) Hill Nubian. This large cluster of relatively small languages, scattered among the Nuba Hills, is represented by wordlists for Dilling (primary source: Kauczor 1920, with Jabr el Dar 2006 used for additional control), Kadaru, Debri (primary source: Thelwall 1978), Karko, and Wali (primary source: Krell 2012). Older data from Carl Meinhof's comparative vocabulary of Nubian languages (Meinhof 1918) have also been consulted for historical purposes, but are unusable as primary sources.

(d) Birgid; primary source — Thelwall 1977, with MacMichael 1920 consulted for control/historical purposes; since this language, constituting a significantly divergent branch of Nubian, has been reported as extinct, every bit of older data on it is extremely valuable.

(e) Midob; primary source — Werber 1993, with Thelwall 1983 consulted for control purposes.

In addition, a wordlist for the Old Nubian language, represented by texts from the 8th – 11th centuries A.D., has also been compiled based on the comprehensive dictionary of Gerald Browne (1996). Although the amount of recovered texts and their lexical content is large enough to permit the use of Old Nubian for lexicostatistical purposes, it has only been possible to fill in 75 out of 100 slots (and a few of these entries remain under serious doubt for various reasons), so any lexicostatistical conclusions on replacement rates between Old Nubian and modern Nubian dialects must be made with caution.

Worse still, although this topic has not been seriously explored so far, there are reasons to suggest that from a lexical perspective, "Old Nubian" is not a concise single dialect, but an amalgamation of several distinct speech varieties: thus, lexical analysis indicates every once in a while the presence of "doublets", in which one word is cognate with its equivalent in modern Kenuzi-Dongolawi and the other one with the equivalent in modern Nobiin (e. g. *ŋul-* vs. *ado-* 'white', or *aman-* vs. *asse- ~ essi-* 'white'). This goes against the general idea of Old Nubian as being specifically the ancestor of modern "Fadidja/Mahas", i. e. Nobiin dialects (Browne 2002: 1), although from a formal statistical perspective, Old Nubian does have more in common with Nobiin than with Kenuzi/Dongolawi, and it makes more sense to assume a number of Kenuzi-Dongolawi interpolations in the Old Nubian corpus rather than to assign Old Nubian to a third separate subbranch of the Nile-Nubian branch (see below for more details on the overall classification of Nubian); this conclusion also agrees with the additional data on the varied nature of Old Nubian texts as adduced in Bechhaus-Gerst 2011: 20–22.

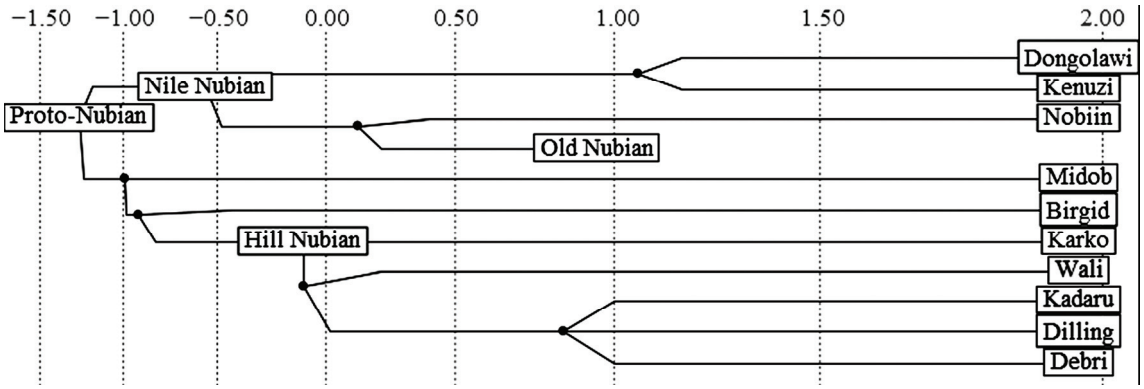
The main principle employed in the construction of a unified wordlist for Old Nubian has been that of statistic frequency. *Hapax legomena* or contextually ambiguous forms were accepted as main entries only in those cases where no other equivalents for the required Swadesh meaning were available. In case of "doublets" where one word is frequently encountered in texts and the other one is basically a *hapax*, only the frequently used word was included in the calculations. Consistent use of this principle showed that the majority of exclusive isoglosses, as a result, is indeed between Old Nubian and Nobiin rather than Old Nubian and Kenuzi-Dongolawi.

Refined lexicostatistical calculations (slightly revised and corrected as compared to the previous analysis in Starostin 2014: 34) yield the following percentage matrix for Nubian (Table 2), which, through the application of Sergei Starostin's revised glottochronological method and the Starling-NJ phylogenetic method (Burlak, Starostin 2005: 162–167; Kassian 2015), may then be converted to the following tree format (Figure 2).

Table 2. Lexicostatistical matrix for Nubian languages (100-item Swadesh wordlists)

	NOB	DNG	KNZ	DIL	KAD	DEB	KRK	WLI	BIR	MID
ONU	0.81	0.63	0.63	0.42	0.43	0.44	0.45	0.42	0.39	0.51
NOB		0.66	0.66	0.40	0.42	0.41	0.41	0.39	0.42	0.51
DNG			0.93	0.59	0.61	0.62	0.55	0.54	0.56	0.57
KNZ				0.60	0.59	0.60	0.55	0.55	0.56	0.57
DIL					0.92	0.91	0.75	0.76	0.64	0.57
KAD						0.92	0.79	0.81	0.60	0.56
DEB							0.80	0.82	0.59	0.57
KRK								0.72	0.56	0.53
WLI									0.59	0.55
BIR										0.56

Figure 2. Phylogenetic tree for Nubian languages (with glottochronological interpretation)



This classification largely agrees with the traditional model as described, e.g., in Bechhaus-Gerst 1985, with a rapid disintegration of Common Nubian into four different branches (Nile-Nubian, Midob, Birgid, and Hill Nubian), but sharply contradicts the later re-classification in Bechhaus-Gerst 1989 and 1996; according to Bechhaus-Gerst, Nobiin should be excluded from Nile-Nubian and positioned as the first branch to split off from Common Nubian, while the increase in lexical and grammatical similarity with Kenuzi-Dongolawi is explained by her as the result of a prolonged period of convergence. This re-classification has been critically scrutinized in Starostin 2014: 93–96, and still more recently in Vasilyev, Starostin 2014, where it was concluded that Nobiin is indeed far more lexically divergent from the rest of Nubian than any other constituent of this group, but that the divergent elements are consistently better interpreted as representing a non-Nubian substrate rather than archaisms inherited from Proto-Nubian[10]; subsequently, the convergence phenomenon must have taken place between Nobiin and some non-Nubian language or languages that used to be spoken to the north of the original Nubian homeland, rather than between Nobiin and Kenuzi-Dongolawi. Results of the analysis convince us that there is no need to dismantle the old Nile-

[10] Precisely the same conclusion has been independently reached by Claude Rilly (2009: 285–288).

Nubian branch, but that there is every reason to treat Nobiin data with caution when it comes to external comparison, particularly if it finds no parallels in other Nubian languages.

The first attempt to establish regular phonetic correspondences between various Nubian languages and set up a Proto-Nubian reconstruction was carried out by Ernst Zyhlarz (1950), but the research was largely inadequate due to lack of sufficient data sources on Hill Nubian, Birgid, and Midob. The first truly significant reconstruction of the Proto-Nubian phonological system, supported by a small etymological vocabulary and still fully relevant today, was carried out by Marianne Bechhaus-Gerst (1985); since then, a somewhat more refined version has been offered by Claude Rilly (2009: 211–288), and additional observations on the complex developments of Proto-Nubian phonology in Hill Nubian languages were made by Angelika Jakobi (2006). The reconstruction system adopted in Starostin 2014 and, consequently, this paper as well, rests largely on the research of Bechhaus-Gerst, but offers a few corrections, for the most part, concerning non-standard consonantal behavior in clusters that appear on morphemic borders; some of these are briefly commented upon below in connection with specific items. In most of the proposed systems, Nile-Nubian languages (and possibly also Birgid) are generally viewed as more phonologically conservative, but data from Hill Nubian and Midob are also essential in order to better assess the distribution of cognates in daughter branches and make more reliable choices for Swadesh meanings on the Proto-Nubian level.

B. *Nara* (= Barea). Nara is typically described as a linguistic isolate, although sources note that the language may be divided in at least two distinct pairs of dialects: Eastern (Higir-Mogoreeb) and Western (Koyta-Saantoorta), with limited mutual intelligibility (Rilly 2005: 1, 2009: 178). Unfortunately, all available sources of significant data concentrate exclusively on Higir as the most widely spoken variety of Nara, which leaves no space for a serious historical reconstruction. The most important of these are Bender 1968, with a 200-item wordlist, and the much earlier descriptive monograph by Leo Reinisch (1874), which also contains a detailed vocabulary. For etymological research, the somewhat later grammatical sketch Thompson 1976 and a few recent works, like Hayward 2000 on the Nara tonal system or Abushush, Hayward 2002 on general phonology, also provide some limited data support.

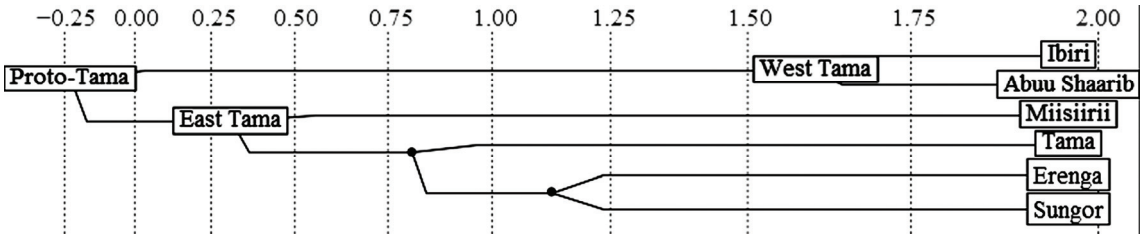
C. *Tama*. Descriptive work on this small, but significantly diversified language group, spoken in Ouaddaï and Dar Fur, has been very scarce so far, with no grammars or dictionaries produced for even a single language. The principal source of data, in fact, remains officially unpublished: it is a comparative vocabulary of all known Tama languages, compiled by John Edgar (1990) from the largest possible variety of sources, including his own field data as well as records stretching all the way back to the late 19th century, and also incorporating data from printed sources such as Lukas 1933 on Ibiri and Lukas 1938 on Sungor. Although made available (by kind courtesy of Roger Blench) in almost print-ready form, the work formally retains the status of a manuscript due to the author's untimely demise; only a few bits of the data appeared in print form, illustrating Edgar's pioneering attempt at a reconstruction of Proto-Tama phonology (Edgar 1991a).

According to Edgar's classification that has also been lexicostatistically confirmed in Starostin 2014, the Tama group is divided into two primary branches: the smaller West Tama cluster, consisting of Ibiri (Mararit) and its satellite dialects such as Abu Sharib, and the larger East Tama cluster, which is itself divided into Miisiirii and Tama-Erenga-Sungor (three closely related dialects). Data for all five varieties, collected in Edgar 1990, are sufficient to construct near-complete Swadesh wordlists that yield the following cognacy matrix (Table 3; also slightly revised as compared to the previous analysis in Starostin 2014: 317), and the following phylogenetic tree (Fig. 3; also constructed by means of the Starling-NJ method).

Table 3. Lexicostatistical matrix for Tama languages (100-item Swadesh wordlists)

	ERE	SUN	MIS	IBI	ASH
TAM	0.89	0.91	0.80	0.69	0.71
ERE		0.94	0.85	0.69	0.69
SUN			0.85	0.70	0.68
MIS				0.70	0.67
IBI					0.99

Figure 3. Phylogenetic tree for Tama languages (with glottochronological interpretation)



It is important to note that Tama gives the (glottochronologically confirmed) impression of a less chronologically deep family than Nubian; consecutively, its 50-item proto-wordlist is easier to reconstruct due to fewer lexical replacements in the principal branches. Nevertheless, some of the languages have still gone through significant phonetic change, not all of which is easy to trace and reliably reconstruct due to limited (and not always accurately transcribed) amounts of data. Our reconstruction of Proto-Tama depends significantly on the rules laid down in Edgar 1991a, with some additions and corrections offered in Starostin 2014: 314–316.

Comparative 50-item wordlists for Proto-Nubian, Nara, and Proto-Tama.

Preliminary notes. Table 4 below does not list the complete data (freely available at the website of the Global Lexicostatistical Database), but only the reconstructed optimal candidates for 50 out of 100 semantically fixed "Swadesh slots" (detailed explanation of semantics for each slot may be found in Kassian et al. 2010) for Proto-Nubian and Proto-Tama; Nara is represented by Higir dialect data from Bender 1968. Numeric indexes that follow individual items reflect their average "stability index" as per Starostin 2010: 113 (ultimately based on the calculations across various genetic lineages in Eurasia, Africa, and Australia as per S. Starostin 2007).

Detailed justifications for all the reconstructions may be found in Starostin 2014; in this paper, due to volume considerations, notes on particular reconstructions will be condensed and restricted to non-trivial cases of phonetic or semantic developments, while the majority of the notes section will concentrate on the justification of etymological matches between PN, PT, and Nara.

We use the following notation symbols to designate various degrees of cognacy estimation:

! — marks pairs or triplets of reconstructions whose simplified phonetic shapes ("consonantal skeletons") match each other according to the Dolgopolsky consonantal class criterion. In cases where two or more reconstructions are more or less equiprobable for one taxon (either because there is no certainty about the phonetic interpretation of a given proto-etymon, or because two different etyma are represented in two primary branches of the family), in the table below we only list the variant that is compatible with potential external cognates.

+ — marks pairs or triplets of reconstructions that represent highly probable etymological cognates. Although at this point, despite the works of M. L. Bender and C. Rilly, it is probably too early to talk about a definitive set of regular phonetic correspondences for East Sudanic as a whole or Northeast Sudanic (Nubian-Nara-Tama) in particular, we provisionally mark the forms as cognate with each other if the consonantal correspondences between them are trivial (i.e. the consonants are exactly the same) *or* may be shown to form a part of a recurrent pattern (e.g. Proto-Nubian **n-* = Proto-Tama **l-*) *or* may be explained as the result of morphophonological or morphological processes. Precise vocalic correspondences are not expected, but the base root vowels should have a certain degree of proximity, i.e. a match between labial vowels **o* and **u* is acceptable, while a match between **a* and **i* is suspicious. Predictably, there will be a serious correlation rate between "automated" and "etymological" cognates, but not a 100% one (see 'drink', 'egg', etc.).

[] — square brackets mark items that have neither "automated" nor "etymological" parallels in any of the other two groups.

ˀ — this special symbol is typically inserted after the initial vowels of VCVC-type stems, typically encountered in Proto-Nubian, more rarely in Nara, and almost never in Proto-Tama. Since the most common type of root structure for all these languages is CVC, this initial vowel, often identical in quality to the main root vowel (cf. in Proto-Nubian: **ubur-* 'ashes', **awar-* 'night', etc.; there are, however, exceptions such as **agul-* 'mouth', etc.), may be suspected of representing an old fossilized prefix, perhaps the trace of one or more older classifiers or determinants, which justifies its formal deletion in the procedure of external comparison. Alternately, this vowel may have been an integral part of the original root, in which case it would be possible to regard the Proto-Nubian system as more archaic in comparison with Nara and Tama, where it became lost due to purely phonetic processes.

Table 4. 50-item wordlist entries for Proto-Nubian, Nara, and Proto-Tama.

#	Word	Proto-Nubian	Nara	Proto-Tama
1	'ashes'$_{38}$	*uˀbur-ti +	hɪbɪd ?	*or-ŋo +
2	'bird'$_{33}$	*kawir- +!	karba +!	[*wig-]
3	'black'$_{48}$	[*uˀdum-]	[sur-ku]	[*kidi-]
4	'blood'$_{20}$	[*əˀger]	[kito]	[*ya-i]
5	'bone'$_{34}$	*kəsi-di +	ketti +	*ki-(ŋa)-ti +
6	'claw/nail'$_{19}$	*suŋ-di ?	ši	*ŋosa- ?
7	'die'$_{13}$	*di:- +!	di:- +!	[*iye] (← Maba?)
8	'dog'$_{16}$	*bəl ?	wəs +!	*wes-i +!
9	'drink'$_{15}$	*ni:- +	li:- +!	*li- +!
10	'dry'$_{24}$	[*sow-]	[dɪsɛ-]	[*lab-]
11	'ear'$_{32}$	*ulgi ?	tus ?	*(ŋ=)us ?
12	'eat'$_{25}$	*kɔl- +!	kʌl- +!	[*ŋan-]
13	'egg'$_{47}$	*kumbu +	[wari]	*kob- +
14	'eye'$_{4}$	*miɲ +!	[no]	*eˀmeɲ- +!
15	'fire'$_{7}$	*usi-gi +!	ši-ta ?	*us-g +!
16	'foot'$_{43}$	[*oy]	[bəla]	[*war]
17	'hair'$_{27}$	[*dɛl-]	[sɛbi]	[*isigi-t]
18	'hand'$_{11}$	*ə-si +	a:(-)t +	*aw-g +

#	Word	Proto-Nubian	Nara	Proto-Tama
19	'head'$_{49}$	*or +!	[kela]	*ur +!
20	'hear'$_{45}$	*giʒ- ?	[wos /Rn./]	*sig- ?
21	'heart'$_{14}$	[*ay-]	aˀsım-a +!	*samil +!
22	'horn'$_{44}$	*ŋəʒi +	[keli]	*ŋawi-ti +
23	'I'$_{3}$	*ə-y +!	a-g +!	[*wa]
24	'kill'$_{42}$	[*pay-]	siː- +!	*siy- +
25	'leaf'$_{41}$	[*ulgi]	[tifini]	[*afol]
26	'louse'$_{17}$	[*i/ɲ/-ti]	ši-ti +	*sin- +
27	'meat'$_{46}$	[*kosi]	[nɔː-]	[*is-]
28	'moon'$_{18}$	[*ɲun-]	[feːta]	[*ayi-]
29	'mouth'$_{31}$	*aˀgul +	aˀwɔlo +	*kul +
30	'name'$_{10}$	[*əˀri]	aːda +!	*at +!
31	'new'$_{23}$	[*ɛːr]	[wɛr- ~ wɔr-]	[*suw-]
32	'night'$_{50}$	*aˀwar +!	[kıš-]	*war +!
33	'nose'$_{29}$	[*esi-ŋ(i)]	[dəmmo]	[*miʒi]
34a	'not'$_{30}$	*m- +!	ma= +!	*m= +!
34b	'not'$_{30}$	[*=a-]	[ka=]	[*=to]
35	'one'$_{21}$	[*bɛy-]	[doku]	[*ku- ~ *ka-]
36	'rain'$_{39}$	*ar- +!	[hala]	*ar- +!
37	'smoke'$_{36}$	[*gume-]	[aˀsuru]	[*turu-]
38	'star'$_{40}$	*waɲe +!	wiːni +!	*miɲ- +
39	'stone'$_{9}$	[*kul-]	[taːna]	[*kad-]
40	'sun'$_{35}$	[*maša-]	[koːs]	[*ari]
41	'tail'$_{26}$	[*ɛːb]	[dawa]	[*gawu-t]
42	'thou'$_{5}$	*e- ~ *i- +!	ı-ŋa +!	*i- +!
43	'tongue'$_{8}$	*ɲalT- +	[haga]	*laɲa-t +
44	'tooth'$_{22}$	[*ɲəl-]	nihi +	*ŋeʒ- +
45	'tree'$_{37}$	[*pər]	[*kel]	[*gaːn]
46	'two'$_{2}$	*awri +!	ari +!	*wari +!
47	'water'$_{28}$	[*əs-]	[mba]	[*kaːl]
48	'we'$_{1}$	*a-y +!	a-gga +!	[*wa-i]
49	'what'$_{12}$	*nwa- ~ *nwi- +	[nda-]	*num +
50	'who'$_{6}$	[*ŋə-y]	na- +!	*na +!

Comments on individual entries.

1. '**Ashes**'. PN *$u^{ˀ}$bur-ti* (Nob. *ùbúr-tí*, Dng. *ubˈur-ti*, Knz. *ubur-ti*; Dil. *ɔp-te*, Kad., Deb. *ɔt-tɛ*, Krk. *ōmì-t*; Bir. *ubur-ti*; Mid. *úfù-dì*) = PT **or-ŋo* (Ere. *orɔŋo*, Sun. *orŋo ~ oruŋo*, Mis. *ʌrŋo*).

The element *-*ŋ*- in the PT form is easily analyzable as a fossilized plural/collective suffix (the same morpheme is frequently found as a productive pluralizer as well). Root morpheme **or*- is derivable through lenition and contraction from an earlier **owur*- ← **obur*-; for similar cases of possible development of labial **b* before labial vowels cf., e.g., PN *$u^{ˀ}$bur* 'hole' = PT **war*- ~ **wor*- id. (although here PT probably reflects a variant without the prefixal vowel).

Nara *hıbıd* (Bd.), *hübet* (R.) could also belong here, provided the *h-* is prothetic and the word-medial cluster has been simplified (**ubur-ti* → **hubir-ti* → **hubit*); however, this is a complicated scenario that needs additional evidence, so we cannot count this as a bona fide match.

2. '**Bird**'. PN **kawar-ti* (ON *kawar-t-*; Nob. *kawar-ti*, Dng. *kˈawır-tɛ ~ kˈauır-tɛ ~ kˈaur-tɛ*, Knz. *kawir-te*; Dil. *kɔmil-ti*, Krk. *kùbùr-àn*; Bir. *kwar-ti*; Mid. *à:béd-dí*) = Nara *kʌrba* (Bd.), *karba* (R.).

Phonetically compatible under the assumption of a metathesis in Nara (**kaw^{a}r-* → **karb-*), which seems typologically plausible and finds no contradictory evidence.

PT **wig-* 'bird' (Tama *wìgí-t*, Ibi. *wígì-t*, etc.) is incompatible with these forms and finds no obvious parallels in either PN or Nara.

3. '**Black**'. No parallels detected between any of the three taxa.

4. '**Blood**'. No parallels detected between any of the three taxa.

5. '**Bone**'. PN **kəsi-di* (ON *gis-ri-*; Nob. *gìsì-r*, Dng. *kıhˈı:-d*, Knz. *ki:-d*; Kad. *kwe-dɛ*, Deb. *kwe-du*, Krk. *kwīè-dà*, Wal. *kwǐ-tù*; Bir. *kìzí-dì*; Mid. *à:-dí*) = Nara *kə-ti* (Bd.), *ke-tti* (R.) = PT **ki-(ŋa)-ti* (Tama *kí-tí*, Ere. *kìŋà:-tí*, Mis. *kıŋı-t*, Ibi. *kìŋí-t*).

In Starostin 2014: 320 it was suggested that the PT paradigm should be reconstructed as **kiŋa-ti* (sg.), **kiŋa-k* (pl.), with vowel reduction and cluster simplification in Tama proper: **kiŋa-ti* → **kiŋ-ti* → *ki-ti*. However, since then I have found no corroborative evidence for the latter development; and considering the relative frequency of *-ŋ-* as a plural marker in Tama languages, it is perfectly plausible to reinterpret this as sg. **ki-ti*, pl. **ki-ŋa*, with subsequent generalization of the plural form in most Tama languages and reformation of the entire paradigm based on it (with new singulative **ki-ŋa-ti* and new plural **ki-ŋa-k*).

This interpretation is in good agreement with Nara data, suggesting a common Tama-Nara root **ki-* or **kə-*. The parallel with Nubian is slightly more problematic, but intervocalic **-s-* on the whole is a fairly unstable consonant in this entire region (cf. lenition and elision in Kenuzi-Dongolawi for this very root, or the regular deletion of **-s-* in East Tama languages), so the assumption of a regular development **kəsi-ti* → **kə-ti* ~ **ki-ti* in Tama and Nara, even without additional evidence for the moment, seems fairly realistic. In any case, at least the Nara-Tama isogloss is unquestionable.

6. '**Claw/nail**'. PN **suŋ-di* (Nob. *sun-ti*, Dng. *sun-tı*, Bir. *suŋ-di*, etc.) and PT **ŋosa-* (Mis. *ŋɔsʌ-t*, Sun. *ŋisi-t*, etc.; see Starostin 2013: 320–322 for a detailed discussion on the complicated fate of this etymon due to its contamination with 'tooth' in the individual languages) may actually be relatable to each other through metathesis, although it is impossible to say which form should be thought of as representing the original consonantal sequence. However, since this kind of metathesis would have to be qualified as an incidental irregularity, it is difficult to count this parallel as a primary piece of etymological or lexicostatistical evidence for the Nubian-Tama relationship.

7. '**Die**'. PN **di:-* (ON *di:-*, Knz.-Dng. *di:-*, Mid. *tí:-*, etc.) is a perfect match with Nara *di:-*. No sign of this root appears in Tama, and, in fact, Tama **iye* is one of the few entries on the list which, instead, shows close phonetic proximity to Maba languages, cf. Masalit *ıy*, Kibet *iy*, Kodoi *yî:*, Maba *ý́y* 'to die' (Edgar 1991b: 391). Borrowing from Maba is not the only possibility (similar forms are also found on some proto-levels in other East Sudanic languages, e. g. East Nilotic **=yɛ-* 'to die'), but, in any case, it is impossible to relate the Tama equivalent to Nara and/or Nubian.

8. '**Dog**'. Nara *wəs* (Rn.: *wos*) is clearly the same as Tama **wes-i* (Ib. *wí:sì*, AS *wis*, Mis. *wus*; Tama *wèí*, Ere. *wi*, Sun. *wɛ:* with regular deletion of intervocalic **-s-*). On the possibility of Nubian **bəl* (Dng. *wɛl*, Dil. *bol*, Bir. *mɛl*, Mid. *pà:l*, etc.) being related to Tama and Nara through a non-trivial consonantal correspondence see below ('ear'); at present, however, we prefer to keep these etyma apart.

9. '**Drink**'. The obvious parallel is between Nara *liː-* and Proto-Tama **li/y/-* (Tama *liː*, Ere. *lɪʒ-ɛ*, Sun. *liy-ɛ*, Mis. *liy-ei*, AS *li*, etc.). However, both forms also regularly correspond to Proto-Nubian **ni-* (ON *ŋi-*, Dng. *niː*, Dil. *di*, Bir. *ɲiː*, Mid. *tìː-*, etc.): Proto-Nubian has no word-initial **l-*, which makes the assumption of regular development **n- → *l-* perfectly plausible, and furthermore, the correspondence may be strengthened by additional examples, even from the basic lexicon (e. g. Proto-Tama **lasi-* 'long' = Proto-Nubian **nas-* 'long').

10. '**Dry**'. No parallels. This is not a stable item in either Nubian or Tama (most subbranches have their own replacements, and precise reconstruction is very difficult).

11. '**Ear**'. This is a complicated case where additional progress might be made in subsequent etymological studies of the Nubian-Nara-Tama family.

In Tama, the root is **us-* (Tama sg. *ú-tù*, pl. *ú-ŋ-òŋ*, Ere. sg. *us-ut*, pl. *us-ɔŋ*, Mis. sg. *us-ut*, pl. *us-ɔŋ*), but in West Tama, it is preceded by a fossilized prefixal element *ŋ=*: Ibi. *ŋùs-í*, AS *ŋgus-i*. The function of this prefix remains obscure, yet its segmentable status is corroborated quite firmly by additional examples (see 'head' and 'name' below). The ability of the root **us-* to combine with a fossilized prefix reasonably begs the question of whether a *different* fossilized prefix (with an equally obscure function) could not be present in Nara *t(=)us* 'ear'. However, unlike Tama, in Nara no additional evidence has been found so far to suggest the idea of a formerly segmentable *t=*; in fact, the only other basic lexicon term with initial *t-* that shows credible outside parallels is Nara *tawa* 'belly' = Proto-Nubian **tu* id., without any signs of segmentation. Therefore, this comparison remains highly questionable and unfit as primary evidence for relationship.

A different problem is tied to Proto-Nubian **ulgi* 'ear' (ON *ulg-*, Dng. *ulug*, Nob. *úkkí*, Mid. *úlgí*, etc.). If we assume that the second syllable is of suffixal origin, the allegedly original root **ul-* would correlate with Tama **us-* precisely the same way that Proto-Nubian **bəl* 'dog' (see above) correlates with Nara *wəs*, Tama **wes-i* — suggesting a non-trivial correspondence «Proto-Nubian **l* : Proto-Tama **s* : Nara *s*» whose most logical phonetic interpretation would be a lateral fricative (**ɬ*). This idea seems worthy of further exploration, but for the moment, no further examples of this correspondence are available, and we cannot qualify either of these parallels as primary evidence.

12. '**Eat**'. PN **kɔl-* (Dng. *kal*, Dil. *kol*, Mid. *ɔ̀l-*, etc.) is perfectly compatible with Nara *kʌl-* (Rn.: *kal-*). The Tama paradigm is completely different: East Tama **ŋan-* is opposed to West Tama suppletive forms: imperfective **gey-* vs. perfective **sin-*. None of the three forms has anything to do with the verb in Nara or Tama.

13. **'Egg'**. Nile-Nubian **kumbu* (ON *kumpu-*, Dng. *kumbu*, Nob. *kúmbúː*), one of several equiprobable candidates for PN 'egg', is comparable with West Tama **kob-* (Ibiri *kób-ìt*, AS *koːb-it*), assuming cluster simplification in the latter (nasal cluster *-mb-* does not seem to be encountered in inherited lexicon in these languages).

14. **'Eye'**. Some of the phonetic shapes in Nubian and Tama languages are almost completely identical, cf. Dng. *mɪssɪ* vs. AS *meše* (Barth), etc. However, detailed etymological analysis of the complete datasets, as presented in Starostin 2014: 50–51 (for Nubian) and 328–329 (for Tama), shows that in both of these groups, there is serious evidence for reconstructing a "weak" palatal nasal in root-final position, prone to elision or assimilation — but still preserved in some Nubian languages (ON *maɲ-*, Nob. *máːɲ*) and, in assimilated form, in such relic plural forms as Ibiri *ímn-íèn* ← **e=meɲ-oŋ*. This means basic compatibility for the reconstructed variants as well, allowing us to posit **miɲ-* ~ **meɲ-* as the optimal equivalent for 'eye' on the proto-level.

In comparison, Nara *no* 'eye' shows no affinity with these forms, but it makes sense to compare the Nubian and Tama items with the Nara verb *minni* 'to flash, shine' (Rn.): if the

etymology is correct, we could be dealing with a Nubian-Tama shared innovation ('to shine' → 'eye') vs. a possibly retained archaism in Nara.

15. **'Fire'**. Here we have a transparent isogloss between Nubian **usi-gi* (cf. especially Bir. *uzug* and Mid. *ússí*; such forms as Old Nubian *eig-*, Nobiin *íːg*, etc., probably represent contractions of the original stem) and Tama **us-g* (Ibiri *ùsùg-í*, AS *usugu*; Tama *ú*, Ere. *ú*, etc., are also contracted variants, with regular deletion of intervocalic **-s-* in these languages). It is unclear if Nara *šita* 'fire' also belongs here, but it is possible: *-ta* may be identified as a fossilized plural suffix (cf. Nara *nɔ́ː-ta* 'meat' [Bd.] vs. the earlier recorded *no* [Rn.]), and the word-initial vowel could be syncopated in a trisyllabic structure (unless it was a detachable prefix from the very beginning). However, both of these assumptions remain rather speculative.

16. **'Foot'**. No parallels detected between any of the three taxa.

17. **'Hair'**. There is distinct phonetic similarity between Tama **isigi-* (Tama *ìgí-t*, Ere. *sigi*, Ibiri *ìsìŋí-t*, AS *isiŋi-t*) and Nobiin *šìgír-tí* 'hair'. However, the latter, even if it is related (with irregular deletion of the word-medial consonant) to Knz. *siːr* 'hair', is far from the optimal carrier of the basic meaning 'hair' in Proto-Nubian. Additionally, its phonetic proximity to various Semitic and Cushitic terms for 'hair' (e. g. Arabic *šaʕr-*, Ethiosemitic **ṣag^{w}ar*, etc.) makes all these items highly questionable as potential genetic markers, so we would not want to consider them as primary evidence.

18. **'Hand'**. All compared forms may be regarded as cognates, although phonetic similarity between them is obscured by the tendency of the original short root to get fused with various suffixes, formerly (or, sometimes, still productively) denoting singulative or plural semantics.

For PN, Rilly (2009: 477) reconstructs **es-i* 'hand', which almost coincides with **əsi* in Starostin 2014: 54; this form is either preserved with minimal phonetic change (Dil. *iši*, Mid. *ə̀ssì*), or is subject to regular weakening and deletion of intervocalic **-s-* (Dng. *ɪː*), or becomes further extended with an additional singulative marker (Nob. *èddì* ← **əsi-ti*). For PT, the suggested reconstruction is **awg* (Starostin 2014: 332), which seems to explain the wide variety of reflexes (Tama *àù*, Ere. *auw ~ oɣ*, Sun. *ao*, Mis. *wiː*, Ibi. *wèí*) somewhat better than Rilly's **(a)wei* (2009: 477), although ultimately the basic consonantal shape of the reconstruction is the same in both cases, since we regard **aw-g* as a transitional fusion of the original root **aw(i)-* with a former plural marker.

All three forms, including Nara *aːt*, can be rather unproblematically traced back to an original root **ay-*, or, perhaps, a bisyllabic stem **ayi*, with the Proto-NNT paradigm **ayi-ti* (sg.) : **ayi-k-* (pl.) conforming to the very common so-called "T/K pattern" of East Sudanic (Bryan 1959). As both forms underwent contraction and fusion in daughter branches, only the first one survived in Nara (**ayi-ti* → *aːt*) and in Nubian, where assimilation with the fricative **-y-* resulted in fricativization of the old stop (**ayi-ti* → **əyti* → **ə(s)si*); PT, on the other hand, generalized the plural form, and, in addition, underwent a dissimilative process: **ayi-k-* → **awi-k-* → **awg-*. This dissimilation is precisely the same as in the case of 'horn' (see below) and may be considered regular.

Although short monoconsonantal stems beset with idiosyncratic issues of morphological fusion could be regarded as questionable evidence for genetic relationship, in this particular case it is worth noting that the word 'hand' also displays very similar patterns of behaviour in other potential East Sudanic languages as well; cf., for instance, the situation in East Nilotic, where the old root **k=ay-* (extended by means of the common nominal prefix *k=*) is still occasionally encountered as a segmentable unit (e. g. sg. *n=kʌ́í-nʌ́*, pl. *n=kàí-k* in Camus), but generally tends to fuse, once and for all, with the old singular marker *-n* (e. g. sg. *ʌ́=kàn*, pl. *ŋʌ́=kán* in Turkana, etc.; see Vossen 1982: 326 for more data). Similar situations are attested in Surmic,

Daju, Nyimang, and Temein: all these groups share the common invariant **a(y)-C-* ~ **e-C-* for the meaning 'hand', where *-C-* is sometimes fused with the old root and sometimes remains as a productive number marker. These external parallels should certainly raise the level of confidence in the correctness of this Nubian-Nara-Tama etymology.

19. **'Head'**. Here we have a clear correlation between Nubian **or* (→ ON *ur-*, Nob. *ùr*, Dng. *ur*, Mid. *òr*, etc.) and Tama **ur* (→ Ibiri *úr-ì*, AS *ur*; other Tama languages show an initial *ŋ=* which must be some sort of fossilized, possibly pronominal or deictic, prefix — Tama *ŋúr*, Mis. *ŋɔr*, etc.; the same prefix is also encountered in 'name', see below).

Nara *kela* certainly does not belong here, but has a phonetically perfect and semantically acceptable parallel in Nile Nubian **kel-* 'end, border, tip' → ON *kel-*, Dng. *kɛ́ːl*, Knz. *keːl*, suggesting a semantic shift in Nara ('tip, end' → 'head') with loss of the original root.

20. **'Hear'**. PN is reconstructed based on an isogloss between Kenuzi-Dongolawi **giʒ-* and such Hill Nubian forms as Dilling *ki-er-* (←**giʒ-* with regular devoicing of initial velar and possibly regular loss of intervocalic **-ʒ-*, although this has not been properly confirmed yet). Phonetic similarity of this stem with PT **sig-* (Tama *ik-*, Sun. *ig-*, Mis. *sug-o*) is observable, but the two could be related only under the assumption of a spontaneous metathesis (cf. a similar possible metathesis between a velar and an alveolar consonant, but with reverse direction, in the case of 'nail'), therefore, we should not accept this evidence as primary.

The Nara equivalent is incompatible, but if initial *w-* is prothetic, the verb *wos-* may actually contain the same root as Tama **ŋ=us* 'ear' (and even *tus* 'ear' in Nara itself, see notes on 11 'ear' above). If so, this would be the same type of development as in Old Nubian *ulg-ir-* 'to hear', Nobiin *úkké-èr* id., a verbalization of PN **ulgi* 'ear'.

21. **'Heart'**. We reconstruct the PT form as **samil* based on Mis. *samil* and forms with regular deletion of **s-* in East Tama (Tama *àmúl*, Ere. *ʌmɔl*, Sun. *amul*). Since triconsonantal roots in East Sudanic languages are a rarity, it is plausible to assume that **-(i)l* here is a fossilized suffix, same as the one that also occurs in some other nominal stems (e. g. Tama *tó-l-òl* 'belly' = Sun. *to-l* id., further perhaps to PN **tu* 'belly' without this marker) and possibly of the same origin as the Common Nubian determinant **-l*. This allows easy comparison with Nara *asɪma*, at least as far as the basic consonantal skeletons are concerned. Some Nubian forms also show a stem with a fossilized determinant (PN **ay-il-* → ON *ai-l-*, Dil. *a-l-du*, etc.), but the root proper is **ay-* (→ Nob. *áy*, Bir. *ai-di*, etc.), not comparable with Nara and PT.

22. **'Horn'**. PN **ŋəʒi* → Nob. *nìːšì*, Dng. *nɪššɪ*, Dil. *dɔ-ti* (regular development **ŋ-* → **n-* → *d-* and probably the same regular deletion of **-ʒ-* as in 'hear' q.v., with a new productive marker added), Bir. *ŋis-ti*, Mid. *kə́ːɕí*. We may plausibly interpret the form **ŋəʒi* as a contraction from an older **ŋəy-ti*, the same way that **əsi* 'hand' is contracted from **əy-ti* (the only difference being that this time around, the initial voiced consonant caused the word-medial consonant to become voiced as well).

PT **ŋawi-ti* is reconstructed based on Tama *ŋó-d* (pl. *ŋó-n*), Ere. *ŋɛ-tɪ*, Sun. *ŋoː-tu*; with the same dissimilation as in 'hand' (**ayi-* → **awi-*, **ŋayi-* → **ŋawi-*), the original root turns out to be plausibly compatible with pre-PN **ŋəy-ti*. Nara *keli* obviously does not belong here and is probably connected instead, through some old suffixal pattern, with *kela* 'head, top' q.v.

23. **'I'**. The basic form of the 1st p. sg. pronoun in most East Sudanic languages is **a-*, usually extended with the suffixal component **-n-* for the Southern groups and with **-k-* for the Northern groups (cf. Bender's division into "En" and "Ek" languages), although some variation does occur. The original variant is most clearly seen in Nara *a-g*; for Nubian **ə-y* (Mid. *ə́y*, ON *ay*, etc.) it is necessary to assume lenition of the velar stop, but the old root without the nominative singular marker is still preserved in some paradigmatic forms (e. g. Mid. accusative *ə́* 'me', etc.).

The biggest puzzle in this arrangement is PT **wa* 'I', reflected as such in most of the modern dialects. The appearance of an unpredictable *w-*, impossible to explain away as a prothetic development or an enigmatic emphatic particle, makes the base pronominal paradigm of Tama incompatible on the whole with Nubian, Nara, and East Sudanic in general. On the other hand, even if one considers the typologically rare scenario of a borrowed origin for a basic personal pronoun, the fact remains that no modern areal neighbors of Tama have anything even remotely close to a *w*-shaped equivalent for 'I' — the closest would probably be Kanuri *wu*, but since there are no other reliable Tama-Kanuri or Tama-Saharan parallels in the basic lexicon, it is preferable to treat this phonetic similarity as coincidental.

One possible explanation comes from a comparison of this form with the paradigmatic peculiarities of the 1st p. pronoun in Hill Nubian, where it frequently takes on a labialized shape in the indirect stem (cf. Tagle *ì:* 'I', gen. *ɔ̀-nná*, Dilling *e*, gen. *ɔ-nɛ*, etc.) and in Nara, where *a-g* 'I' is opposed to the genitive/dative stem *(w)o*. In light of this evidence, Claude Rilly has proposed to reconstruct a direct stem **a-(i)* and an indirect stem **o-* for Proto-Northeast Sudanic (Rilly 2009: 467), with analogical levelling in Proto-Tama (where languages such as Ibiri also show a separate genitive form *ho-n*). This does not quite explain why the nominative stem is *wa* and not the expected **o*, but the presence of these labialised indirect forms in Nubian and Nara is hardly accidental.

24. '**Kill**'. Nara *si-* and PT **siy-* (→ Mis. *siy-ɔ*, Ere. *šɪ-o*; Ibi. *ey*, Tama *ìy-ɛ́* with regular loss of word-initial *s-*) present a perfect match. PN **pay-* (→ Nob. *fá:y-*, Mid. *pé-*, etc.) is not related and finds no clear correlates in the other two groups.

25. '**Leaf**'. Excluded from comparison. Most of the attested equivalents are either derived from the word for 'ear' (a very common typological development for the entire area) or are of obscure origin.

26. '**Louse**'. Nara *ši-ti* 'louse' has precisely the same phonetic shape as Ere., Sun. *ši-ti* id. (cf. also Tama *í-tì* with regular deletion of initial **s-*), although for PT, the original root shape has to be reconstructed as **sin-* based on Mis. *šin-ti* (the plural form is simply *šin*; special marking of the sg. rather than pl. number for this item is hardly surprising). Cluster simplification in Nara (**sin-ti* → *ši/t/ti*) is neither confirmed nor contradicted by additional examples, but is typologically plausible.

It is tempting to find some connection between these forms and PN **iti-di* 'louse' → Nob., Knz. *issi*, Dng. *ɪssɪ*, Dil. *iti-d*, Mid. *ì:dì*, where **-di* is a relatively recent marker of the singulative, common in Nubian nominal stems. Theoretically, the remaining root **iti-* itself may be an old contraction from **iŋ-ti*, but there is no evidence that the initial sibilant could be deleted in PN just as it was (regularly) deleted in Tama; therefore, at the present stage the exact phonetic resemblance between such forms as Dilling *iti-* and Tama *ítì* should rather be deemed a coincidence.

27. **'Meat'**. No parallels detected between any of the three taxa.

28. **'Moon'**. No parallels detected between any of the three taxa.

29. '**Mouth**'. The PN form is reconstructed rather securely as **agul-* (Knz., Dng. *agil*, Dil. *ɔgul*, Bir. *agal*, Mid. *á:l* with contraction; Nob. *ág* with seemingly regular deletion of stem-final **-l*). In Tama, the situation is more complicated: here, Eastern **kul* (Tama *kùl*, Ere., Sun., Mis. *kul*) seems poorly compatible with such Western forms as Ibi. *úlì ~ awal*, AS *o:l ~ awl*. However, in Starostin 2014: 345 it was argued that both variants may still be reconciled under the assumption of two morphological variants in PT — simple **kul-* and its prefixal counterpart **V=kul-*, only the latter of which was preserved in the Western branch (with vocalic reduction and consonantal lenition: **V=kul-* → **awl-*). This solution remains hard to prove, but is nevertheless realistic (monovocalic fossilized prefixes were at least as likely to exist in PT as they were in PN), and makes the final reconstruction even more compatible with Nubian

data, since PT **V=kul-* may indeed have contained the very same prefix that is also preserved in PN **aˀgul-*.

For Nara, it is essential to pay attention to the dialectal forms listed in Rilly 2009: 178: Higir *awlo*, Mogoreeb *àlkò*, Koyta *aulo*, Saantoorta *agura*. Although we do not have enough dialectal evidence to confirm this as part of a regular pattern, the only plausible way to explain the divergence is to set up the protoform **agulo* or **agula*, best preserved in Saantoorta (with a presumably regular *l* → *r* development) but undergoing reduction → **aglo* in the other dialects, with a subsequent metathesis in Mogoreeb and lenition → **aɣlo* → **awlo* in the other two dialects. This makes the form perfectly compatible with Nubian and Tama data.

30. **'Name'**. Nara *a:d-a* is obviously compatible with Proto-West Tama **a:t* (Ibi. *á:t*, AS *a:t*); Proto-East Tama **ŋa:t* (Ere., Mis. *ŋa:t*, Tama *ŋát*, Sun. *ŋat*) probably belongs here as well, provided that initial *ŋ=* may be viewed here as the same fossilized prefix that was already encountered above in 'head'.

Rilly (2009: 486) suggests that both of these items are further compatible with PN **aˀri*, but there are too many unresolved problems with this comparison: even if the common Nara-Tama root is to be reconstructed as **a:d-*, there is no strong evidence that PNNT **-d-* could yield PN **-r-* in intervocalic position. Provisionally, we treat these etyma as different items.

31. **'New'**. No convincing parallels. Phonetic similarity is detected between PN **ɛ:r* (Knz. *e:r*, Dng. *ɛ:r*, Dil. *er*, Bir. *e:r*, etc.; replaced by a substrate element in ON *miri-*, Nob. *míríː*) and Nara *wor-ko* (Rn.), *wɔr-ku* (Bd.; also listed as *wɛr-* with a front vowel in Bender 1971: 268), but even if Nara *w-* is prothetic (of which there is no certainty), the significant difference in vocalism quality remains unexplained, so we provisionally reject this pair as a potential etymological match.

32. **'Night'**. A transparent isogloss between PN **aˀwar* (ON *oyar-*, Nob. *áwá*; Mid. *òːd*; replaced in other branches by different innovations) and PT **war* (Tama *wàr*, Ere. *wa:r*, Sun. *war-dɛ*, Mis. *war*). In Nara, the old word was replaced by *kiše ~ kis-ne* (Rn.), *kɪši-ŋa* (Bd.), bearing some resemblance to West Tama forms: Ibi. *íšè*, AS *i:še*. The latter, however, are transparent borrowings from nearby Maba (*íšè* 'night'), and since Maba-Nara contacts are geographically impossible, it is probably better to interpret the partial Nara – West Tama similarity as due to chance.

33. **'Nose'**. No parallels detected between any of the three taxa.

34. **'Not'**. All three taxa present evidence for at least two different morphemes that could mark indicative negation on the proto-level, but only one of them is compatible: PN **m-*, functioning as part of the negative verbal stem **mun-* ~ **min-* 'not (to be)' in Nile-Nubian and Hill Nubian and as a negative suffix in Birgid = Nara *ma* (negation marker in perfective forms) = PT **m-* (basic negative prefix in West Tama, also encountered as a prefix in certain adjectival stems in East Tama, cf. Sun. *áŋgé* 'a lot' vs. *m=aŋge* 'a little'). The others are different in all three taxa — PN monovocalic suffix **=a-*, fully preserved only in Midob but looking quite archaic in nature; Nara *ka* (negation marker in imperfective forms); and West Tama suffix **-to*. It is worth noting, however, that out of all East Sudanic languages, the only other family that shows signs of a proto-level **m-*shaped negative marker is Nilotic, so it is justified to regard this isogloss as significant.

35. **'One'**. No direct parallels detected between any of the three taxa. However, PN **bɛy- ~ *bɛy-ir* 'one' (ON *we-l- ~ we-r-*, Nob. *wèː ~ wèː-l ~ wèː-r*, Knz. *weː-r*, Dng. *wɛː-r*, Bir. *meː-l-ug*, Mid. *pèː-r*; cf. also Mid. *pèː* 'somebody') is well compatible, phonetically and semantically, with Nara *bi-ko* (Rn.), *biː-k* (Bd.) 'other'. Nara *doku* and PT **kV-* 'one' could only be related if *do-* in Nara were shown to be a prefixal component, which currently seems impossible.

36. **'Rain'**. PN **ar-* (ON *aru-*, Nob. *áwwí*, Dng. *aru*, Dil. *are*, Bir. *a:le*, Mid. *ár-*) is clearly the same root as East Tama **ar* (Tama *àɽ*, Ere., Mis. *aɽ*, Sun. *ar*). Whether Nara *hala* can belong here

as well is debatable: Rilly (2009: 501) lists the dialectal form *hàrà* from Saantoorta, but this seems to be the same dialectal development **l* → *r* as in 'mouth' (see above), and there are no other known cases of Nara *l* corresponding to PN and PT **r*. Initial *h-* also presents a problem; according to Rilly (2009: 302), it is an irregular reflex of PNNT **k-*, which seems to be well confirmed by several examples, so the overall correspondence for 'rain' in Nubian would be something like **kal-* rather than **ar-*.

37. **'Smoke'**. No direct parallels detected between any of the three taxa. However, it is permissible to compare PT **turu-* 'smoke' (preserved in Tama *túrú-t* and possibly in Ibi. *dùlód-à*, AS *dulud-a*, although correspondences are somewhat problematic) directly with Midob *tùrùd* 'fog, mist' (glossed this way in Werner 1993: 135, but mistakenly glossed as 'smoke' in Rilly 2009: 459).

38. **'Star'**. PN **waɲ-* is best preserved in Birgid (*waːɲ-di*) and, with various contractions and assimilations, is also found in Hill Nubian (Kad. *wonɔ-ntu*, Deb. *won-du-nu*), Midob (*òɲè-dì*) and Nile-Nubian **wiɲ-di* ← **waɲ-i-di* (ON *wiɲǯ-*, Nob. *wìnʒì*, Knz. *wissi*, Dng. *wɪssɪ*). All these forms are naturally compatible with Nara *wini* (Rn.) 'star' (Bender quotes the form *hū=wīnī*, where the first component is possibly the adjectival root 'round', cf. *hu-e* (Rn.) 'to be round').

More problematic is the relationship of these forms to PT **miɲ-* 'star' (Tama *míɲì-t*, Ere. *miŋi-t*, Sun. *miɲ-a*; Ibi. *ɲìŋí-t*, AS *ŋin-ti* with assimilation **m-* → *ɲ-* due to the influence of the palatal nasal in word-medial position). On one hand, the most straightforward correspondence for this is the Nara verb *minni-* (Rn.) 'to shine'. On the other hand, Tama data collected by Edgar shows a near-complete lack of native roots with the general structure **wVN-*, meaning that assimilation **wiɲ-* → **miɲ-* would be perfectly natural in this protolanguage. Additionally, both Nara *wini* and PT **miɲ-* display the same interesting polysemy 'star/fly (n.)' (not shared, however, by Nubian). In light of these observations, PT **miɲ-* is judged as formally compatible with both PN and Nara and may be used as evidence for descent from the same common ancestral form (presumably **waɲ-*, as in PN).

39. **'Stone'**. No direct parallels detected between any of the three taxa. ON *kit*, Nob. *kìd* 'stone' are formally comparable with PT **kad-* (Mis. *kʌt*, Ere. *kʌdda*, Sun. *kada*), but the Nubian word is restricted to the Nobiin branch of Nile-Nubian, whereas the optimal distributional candidate for PN 'stone' is **kul-*, found in Kenuzi-Dongolawi, Birgid, and Midob; additionally, vocalic discrepancies are too severe here to make the Nobiin – Tama match a valid etymology.

40. **'Sun'**. No parallels detected between any of the three taxa.

41. **'Tail'**. No parallels detected between any of the three taxa.

42. **'Thou'**. The 2nd p. pronoun, unlike the 1st p., matches nicely across all three families, allowing to reconstruct **i-* as the simple root morpheme for PNNT (inherited from Common East Sudanic). In Nubian, **i-* shifts to **e-* in Birgid and in Kenuzi-Dongolawi (and then further to **a-* in Hill Nubian), but the original articulation is still well preserved in Nobiin and Midob. The oblique (genitive) stem **i-n-* is also common for PN and PT (Rilly 2009: 519).

43. **'Tongue'**. Our reconstruction **ɲalT-* for PN is significantly different from Bechhaus-Gerst's **ʒardi*, but much closer to Rilly's **ŋal*. The word-initial phoneme here is reflected as **n-* in Nile-Nubian (Nob. *nàr*, Knz. *ned*, Dng. *nɛd*), as *ʒ-* in most Hill Nubian languages (Dil. *ʒal-e*, Kad. *ʒal-do*, Karko *ʒâr-è*, etc.), as *n-* in Birgid (*nat-ti*) and as *k-* (← **ŋ-*) in Midob (*kàd-ì ~ kàd-àŋì*); Bechhaus-Gerst interprets it as **ʒ-*, but this in no way explains the pervasive nasal reflexes. On the other hand, **ŋ-* is also excluded, since it is supposed to be preserved, not palatalized, in Hill Nubian. Based on the phonetic qualities of the different reflexes (coronal/velar nasals vs. palatal affricates), the optimal choice for reconstruction here is palatal **ɲ-*, and it seems to have been preserved in at least one Hill Nubian language: cf. Debri *ɲal-do* from Robin Thelwall's field data (unless this is a misprint instead of **jaldo*).

Word-medially, we agree with Rilly that **-l-* rather than **-r-* should be reconstructed, since **-r-* is a highly stable phoneme in Nubian; however, a simple reconstruction of the root **ɲal-* (with a complex singulative correlate **ɲal-di*) does not suffice, since reflexes in individual languages are widely different from those of the similar stem **ɲil-di* 'tooth' (see below). Already in PN, the root itself must have contained a cluster (**ɲalt- ~ *ɲald-*) or have been bisyllabic (**ɲalaT-*), which explains the loss of resonant articulation in Birgid (*nat-ti* ← **ɲalT-ti*) and word-final *-r* / *-d* in Nile Nubian (which usually appears in original **CVCV-ti* type structures, cf. 'bone' above).

This turns out to be significant on the level of external comparison, when the Nubian word for 'tongue' is compared with forms in Tama languages: Tama *àr(r)à-t*, Ere. *là:t*, Sun. *lat*, Mis. *le:t*, Ibi. *lé:d* (also *la:t* and *laed* in alternate sources), AS *let*. This item is reconstructed as PT **la:t* by Rilly, but the reconstruction does not explain the front vowel in Mis. and Ibi., not to mention the odd diphthong *-ae-* in H. Barth's and P. Doornbos' transcriptions of West Tama material. In Starostin 2014: 360, it is argued that the discrepancies in vocalism and the diphthong-containing transcriptions can only be explained if **la:t* is traced back to an older **laCat*, where **-C-* is a weak consonant with palatalizing effect, most likely **-ɲ-* (since glides like *-y-*, *-w-* do not regularly elide in intervocalic position).

The resulting reconstructions, PN **ɲalT-* (**ɲalat-* ?) and PT **laɲat*, are compatible under a simple metathesis scenario; the actual metathesis must have happened in Tama, as is indirectly hinted at by external data from other East Sudanic languages (cf. Nyimang *ŋìldì*, etc.). Admittedly, this etymology rests rather heavily on intricacies of internal reconstructions in both Nubian and Tama, as well as upon assumption of irregular metathesis; however, irregularities and non-trivial developments are fairly typical of the word 'tongue' in numerous families all over the world. In any case, PN and PT are clearly more compatible with each other than Nara *haga*, an isolated form with no external parallels.

44. **'Tooth'**. A common feature of all three compared taxa is that they all share a nasal as the first consonant in the word for 'tooth': PN **ɲəl-* (Knz. *nel*, Dng. *nɛl*, Bir. *ɲil-di*; Hill Nubian **ʒil-* → Dil. *ʒil-i*, Kad. *ʒıl-du*, etc.; Mid. *kàd-dì* ← **ŋəl-di*; Nob. *nì:d* ← **ɲil-d*), Nara *nihi*, PT **ŋes-* or **ŋeʒ-* (Ere. *ɲisi-t*, Sun. *ɲísì-t*, Mis. *ŋɛsi-t*; in Starostin 2014: 361, these forms are further compared with Ibi. *ŋóŋì-t*, AS *ŋoɲi-t* under a complex scenario of development from PT **ŋeʒ-*).

It seems, however, impossible to trace all three forms back to the same common source. There are two potential pathways here: (a) if the PT form is to be reconstructed as **ŋes-*, one could think of a common origin with PN **ɲəl-*, showing the same hypothetical correspondence that had already been suggested earlier with 'dog' and 'ear', i.e. going back to PNNT **ŋeɬ-*; (b) since Nara *nihi* must go back to **niKi* with an intervocalic velar stop, it might be compared with PT **ŋeʒ-* under the assumption of palatalization in PT (**ŋegi* → **ŋeʒi*); unfortunately, there are currently no additional examples to support such an assumption. Curiously, external data from other East Sudanic languages provides evidence for *both* solutions: velar-medial forms are attested in Surmic (Southwest Surmic **ɲıgı-t*, Southeast Surmic **ɲigi*), Jebel (**ɲigi*), and Daju (**ɲiɣi*) languages, whereas the lateral-medial form is seen in Nyimang (**ŋil-*; see tables in Starostin 2014: 722–729).

For the sake of uniformity, since we have not officially endorsed the correspondence of PN **l* to PT **s* yet, it is more prudent to go with the less radical variant (b) for the moment. Alternately, one could consider all three forms unrelated, but in the overall context of the situation, accidental similarity on all sides is hardly likely.

45. '**Tree**'. For PN, C. Rilly (2009: 423) reconstructs **ko:r-i* 'tree' vs. **ber-* 'wood'; in Starostin 2014: 82, these reconstructions are amended to **koy/i/d* and **pər* respectively, and it is also pointed out that the latter word sometimes displays the polysemy 'tree/wood' (e. g. in

Hill Nubian or in old lexical materials on Kenuzi-Dongolawi) and should probably be projected in the meaning ‘tree’ onto the PN level, whereas the original meaning of **koy/i/d* may have been more narrow (e. g. = ‘Ziziphus spina-christi’ in Dng.). Recent innovation is also perceived in Nara, where Bender's *kel* contrasts with *tûm* (Rilly's spelling) ‘wood’, a word that is glossed as *tüm* ‘tree, wood’ in the old dictionary of Reinisch and is typologically likely to represent the older equivalent for ‘tree’. Even in Tama, the protoform **gan* ‘tree’ seems to be connected with the verbal root *ge- ~ gi-* ‘to rise, to stand up’ (diachronically, ‘to stand up, to be vertical’ is a well-known possible source for ‘tree’ as ‘vertically planted wood’, e. g. Chinese *shù*) and is probably secondary next to the old root **kiŋ-* ‘wood’.

In any case, none of these forms match with each other, although some (Nara *tûm* and Tama **kiŋ-*, in particular) may have interesting parallels in other branches of East Sudanic.

46. **‘Two’**. Here, all the forms are compatible. In the case of Nubian, the most archaic form is found in Haraza Nubian *auri-yah* (Bell 1975: 84), which explains the non-trivial correspondence of Nile-Nubian **-ww-* (ON *uwo-*, Nob. *úwwó*, Knz. *owwi*, Dng. *owwɪ*) to Hill Nubian **-r-* (Dil. *ore-n*, Kad. *ɔrro*, Deb. *ɔrrɔ*, Karko *ārè*). In Nara *ari-ga*, the labial element is missing (probably due to cluster simplification), but in PT **wari* (Tama *wárì*, Ere. *wàrrí*, Sun. *warri*, Mis. *woɽa*, Ibi. *wàrí*, AS *werre*) it is found in word-initial position, suggesting metathesis: **awri* → **wari*.

47. **‘Water’**. No direct parallels detected between any of the three taxa. PT **ka:l* (Tama, Ere. *ká:l*, Mis. *qal*, Ibi. *kàr-áŋ*, AS *kar-aŋ*) is etymologically comparable with Nara *kalli* (Rn.), *kəlli* (Bd.) ‘cold’, since the semantic shift from ‘cold’ to ‘water’ is typologically plausible. External data from other East Sudanic languages suggest that Nara *mba* might be the most archaic form here (cf. Surmic **ma:m ~ *maw*, Daju **ama ~ *uma*, etc.), but comparable forms are not attested in either Nubian or Tama. The only possible exception is Old Nubian *aman-*, Nobiin *ámán* ‘water, river, Nile’; however, distribution-wise this word belongs to the same layer of «Para-Nobiin substrate» as many other forms without Common Nubian etymologies, and cannot be reliably traced back to Proto-Nubian, let alone etymologically compared with Nara *mba*.

48. **‘We’**. The PN reconstruction **a-y* is justified in detail in Starostin 2014: 86–90, where it is also argued that the clusivity opposition in certain Nubian languages (Midob, Old Nubian) is secondary and cannot be traced back to the PN level. It is quite tempting to put forward a plausible scenario in which PN **ə-y* ‘I’ / **a-y* ‘we’ would directly correlate with Nara *a-g* ‘I’ / *a-gga* ‘we’ (e. g. PNNT **ag* → **aɣ* → **əy*, but PNNT **aga* → **aɣa* → **ay* without vocalic change), but it is hardly possible to back it with additional evidence. In any case, the pronouns here quite clearly match each other on the root level. As for Tama, **wa-yi* seems to be derived from **wa* ‘I’ (sg.), meaning that there are the same problems with trying to relate it to Nubian-Nara **a-* as with the singular correlate (see above).

49. **‘What’**. In Nubian, there are two main groups of forms with the meaning of ‘what?’: one beginning with *m-* (in Nile-Nubian: Old Nubian *mi-*, Nob. *mì-n*, Knz. *mi-n*, Dng. *mɪ-n-*) and one beginning with *n-* (Dil. *na*, Kad. *na-*, Bir. *na-ta*, Mid. *nè:-*, etc.). Rilly regards them as etymologically distinct, reconstructing **mi-n* and **na: ~ *ne:* respectively. However, the second reconstruction is insecure, considering that the regular reflex of **n-* in Hill Nubian languages is *d-*, and in Midob it is *t-* (see ‘drink’ above). In Starostin 2014: 91, it is argued that the preservation of **n-* in this pronoun can only be due to some outstanding circumstances, and that under these circumstances the two forms may be traced back to a common protoform, provisionally given as **nWV-*, where **-W-* is an original labial glide or nasal. Such a form in itself could only be contracted from an earlier **nVwV-* or **nVmV-*, and this, in turn, makes it into an excellent match with Tama *nùmú-*, Ere. *numɔ-*, Sun. *nomo-*, Mis. *nùmá-*, Ibi. *nʌmʌ*, AS *nɛm-* ‘what’ ← PT **num*. Whether Nara *nda-* belongs here as well is far more debatable.

50. **'Who'**. Nara *na* and PT **na* (Tama *na-yɛ*, Sun., Mis. *na*, Ibi. *nà-n*, AS *na:-*) obviously match with each other. PN **ŋə-y* is reconstructed with an initial velar nasal (this is most clearly seen in the Mid. reflex *kà:-*), which makes it hard to relate this root at least to PT **na*, since initial **ŋ-* is quite frequent in PT, and there are no obvious factors here that would explain the fronting **ŋ- → *n-* in PT. For now, we only count the Nara/Tama match as etymologically significant.

Conclusions

Taking into consideration the importance of stratifying etymological and lexicostatistical matches to reflect their proportional representation across more and less stable layers of the basic lexicon, we separate the 50-item wordlist into a more stable and a less stable (on the average) half, based on the respective stability indexes of each item (see Table 1); Table 5 below summarizes the pairwise matchings in both halves found between all three taxa. Note that only the items that are marked with a + sign (i.e. credible etymological matches) in Table 1 are included in the calculations.

Table 5. Number of lexicostatistical matches between Nubian, Nara, and Tama.

	Nara			Tama		
	Items 1–25	Items 26–50	Overall	Items 1–25	Items 26–50	Overall
Nubian	8	5	**13**	8	10	**18**
Nara	—			10	5	**15**

The following conclusions may be drawn from the table itself, as well as from further analysis of some of the individual matches concealed behind the numbers.

1. The highest number of matches is between PN and PT: 18/50 = 36%. This is much higher than the 20% figure given in Starostin 2014: 677, where only the automatically detected pseudo-cognates were counted. However, both of these figures are statistically significant (based on empiric evidence from comparing multiple random pairs of unrelated languages, we accept a threshold of 5-6 matches out of 50 to rule out accidental similarity), and the same is true for the other two pairs as well[11].

2. Using Indo-European as a comparative benchmark, we may select, e.g., Old Indian as the approximate chronological equivalent of PN, and Proto-Germanic or Latin as the approximate chronological equivalent for the somewhat younger PT. In this case, the figure of 36% will be significantly lower than the corresponding numbers for Sanskrit vs. Latin (57%) or Sanskrit vs. Germanic (56%)[12]. This means that if Nubian and Tama languages are genetically related, their common ancestor must have probably been older than Proto-Indo-European (e.g., Sergei Starostin's recalibrated glottochronological formula in this case yields a dating of approximately 4700 BC).

[11] Had this number of parallels been seriously lower (e.g. in the range of 8–10 matches out of 50), it would have made sense to apply the same kind of permutation test as performed in, e.g., Kassian, Zhivlov, Starostin 2015, in order to establish statistical significance on a formally rigorous basis. With this amount of evidence, however, it hardly seems worth the bother.

[12] These numbers are based on preliminary 50-item wordlists, reconstructed or collected for various small language groups of Eurasia and publicly available on the Global Lexicostatistical Database website: http://starling.rinet.ru/new100/eurasia.xls.

3. The overall numeric correlations between Nubian, Nara, and Tama give no definitive answer to the question of the internal structure of their phylogenetic tree. Although 18 matches between Nubian and Tama is a significantly higher number than 13 matches between Nubian and Nara, this is primarily explicable by the fact that Nara is a modern language, while PT is a reconstruction that pushes us back about 2000 years, so that, even if all three branches split from their common source at the same time, we would naturally expect Nara to show less in common with PN and PT than both of them have in common with each other. At the moment, all three taxa appear to be more or less equidistant; future studies will let us understand better if there are any truly decisive shared innovations in between any two out of three branches of the family.

4. The distribution of cognates across the various stability groups correlates very well with our expectations (more cognates in the more stable part, fewer cognates in the less stable part) in the case of Nubian-Nara (8 against 5) and Nara-Tama (10 against 5), but not in the case of Nubian-Tama (8 against 10) — due to such shared items as 'ashes', 'egg', 'head', 'horn', 'night', 'rain' that have no parallels in Nara. Although the discrepancy is not altogether tragic, it does suggest that at least a few of these matches might ultimately be areal rather than genetic in origin: for instance, the word **ar-* for 'rain / sky' has a rather wide areal distribution and could represent a cultural *Wanderwort* rather than an inherited term.

5. On the other hand, it is notable that cognates are encountered across all semantic and functional classes of words — including body part terms, verbs, personal and interrogative pronouns, and even the negation marker. Combined with additional etymologies and occasional grammatical isoglosses that were previously published in J. Greenberg's, M. L. Bender's, and C. Rilly's works, this makes the scenario of common descent from a Proto-Nubian-Nara-Tama ancestor far more plausible than the opposite scenario of areal diffusion.

It must be stressed that, although the absolute majority of lexical parallels commented upon in this paper had previously been suggested by at least one or more of the abovementioned authors, the sort of etymological / lexicostatistical refining conducted here — where only direct semantic matches are taken into consideration, and each candidate for comparison is vetted on the issue of reconstructibility for proto-status, to reduce the risk of accidental matches — has been performed for the first time. In our opinion, the Nubian-Nara-Tama connection passes this restrictive test with flying colors. On the other hand, the question of whether it makes practical sense to try to produce a large etymological corpus solely for PNNT without taking into consideration the data from other East Sudanic languages is still open: as we have seen, NNT is a fairly deep family, probably older than Indo-European by at least one millennium, and this, combined with the relative scarceness of data on Nara and Tama as well as several millennia of areal interference, means that positive identification of large numbers of cognates is going to be a very hard task without assessing the hypothesis in an even larger context. The next logical step for such an assessment would be to investigate the position of Nyimang, a minor language group of Kordofan whose ties to NNT seem to be counterbalanced with its ties to the neighboring Temein languages; we plan to cover this issue in our next publication on East Sudanic lexicostatistics.

Abbreviations

AS — Abu Sharib; Bd. — Bender 1968; Bir. — Birgid; Deb. — Debri; Dil. — Dilling; Dng. — Dongolawi; Ere. — Erenga; Ibi. — Ibiri; Kad. — Kadaru; Knz. — Kenuzi; Krk. — Karko; Mid. — Midob; Mis. — Miisiirii; Nob. — Nobiin; ON — Old Nubian; PN — Proto-Nubian; PNNT — Proto-Nubian-Nara-Tama; PT — Proto-Tama; Rn. — Reinisch 1874; Sun. — Sungor; Wal. — Wali.

References

Abushush, Dawd, Richard J. Hayward. 2002. Nara. *Journal of the International Phonetic Association* 32(2): 249–255.

Armbruster, Charles H. 1965. *Dongolese Nubian: A Lexicon*. Cambridge University Press.

Bechhaus-Gerst, Marianne. 1985. Sprachliche und historische Rekonstruktionen im Bereich des Nubischen unter besonderer Berücksichtigung des Nilnubischen. *Sprache und Geschichte in Afrika* 6(1984/5): 7–134.

Bechhaus-Gerst, Marianne. 1989. "Nile-Nubian" reconsidered. In: M. Lionel Bender (ed.) *Topics in Nilo-Saharan Linguistics*: 85–96. Hamburg: Helmut Buske Verlag.

Bechhaus-Gerst, Marianne. 1996. *Sprachwandel durch Sprachkontakt am Beispiel des Nubischen im Niltal: Möglichkeiten und Grenzen einer diachronen Soziolinguistik*. Köln: Rüdiger Köppe Verlag.

Bechhaus-Gerst, Marianne. 2011. *The (Hi)story of Nobiin — 1000 Years of Language Change*. Frankfurt am Main: Peter Lang.

Bell, Herman. 1970. The Phonology of Nobiin Nubian. *African Language Review* 9: 115–139.

Bell, Herman. 1975. Documentary evidence on the Haraza Nubian language. *Sudan Notes And Records* 56: 1–35.

Bender, Lionel M. 1968. Analysis of a Barya wordlist. *Anthropological Linguistics* 10(9): 1–24.

Bender, Lionel M. 1971. The Languages of Ethiopia: A New Lexicostatistic Classification and Some Problems of Diffusion. *Anthropological Linguistics* 13(5): 165–288.

Bender, Lionel M. 1997. *The Nilo-Saharan Languages: A Comparative Essay*. München – Newcastle: LINCOM Europa.

Bender, Lionel M. 2005. *The East Sudanic Languages: Lexicon and Phonology*. Southern Illinois University: SIU Printing/Duplicating.

Blench, Roger. 2011. *Can Sino-Tibetan and Austroasiatic help us understand the evolution of Niger-Congo noun classes*? Ms. (presentation given at the 41st CALL, Leiden, August 29–31, 2011).

Browne, Gerald M. 1996. *Old Nubian Dictionary*. Leuven: Peeters.

Browne, Gerald M. 2002. *A Grammar of Old Nubian*. Münich: LINCOM.

Bryan, Margaret A. 1959. The T-K Languages. A New Substratum. *Africa: Journal of the International African Institute* 29(1): 1-21.

Burlak, S. A., S. A. Starostin. 2005. *Sravnitel'no-istoricheskoje jazykoznanije* [*Comparative-Historical Linguistics*]. Moscow: Academia.

Edgar, John T. 1990. *Tama Group Lexicon*. Ms.

Edgar, John T. 1991a. First Steps Toward Proto-Tama. In: M. Lionel Bender (ed.). *Proceedings of the 4th Nilo-Saharan Conference. Bayreuth, Aug. 30 – Sep. 2, 1989*: 111–131. Hamburg: Helmut Buske Verlag.

Edgar, John T. 1991b. *Maba-Group Lexicon*. Berlin: Dietrich Reimer Verlag.

Ehret, Christopher. 2001. *A Historical-Comparative Reconstruction of Nilo-Saharan*. Köln: Rüdiger Köppe Verlag.

Fleming, Harold. 1983. Kuliak External Relationships: Step One. In: R. Voßen & M. Bechhaus-Gerst (eds.). *Nilotic Studies*: 423–478. Berlin: Dietrich Reimer Verlag.

Greenberg, Joseph H. 1966. *The Languages of Africa*. Bloomington, Indiana University; Mouton & Co., The Hague.

Hammarström, Harald, Robert Forkel, Martin Haspelmath. 2017. *Glottolog 3.0*. Jena: Max Planck Institute for the Science of Human History. Available online at http://glottolog.org; accessed on 2017-05-21.

Hayward, Richard J. 2000. Observations on Tone in the Higir Dialect of Nara. In: R. Vossen, A. Mietzner, A. Meissner (eds.). *"Mehr als nur Worte...": Afrikanistische Beiträge zum 65. Geburtstag von Franz Rottland*: 247–267. Köln: Rüdiger Köppe Verlag.

Heusing, Gerald. 2004. *Die südlichen Lwoo-Sprachen. Beschreibung, Vergleich und Rekonstruktion*. Köln: Rüdiger Köppe Verlag.

Hofmann, Inge. 1986. *Nubisches Wörterverzeichnis: Nubisch-deutsches und deutsch-nubisches Wörterverzeichnis nach dem Kenzi-Material des Samuêl Alî Hisên (1863-1927)*. Berlin: Dietrich Reimer Verlag.

Jabr el Dar, Khaliifa. 2006. Towards a general orthography of the Ajang languages. In: Al-Amin Abu-Manga, Leoma Gilley, & Anne Storch (eds.). *Insights into Nilo-Saharan Language, History and Culture*: 183–198. Köln: Rüdiger Köppe Verlag.

Jakobi, Angelika. 2006. The loss of syllable-final proto-Nubian consonants. In: Al-Amin Abu-Manga, Leoma Gilley, & Anne Storch (eds.). *Insights into Nilo-Saharan Language, History and Culture*: 215–228. Köln: Rüdiger Köppe Verlag.

Kassian, Alexei. 2015. Towards a Formal Genealogical Classification of the Lezgian Languages (North Caucasus): Testing Various Phylogenetic Methods on Lexical Data. *PLoS ONE* 10(2): e0116950 (https://doi.org/10.1371/journal.pone.0116950).

Kassian, Alexei, George Starostin, Anna Dybo, Vasily Chernov. 2010. The Swadesh wordlist: an attempt at semantic specification. *Journal of Language Relationship* 4: 46–89.

Kassian, Alexei, Mikhail Zhivlov, George Starostin. 2015. Proto-Indo-European-Uralic Comparison from the Probabilistic Point of View. *Journal of Indo-European Studies* 43: 60–80.

Kauczor, P. D. 1920. *Die Bergnubische Sprache (Dialekt von Gebel Delen)*. Wien: Alfred Hölder.

Khalil, Mokhtar M. 1996. *Wörterbuch der nubischen Sprache (Fadidja/Maḥas Dialekt)*. Warszawa: Piotr O. Scholz.

Krell, Amy. 2012. *Rapid Appraisal Sociolinguistic Survey Among Ama, Karko, and Wali Language Groups (Southern Kordofan, Sudan)*. SIL International.

Lepsius, Carl Richard. 1880. *Nubische Grammatik. Mit einer Einleitung über die Völker und Sprachen Afrikas*. Berlin: W. Hertz.

Lukas, Johannes. 1933. Beiträge zur Kenntnis der Sprachen von Wadai. *Journal de la Société des Africanistes* 3(1): 25–55.

Lukas, Johannes. 1938. Die Sprache der Sungor in Wadai (Aus Nachtigals Nachlaß). *Ausland Hochschule Mitteilungen* 41(3): 171–246.

MacMichael, H. A. 1920. Darfur Linguistics. *Sudan Notes and Records* 3(3): 197–216.

Massenbach, Gertrud von. 1962. *Nubische Texte im Dialekt der Kunuzi und der Dongolawi, mit Glossar*. Wiesbaden: Deutsche Morgenländische Gesellschaft.

Meinhof, Carl. 1918. Sprachstudien im egyptischen Sudan. 33. Dulman. 34. Garko. 36. Kadero. 37. Koldegi. 38. Zur Formenlehre der nubischen Dialekte. 39. Vergleichendes Wörterverzeichnis: Deutsch-Nubisch. 40. Kenuzi. 41. Dongola. 42. Fadidja. 43. Mahas. 44. Bedauye. *Zeitschrift für Kolonialsprachen* IX: 43–64, 89–117, 167–204, 226–255.

Reinisch, Leo. 1874. *Die Barea-Sprache. Grammatik, Text und Wörterbuch. Nach den handschriftlichen Materialien von Werner Munzinger Pascha*. Wien: Wilhelm Braumüller.

Rilly, Claude. 2005. The Classification of Nara Language. *Journal of Erithrean Studies* 4: 1–27.

Rilly, Claude. 2009. *Le Méroïtique et sa famille linguistique*. Louvain – Paris – Dudley, MA: Peeters.

Starostin, George. 2010. Preliminary lexicostatistics as a basis for language classification: a new approach. *Journal of Language Relationship* 3: 79–117.

Starostin, George. 2016. From wordlists to proto-wordlists: reconstruction as 'optimal selection'. In: K. Pozdniakov (ed.). *Comparatisme et reconstruction: tendances actuelles. Faits de Langues 47*: 177–200. Berne: Peter Lang.

Starostin, Georgij. 2013. *Jazyki Afriki. Opyt postrojenija leksikostatisticheskoj klassifikacii. Tom I: Metodologija. Kojsanskije jazyki* [*Languages of Africa. An attempt at a lexicostatistical classification. Volume I: Methodology. Khoisan languages*]. Moscow: Jazyki slav'anskoj kul'tury.

Starostin, Georgij. 2014. *Jazyki Afriki. Opyt postrojenija leksikostatisticheskoj klassifikacii. Tom II: Vostochnosudanskije jazyki* [*Languages of Africa. An attempt at a lexicostatistical classification. Volume II: East Sudanic languages*]. Moscow: Jazyki slav'anskoj kul'tury.

Starostin, Sergei. 2000. Comparative-historical Linguistics and Lexicostatistics. In: C. Renfrew, A. MacMahon, L. Trask (eds.). *Time Depth In Historical Linguistics*: 223–259. McDonald Institute for Archaeological Research, Oxford Publishing Press.

Starostin, Sergei. 2007. Opredelenije ustojčivosti bazisnoj leksiki [Defining the Stability of Basic Lexicon]. In: Sergei Starostin. *Trudy po jazykoznaniju [Works in Linguistics]*: 825–839. Moscow: Jazyki slav'anskix kul'tur.

Thelwall, Robin. 1977. A Birgid vocabulary list and its links with Daju. In: H. Ganslmayr & H. Jungraithmayr (eds.). *Gedenkschrift Gustav Nachtigall 1874-1974*: 197–210. Bremen: Übersee-Museum.

Thelwall, Robin. 1978. Lexicostatistical relations between Nubian, Daju and Dinka. In: J. Leclant, J. Vercoutter (eds.). *Etudes Nubiennes. Colloque de Chantilly, 2-6 Juillet 1975*: 265–286. Caire: L'Institut Français d'Archéologie Orientale du Caire.

Thelwall, Robin. 1983. Meidob Nubian: Phonology, Grammatical Notes and Basic Vocabulary. In: Lionel M. Bender (ed.). *Nilo-Saharan Language Studies*: 97–113. Michigan: East Lansing.

Thompson, E. David. 1976. Nera. In: M. Lionel Bender (ed.). *The Non-Semitic Languages of Ethiopia*: 484–494. East Lansing: Michigan State University.

Tucker, Archibald, Margaret A. Bryan. 1956. *The Non-Bantu Languages of North-Eastern Africa*. Oxford University Press.

Vasilyev, Mikhail, George Starostin. 2014. Leksikostatisticheskaja klassifikacija nubijskikh jazykov: k voprosu o nil'sko-nubijskoj jazykovoj obshnosti [Lexicostatistical classification of the Nubian languages and the issue of the Nile-Nubian genetic unity]. *Journal of Language Relationship* 12: 51–72.

Vossen, Rainer. 1982. *The Eastern Nilotes. Linguistic and historical reconstructions*. Berlin: Dietrich Reimer Verlag.

Werner, Roland. 1987. *Grammatik des Nobiin (Nilnubisch). Phonologie, Tonologie und Morphologie*. Hamburg: Helmut Buske Verlag.

Werner, Roland. 1993. *Tìdn-áal: A Study of Midob (Darfur Nubian)*. Berlin: Dietrich Reimer Verlag.

Youn, Hyejin, Logan Sutton, Eric Smith, Cristopher Moore, Jon F. Wilkins, Ian Maddieson, William Croft, Tanmoy Bhattacharya. 2016. On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences of the United States of America* 113(7): 1766–1771.

Zyhlarz, Ernst. 1950. Die Lautverschiebung des Nubischen. *Zeitschrift für Eingeborene Sprachen* 35: 1–20, 128–146, 280–313.

Г. С. Старостин. Лексикостатистические исследования по восточносуданским языкам I: к вопросу о нубийско-нара-тама генетической общности

В статье дается подробный лексикостатистический обзор реконструированных 50-словных списков (сокращенный вариант классического списка Сводеша, состоящий из более устойчивых элементов) по трем языковым группам северо-восточной Африки — нубийской, нара и тама. Эти группы традиционно относятся к восточносуданской семье и в большинстве существующих классификаций описываются как особенно близко родственные друг другу. В обзоре продемонстрировано, что как в количественном, так и в качественном отношении лексикостатистические параллели между нубийскими, нара и тама языками убедительно интерпретируются как следы общего происхождения (а не ареальной близости) этих групп, что формально подтверждает гипотезу, которой придерживались Дж. Гринберг, М. Л. Бендер, К. Рильи и другие исследователи. При этом глоттохронологическая оценка гипотезы показывает, что пранубийско-нара-тама язык следует относить к периоду не позднее 5-го тыс. до н. э., т. е. семья в целом оказывается даже более древней, чем праиндоевропейская, и насколько детально можно будет реконструировать для нее этимологический корпус, остается неясным. Статья представляет собой первую публикацию из серии, которую предполагается посвятить комплексной этимолого-лексикостатистической оценке восточносуданской гипотезы.

Ключевые слова: нило-сахарские языки, восточносуданские языки, нубийские языки, языки тама, африканское историческое языкознание.

М. Е. Васильев[†], *М. Н. Саенко*[‡]
[†] Институт славяноведения РАН (Россия, Москва); mvhumanity@gmail.com,
[‡] Институт славяноведения РАН (Россия, Москва); veraetatis@yandex.ru

К вопросу о точности глоттохронологии: датирование языковой дивергенции по данным романских языков

Статья продолжает начатое раннее исследование, посвященное определению точности лигвистических датировок, получаемых с помощью глоттохронологии. Предметом рассмотрения является датирование языковой дивергенции (процесса разделения двух или нескольких идиомов), которое производится на материале 110-словных списков современных романских языков. Предметом рассмотрения является датирование языковой дивергенции — т. е. определение даты разделения двух или нескольких современных языков. В статье сопоставляются как традиционные, так и вновь предложенные модели глоттохронологии. При этом особое внимание уделяется величине погрешности и надёжности глоттохронологических вычислений на различных временных глубинах. Результаты проведенного исследования позволяют не только количественно оценить границы точности глоттохронологии при датировании романских языков, но также сделать ряд практических выводов, касающихся возможностей применения глоттохронологии на любом другом языковом материале.

Ключевые слова: глоттохронология, лексикостатистика, список Сводеша, романские языки

Данная статья является продолжением нашего исследования, цель которого — оценка точности и надежности лингвистических датировок, получаемых с использованием глоттохронологических расчетов. Первая часть работы (Васильев, Саенко 2016) была посвящена рассмотрению лексических изменений, происходящих в списке базисной лексики одного языка с течением времени, и определению временной дистанции между языком-предком и его потомками на основе нескольких различных глоттохронологических методов. Во второй части мы обратимся к процессу языковой дивергенции — т. е. независимому изменению лексики родственных идиомов после их разделения.

Датирование языковой дивергенции представляет наибольший практический интерес для сравнительно-исторического языкознания (в частности, при изучении дальнего языкового родства), так как дает возможность получить датированные генеалогические деревья и тем самым — сопоставить результаты праязыковой реконструкции с историческими или доисторическими событиями, не имеющими письменных свидетельств. При этом глоттохронология, несмотря на многочисленные критические замечания, до сих остается наиболее популярным, а в случае с малоизученными семьями языков — единственно доступным методом для получения лингвистических датировок. Нередко глоттохронологические датировки, опубликованные в узкопрофильных статьях и имеющие, как правило, лишь предварительный, оценочный характер, используются также в научных и научно-популярных работах по смежным дисциплинам (история, археология и др.), что способствует их распространению и популяризации за пределами сравнительно-исторического языкознания. Зачастую такие данные преподносятся читателю уже как установленный научный факт, подтвержденный строгим лингвистическим методом[1].

[1] См., например, (Алексеев 2013: 63; Багаев 2015: 127). Более того, даже специалисты, известные критическим отношением к глоттохронологии, нередко сами пользуются её результатами в своих работах. См. например, работы *Л.* С. Клейна (Клейн 2010: 33–36, 122, 305–306, 466).

В этих обстоятельствах исследование точности глоттохронологических моделей приобретает особое значение как для профессионального лингвистического сообщества, так и для широкой аудитории, поскольку, с одной стороны, оно позволит специалистам получить представление о надежности и границах применимости метода при датировании языковой дивергенции, а неспециалистам — поможет избежать лишних разочарований, связанных главным образом с непониманием «действительных и мнимых»[2] возможностей глоттохронологии.

В настоящей статье представлена попытка такого исследования, выполненная на актуальном романском материале, представленном списками базисной лексики 56 языков и диалектов. Для датирования языковой дивергенции нами будут использованы три различных глоттохронологических метода: классическая глоттохронология М. Сводеша, усовершенствованная методика, разработанная С. А. Старостиным, и недавно предложенная модель, основанная на потоковом подходе к процессу лексических замен. При этом основные цели и задачи исследования останутся теми же, что и в первой части работы:

1. Сопоставить результаты применения известных глоттохронологических моделей (М. Сводеша, С. А. Старостина, потоковой модели) на романском материале.
2. Определить точность полученных датировок и при необходимости провести калибровку рассматриваемых моделей по имеющимся исходным данным (опорным точкам).
3. Оценить фактическую и теоретическую погрешность рассматриваемых моделей при датировании дивергенции между романскими идиомами и группами идиомов.

По итогам рассмотрения мы сделаем выводы о пределах точности и практической ценности глоттохронологических расчетов при датировании романских языков, а также о возможностях применения глоттохронологии на другом языковом материале и бóльших временных глубинах.

1. Исходные данные

Обе части нашего исследования проводятся на материале обновленной лексикостатистической базы романских языков, включающей в себя этимологизированные 110-словные списки 52-х современных и 4-х исторических литературных идиомов (архаическая и классическая латынь, староитальянский и старофранцузский)[3]. Используя приведенные в базе проценты совпадений[4] между парами или группами языков, а также сведения о дате их разделения, известные из экстралингвистических источников, сформируем набор исходных данных (или «опорных точек»), необходимых для измерения скорости расхождения языков, а также определения других параметров дивергенции — см. табл. 1. В полученной таблице для каждой сравниваемой пары идиомов (или групп идиомов[5])

[2] Из заглавия тезисов к докладу В. М. Иллич-Свитыча: «Мнимые и действительные возможности лексикостатистики» (Иллич-Свитыч 1966).

[3] Подробнее о составе и принципах формирования базы, а также методике сбора списков см. в первой части работы — Васильев, Саенко 2016: 262–263.

[4] Полная таблица попарных совпадений между романскими идиомами приводится в дополнительных материалах к вышеуказанной статье (там же), которые доступны на сайте журнала www.jolr.ru.

[5] При сравнении нескольких пар языков (строки 5, 7, 8), в таблице приводится минимальное, максимальное и среднее арифметическое значение совпадений между соответствующими списками. Среднее арифметическое значение рассчитывается как сумма всех полученных процентов совпадений, деленная на количество слагаемых. Подробнее см. в сноске 2 к статье Васильев, Саенко 2016: 263.

Таблица 1. Исходные данные для определения скорости дивергенции романских языков
(по данным 110-словных списков для 54-х романских идиомов)

№	Сравниваемые списки	Мин. % совп.	Средний % совп.	Макс. % совп.	Дата разделения, год	Время дивергенции, лет
1	*Любые идиомы*	-	100	-	0	0
2	Португальский — галисийский	-	97	-	1400	600
3	Старофранцузский (Кретьен де Труа, 1140 г. н.э.) — староитальянский (*Данте*, 1270 г.)	-	91	-	480	790*
4	Румынский — арумынский	-	89	-	900	1100
5	Франко-провансальский — пикардский, валлонский	86	87,5	89	850	1150
6	Португальский – кастильский испанский	-	86	-	710	1290
7**	Португальский — фриульский, ладинские, руманшские, итало-романские, франко-провансальский, окситанский, галло-романские	71	78,2	83	480	1520
	Каталанские — фриульский, ладинские, руманшские, итало-романские, галло-романские	73	79,3	85		
	Пикардский и валлонский — фриульский, ладинские, руманшские, итало-романские, каталанские, иберийские	69	78,0	83		
	Общее значение:	69	78,5	85		
8	Балкано-романские — остальные романские	61	69,4	78	270	1730

* Значение рассчитано относительно даты фиксации для языка *Данте* — 1270 г.

** Следует отдельно пояснить выбор идиомов в строке 7. Поскольку основное ядро романских языков представляет собой практически непрерывный диалектный континуум, процент совпадений между географически соседними идиомами может иметь некий «подскок» в силу наличия ареальных инноваций. Для смягчения нежелательного влияния завышенных значений на результаты глоттохронологических расчетов использовался следующий прием: при рассмотрении какого-либо звена диалектный цепи соседствующие с ним звенья исключались и проводилось сравнение только с географически не примыкающими идиомами. Например, португальский сопоставлялся со всеми остальными «ядерными» (т. е. всеми, кроме балкано-романских и сардинских) языками, кроме соседствующих с ним идиомов Испании.

указаны проценты совпадений между их 110-словными списками, предполагаемая дата их разделения, реконструируемая по экстралингвистическим данным, а также время их развития после разделения — т. е. собственно дивергенции.

Например, доля совпадений между списками франко-провансальского и пикардского составляет 89%, франко-провансальского и валлонского — 86%. Сложив оба значения и разделив сумму пополам, получаем среднее значение — 87,5%, представленное в таблице.

Для датирования разделения этих и других идиомов были выбраны исторические события, в значительной мере повлиявшие на историю романского мира, а тем самым — и на судьбу носителей романских языков. К примеру, разделение франко-провансальского с пикардским и валлонским, по всей вероятности, следует связывать с обособлением Прованса в ходе распада империи Каролингов в 850–860-х гг. Аналогичным образом можно соотнести отделение балкано-романской группы от основного массива романских языков — с выводом римских легионов из Дакии в 271 г.; распад основной романской общности — с крушением Западной Римской империи в 476 г.; расхождение португальского и кастильского испанского — с арабским завоеванием Пиренейского полуострова в 711–718 гг. Разделение румынского и арумынского в 900 г. датируется на основе свидетельств об упоминании арумын в качестве отдельной этнической группы в византийских хрониках IX в.[6]

Разумеется, любые попытки установить хронологическую корреляцию между лингвистическими изменениями и историческим контекстом могут вызвать обоснованные возражения. Прежде всего очевидно, что начало языковой дивергенции не всегда связано с переломными историческими событиями (такими как завоевания, миграции, природные бедствия и т. д.), а может происходить вследствие внутренних культурных, социальных, экономических и других причин. Таким образом разделение языков может как предшествовать физическому разделению их носителей, так и произойти спустя некоторое (иногда — продолжительное) время после него — например, при условии поддержания культурных контактов между ними[7].

Во-вторых, сам термин «дата разделения» можно применять лишь условно, поскольку в действительности начало дивергенции не является одномоментным событием, а представляет собой процесс, происходящий постепенно с течением времени. Иначе говоря, было бы корректнее говорить не о «дате», а о некотором «периоде» дивергенции, по завершении которого мы можем зафиксировать те или иные различия, свидетельствующие о независимом развитии идиомов. При этом, как показывают конкретные случаи дивергенции, момент первой замены не всегда являются надежным свидетельством начала разделения[8]. Если же связывать фактическое разделение с накоплением определённого числа различий между идиомами, то встаёт вопрос, какое количество (или качество) различий считать критическим.

В то же время следует отметить, что все известные на сегодняшний день глоттохронологические модели были получены с использованием (пусть и в имплицитном виде)

[6] См. подробнее в Нарумов 2001: 638.

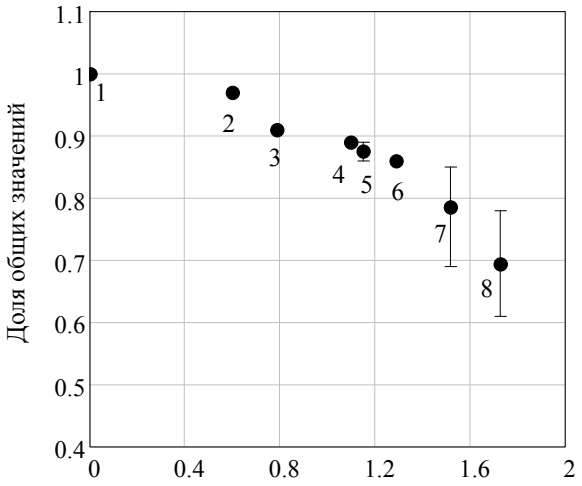
[7] Этот факт, в частности, может проявляться в значительном разбросе процентов совпадений, полученных для одной и той же точки исходных данных.

[8] Например, в сардинских идиомах континуанты *magnus* по-прежнему являются основным словом для *big*, в то время, как в остальном романском мире *magnus* было вытеснено *grandis*. Мы знаем, что эта замена является довольно старой, поскольку для языка Апулея базовым следует считать именно *grandis*, а не *magnus*. Однако связывать отделение сардинских от основного ядра романских языков с заменой *magnus* > *grandis*, было бы неправильным: в сардинских мы находим целый ряд более поздних романских инноваций (*ignis* > *focus*; *iecur* > *ficatum*; *vir* > *homo*; *os* > *bucca*; *cutis* > *pellis*; *brevis* > *curtus* и др.).

допущения о некотором *моменте времени*, соответствующем началу разделения языков в ходе дивергенции. При этом поиск и калибровка параметров моделей производилась, как правило, на основе исторических или доисторических (например, полученных с помощью археологии) сведений о жизни носителей рассматриваемых языков. Очевидным образом отказ от данного допущения и методики калибровки привел бы к невозможности получения числовых параметров моделей, а следовательно — к невозможности применения глоттохронологии в целом. Поэтому, осознавая всю проблематичность и несовершенство данного подхода, мы, тем не менее, должны признать его использование целесообразным и оправданным в рамках нашего исследования.

Полученные исходные данные можно представить в виде диаграммы, на которой каждая опорная точка соответствует строке таблицы 1 с тем же номером.

Рисунок 1. Изменение доли совпадений в базисной лексике романских языков в зависимости от времени дивергенции. Для точек 5, 7 и 8 показан диапазон разброса долей совпадений и среднее арифметическое значение.



На рисунке видно, что наблюдаемый процесс дивергенции (так же как и процесс изменения лексики одного языка, рассмотренный нами ранее[9]), имеет статистическую природу. В частности, для точки 7 доля совпадений между различными идиомами с одной и той же предполагаемой датой разделения (480 г.) варьируется от 69 до 85 %[10], что указывает на вероятностный характер лексических замен. Отметим также, что все опорные точки лежат в относительно небольшом временном диапазоне (до 2000 лет), которым и будут в основном ограничены хронологические рамки нашего рассмотрения. При этом мы наблюдаем компактное расположение опорных точек вдоль некоторой линии регрессии[11] на всем рассматриваемом интервале времени, что дает основания говорить о наличии *значимой* статистической зависимости между долей совпадений в лексике разделившихся языков и временем их дивергенции. Определение свойств этой зависимости мы будем производить на основе трех различных глоттохронологических методов,

[9] Ср. с аналогичной диаграммой на рис. 2 (Васильев, Саенко 2016: 264–265).

[10] Соответствующий диапазон для точки 8 (270 г.) еще шире — от 61 до 78%.

[11] Кривая, наиболее точно отражающая распределение экспериментальных данных.

два из которых (методика М. Сводеша и С. А. Старостина) хорошо известны и уже рассматривались в первой части нашей работы, а третий (потоковая модель) предложен относительно недавно[12]. Для этого мы сопоставим значения каждой модели с исходными данными и при необходимости проведем калибровку их параметров, а затем сделаем выводы о соответствии или несоответствии полученных моделей общему характеру процесса дивергенции и его особенностям.

2. Анализ глоттохронологических моделей

2.1. Глоттохронология М. Сводеша

В соответствии с методикой М. Сводеша развитие языков-потомков после распада их общего предка (праязыка), происходит независимо друг от друга, что отражено в одном из главных постулатов классической глоттохронологии:

> Вероятность того, что слово из О[сновного] С[писка] праязыка сохранится в О[сновном] С[писке] одного языка-потомка, не зависит от вероятности его сохранения в аналогичном списке другого языка-потомка (Арапов, Херц 1974: 25).

Данное утверждение позволило Сводешу перейти от общего уравнения глоттохронологии, имеющего вид $N_{Sw}(t) = e^{-\lambda \cdot t}$, к модели дивергенции двух или нескольких языков-потомков путем возведения исходного выражения в соответствующую степень. В частности, для двух идиомов мы получаем формулу:

$$N2_{Sw}(t) = N_{Sw}(t)^2 = e^{-2\lambda \cdot t}.$$

Таким образом, «скорость» дивергенции двух языков относительно друг друга (2λ) оказывается вдвое больше «скорости» изменения одного языка относительно своего предка (λ), что соответствует принятому утверждению о независимом развитии идиомов после их разделения.

Подставляя в качестве «коэффициента потерь» λ[13] значение 0,16, предложенное Сводешем[14], получим итоговую модель:

$$N2_{Sw}(t) = e^{-2 \cdot 0{,}16 \cdot t}$$

Используя данную формулу, можно подсчитать время дивергенции двух родственных языков (t), по известному проценту совпадений между их основными списками (N). Например, согласно этой модели разделение румынского и арумынского с долей совпадения 89% (N=0,89), должно было произойти около 370 лет назад:

---

[12] Впервые данная методика была описана в статье Васильев, Милитарёв, 2008: 509–536.

[13] Коэффициент потерь (λ) в формуле Сводеша определяет темп замен в базисной лексике языка: чем больше λ, тем больший процент значений изменится в списке за определенный промежуток времени. Не следует путать «коэффициент потерь» с «коэффициентом сохраняемости» (r), который также часто используется в работах по глоттохронологии и означает долю слов, сохранившихся (т.е. оставшихся неизменными) в списке за 1000 лет.

[14] Данное значение λ соответствует «коэффициенту сохраняемости» r=0,85, первоначально вычисленному Сводешем для 200-словных списков (Сводеш 1960: 34). Позднее величина λ неоднократно уточнялась и корректировалась (в том числе — по стословным спискам). Тем не менее, наибольшую известность приобрело именно исходное значение, которое еще долго использовалось в дискуссиях как сторонниками, так и критиками глоттохронологии. Подробнее см. в Васильев, Саенко 2016: 260–261.

$$t = -\frac{\ln(N)}{2 \cdot \lambda} = -\frac{\ln 0{,}89}{2 \cdot 0{,}16} = \frac{0{,}117}{0{,}32} = 0{,}366 \text{ тыс. лет},$$

— т. е. примерно в XVII в., что существенно позже предполагаемой даты — IX в. (см. табл. 1, строка 4).

Результаты аналогичных расчетов, проведенных для диапазона возможных значений N, представлены на рис. 2.

Рисунок 2. Сравнение модели М. Сводеша с исходными данными: $N2_{Sw}(t) = e^{-2 \cdot 0{,}16 \cdot t}$.

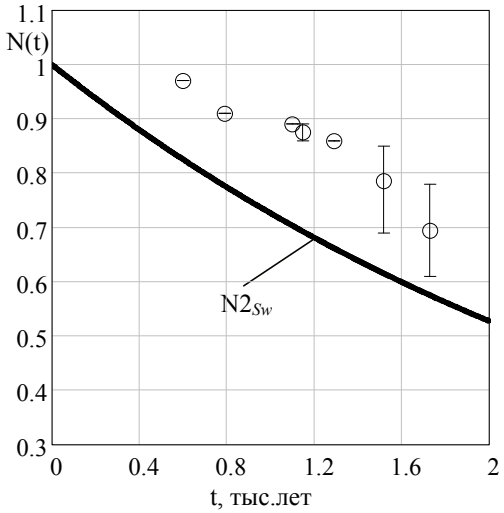


Рисунок показывает, что на всем рассматриваемом интервале использование модели приводит к существенному (в 2 и более раз) «омоложению» расчетных датировок по отношению к предполагаемым датам разделения. Величина отклонения выглядит особенно внушительно при сравнении с результатами, полученными при моделировании процесса замен в лексике одного языка[15], где применение формулы Сводеша с тем же коэффициентом λ=0,16 и на том же романском материале обеспечило очень хорошее совпадение расчетных и фактических значений[16].

Для повышения точности модели попробуем провести калибровку коэффициента λ по имеющимся исходным данным. Для этого воспользуемся уже знакомым нам методом наименьших квадратов[17]. Смысл метода заключается в поиске такого значения λ, при котором суммарное отклонение (ε) между фактическими и расчетными долями совпадений, вычисленное для всех опорных точек, окажется минимальным. В общем виде формулу для поиска оптимального значения λ можно представить следующим образом:

$$\varepsilon = \sum_{i} (N_{p,i} - N_{\phi,i})^2 \rightarrow \min,$$

где N_p — расчетное значение доли совпадений, вычисленное по формуле $N2_{Sw}(t) = e^{-2 \cdot \lambda \cdot t}$, i — номер опорной точки, а N_ϕ и t — фактические значения доли совпадений и времени дивергенции[18].

[15] Ср. с аналогичным графиком на рис.3 (Васильев, Саенко 2016: 266).

[16] Напомним, что соответствие оказалось настолько точным, что после калибровки модели по исходным данным величина коэффициента λ не изменилась и совпала с исходным значением — 0,16.

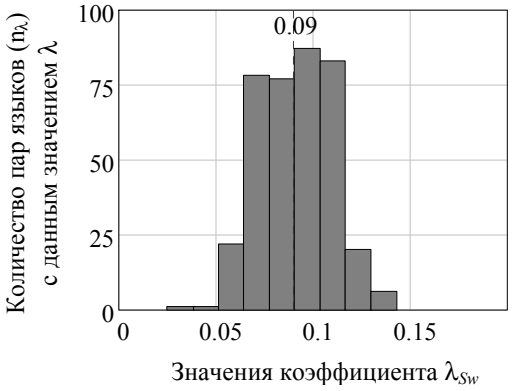
[17] См. подробное описание метода с примерами его использования в первой части исследования (там же: 265–267).

[18] Например, для случая с румынским и арумынским языками доля совпадений между их списками составляет 89% (N_ϕ=0,89), а время дивергенции 1100 лет (t=1,1). Подставляя значение времени в формулу

Подставляя в данную формулу данные из табл. 1 и выполнив необходимые вычисления[19], получаем коэффициент λ=0,09, удовлетворяющий условию наименьшего суммарного отклонения, величина которого составила ε=0,76 (см. рис. 3). Таким образом, калиброванная модель Сводеша будет иметь вид:

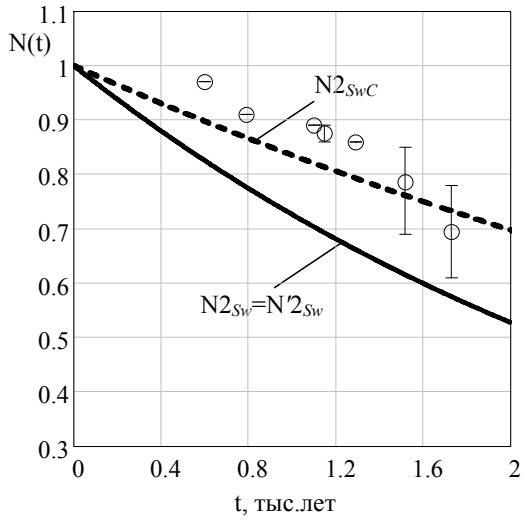
$$N2_{SwC}(t) = e^{-2\cdot 0,09\cdot t}.$$

Рисунок 3. Распределение значений коэффициента λ_{Sw}, рассчитанных по опорным точкам (табл. 1) с помощью модели Сводеша. Найденное оптимальное значение λ_{Sw} соответствует математическому ожиданию 0,09 при среднем квадратическом отклонении σ_λ=0,02.



Для сопоставления исходной и новой модели, а также оценки результатов калибрования обратимся к диаграмме на рис. 4.

Рисунок 4. Сравнение исходной и калиброванной моделей Сводеша с исходными данными:
$N2_{Sw}(t) = e^{-2\cdot 0,16\cdot t}$ (ε=8,46) — исходная модель Сводеша[20];
$N2_{SwC}(t) = e^{-2\cdot 0,09\cdot t}$ (ε=0,76) — калиброванная модель Сводеша.



Сводеша, мы получаем: $N_p=N2_{Sw}(t)=e^{-2\cdot\lambda\cdot 1,1}$. Очевидно, для наилучшего соответствия между фактическим и расчетным значением необходимо найти такое значение λ, при котором величина отклонения ε будет минимальной:

$$\varepsilon = (e^{-2\cdot\lambda\cdot 1,1} - 0,89)^2 \rightarrow \min.$$

[19] Большинство расчетов и построение графиков проводилось с помощью системы MathCad.

[20] Как уже говорилось выше (сноска 16), при калибровке коэффициента λ по исходным данным для изменения лексики одного языка его значение совпало с исходным (0,16). Таким образом, вид моделей $N2_{Sw}$ и $N'2_{Sw}$ (с исходным и калиброванным коэффициентами) также совпадает (см. в первой части работы — там же: 267).

Как следует из представленной диаграммы, переход к калиброванной модели с коэффициентом λ=0,09 позволил заметно уменьшить расхождение между расчетными и фактическими значениями[21]. В то же время для большинства опорных точек полученные датировки по-прежнему оказались моложе ожидаемых. При этом форма полученной кривой указывает, что дальнейшая калибровка модели не позволит добиться существенного улучшения результатов в силу принципиального несоответствия между простой экспоненциальной зависимостью и общим характером процесса дивергенции.

Попытка преодолеть этот сущностный недостаток классической глоттохронологии Сводеша была осуществлена в рамках усовершенствованной методики С. А. Старостина, к рассмотрению которой мы переходим.

2.2. Глоттохронологический метод С. А. Старостина

В работе (Starostin 2000: 233–259) С. А. Старостин устанавливает, что основной причиной неудач методики Сводеша является невыполнение двух основных постулатов глоттохронологии: о постоянной скорости лексических изменений и одинаковой стабильности значений в основном списке. Анализируя конкретные примеры развития базисной лексики, он предлагает ввести в исходную глоттохронологическую формулу $N_{Sw}(t) = e^{-\lambda \cdot t}$ две поправки: *замедляющую*, связанную с проявлением в списке наиболее устойчивой части лексики (λ=λ·N(t)), и *ускоряющую* — отражающую «устаревание» сохранившейся лексики, а следовательно — ускорение замен среди устаревших значений (λ=λ·t). Таким образом, процесс лексической дивергенции в одном языке должен описываться более сложным соотношением:

$$N_{St}(t) = e^{-\lambda \cdot N_{St} \cdot t^2}$$

При этом Старостин вслед за Сводешем принимает постулат о независимости развитии языков-потомков[22], что позволяет использовать такую же методику получения модели дивергенции — путем возведения исходной формулы во вторую степень:

$$N2_{St}(t) = N_{St}(t)^2 = e^{-2 \cdot \lambda \cdot N_{St} \cdot t^2} = e^{-2 \cdot \lambda \cdot \sqrt{N2_{St}} \cdot t^2}.$$

Апробируя полученную формулу на различном языковом материале (в том числе романском), Старостин определяет также константу λ, которая по разным подсчетам варьируется около величины 0,05. Таким образом, конечная модель для датирования относительного развития языков будет иметь вид:

$$N2_{St}(t) = e^{-2 \cdot 0{,}05 \cdot \sqrt{N2_{St}} \cdot t^2}.$$

Сопоставление графика полученной модели с опорными точками (рис. 5) подтверждает, что введение поправок позволило добиться значительно лучшего соответствия между фактическими и расчётными датировками, причем на всем рассматриваемом временном диапазоне. При этом параметры модели, найденные нами в ходе калибровки, существенно не отличаются от предложенных. Так, с помощью метода наименьших квадратов получаем коэффициент λ=0,07 (см. рис. 6), близкий к исходному значению 0,05[23].

---

[21] Об этом же свидетельствует изменение величины суммарного отклонения, которое уменьшилось с ε=8,46 для исходной модели до ε=0,76 для калиброванной формулы.

[22] См. п. 2.1 выше.

[23] Наблюдаемое при этом заметное уменьшение суммарного отклонения (которое снизилось с 1,97 до 0,72 после калибровки модели), объясняется в первую очередь неравным количеством сравниваемых идиомов в различных опорных точках. Так, опорные точки 7 и 8 содержат результаты сравнения для нескольких

Рисунок 5. Сравнение моделей Старостина с различными коэффициентами λ:
$N2_{St}(t) = e^{-2\cdot0,05\cdot\sqrt{N2_{St}}\cdot t^2}$ (ε=1,97) — исходная модель Старостина;
$N2_{StC}(t) = e^{-2\cdot0,07\sqrt{N2_{StC}}\cdot t^2}$ (ε=0,72) — калиброванная модель Старостина;
$N'2_{St}(t) = e^{-2\cdot0,11\cdot\sqrt{N'2_{St}}\cdot t^2}$ (ε=7,07) — модель Старостина с коэффициентом λ, калиброванным по данным для развития лексики одного языка.

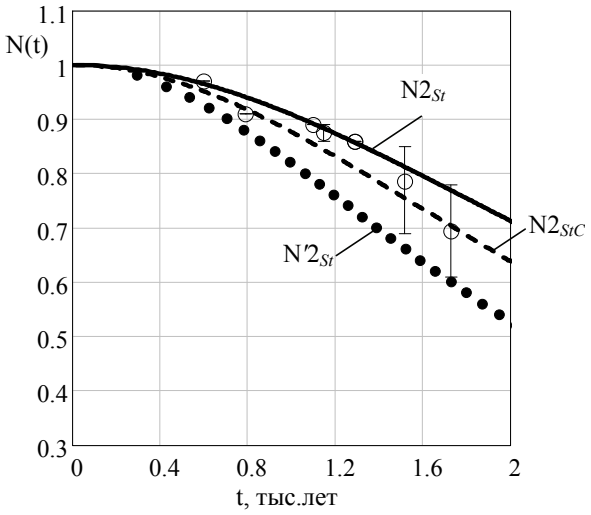
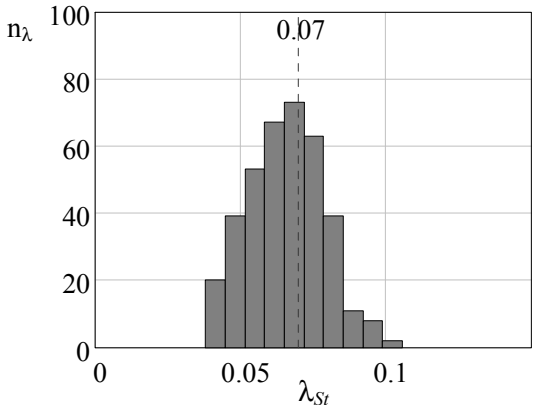


Рисунок 6. Распределение значений коэффициента λ, рассчитанных для опорных точек (табл. 1) по модели Старостина. Математическое ожидание коэффициента λ_{St} составляет 0,07; значение среднего квадратического отклонения σ_λ=0,013.



Примечательно, что при использовании найденного в первой части работы[24] коэффициента λ=0,11, полученного для модели Старостина по романским данным для дивергенции одного языка, результаты оказываются даже хуже, чем в случае с исходной константой (0,05) — см. рис. 5. Это несоответствие указывает на некорректность[25] применяемой как Сводешем, так и Старостиным методики перехода от модели независимого

десятков пар языков, в то время как предыдущее — всего для одной-двух пар. Таким образом, при вычислении суммарного отклонения, соответствие кривой последним двум точкам имеет гораздо больший «вес», чем всем остальным вместе взятым.

[24] См. Васильев, Саенко 2016: 268–269.

[25] Очевидно, в противном случае найденные коэффициенты λ должны были совпасть или иметь сходные значения.

развития одного идиома к модели относительной дивергенции и тем самым свидетельствует о невыполнении постулата Сводеша о независимом развитии языков-потомков. В свою очередь отказ от третьего постулата приводит к необходимости учитывать согласованные изменения в базисной лексике родственных языков после их разделения. Примером такого подхода к описанию процесса дивергенции является потоковая модель, которую мы рассмотрим далее.

2.3. Потоковая глоттохронологическая модель[26]

В отличие от представленных выше методик М. Сводеша и С. А. Старостина потоковая модель базируется на предположении, что развитие двух языков-потомков обладает определенной согласованностью, в результате чего даже спустя значительное время после их разделения в списках обоих идиомов могут заменяться одни и те же значения. При этом в каждом из списков можно выделить две составляющие, одна из которых соответствует значениям, которые развиваются сходным образом в обоих языках, а вторая — независимо развивающейся части списка. Причем в начальный момент разделения все значения будут развиваться согласованно (т. е. относиться к связанной составляющей), а в ходе дальнейшего развития — постепенно утрачивать эту согласованность и переходить в независимую составляющую, увеличение которой и будет соответствовать собственно дивергенции двух языков[27]. Если предположить, что убывание связанной составляющей происходит экспоненциально, а сам процесс замен внутри каждой из составляющих представляет собой сумму потоков[28] случайных событий, то формула, описывающая общий процесс дивергенции между двумя языками, примет вид[29]:

$$N2_P(t) = c_0 + c_1\left(\frac{\mu}{\mu-\eta}\cdot e^{-\eta t} + \frac{\eta}{\eta-\mu}\cdot e^{-\mu t}\right),$$

где константы c_0 и c_1 соответствуют количеству значений в наиболее устойчивой и изменяющейся частях списков, коэффициент η определяет скорость потерь в изменяющейся части списка, а величина μ является показателем увеличения его независимой составляющей.

---

[26] На всякий случай подчеркнем, что обсуждаемая ниже модель дивергенции *не идентична* потоковой модели, описывающей процесс замен в лексике одного языка и рассмотренной в первой части работы (Васильев, Саенко 2016: 269–270). Использование термина «потоковая» применительно к обеим моделям отражает тот факт, что в их основе лежит одно и то же представление о процессе замен в базисной лексике как о совокупности потоков редких случайных событий, каждое последующее из которых не зависит от предыдущего. Несмотря на теоретический характер, данное представление имеет ряд практических следствий, непосредственно влияющих на результаты и саму методику проводимого исследования. В частности, численная оценка точности и надежности глоттохронологических моделей, становится возможной благодаря известным статистическим свойствам стационарных потоков, используемых при моделировании.

[27] Более подробное описание и теоретическое обоснование модели содержится в статье Васильев, Милитарев 2008: 523–529.

[28] Каждый из таких потоков соответствует процессу замен одного из значений списка.

[29] Полный вывод формулы дается в Приложении к статье (Васильев, Милитарев 2008: 535–536). Особый интерес представляет содержательный анализ этой формулы и, в частности, вопрос соотношения значений η и μ, а также возможность перехода к упрощенному виду формулы при их равенстве ($\eta=\mu$). Обсуждение этих особенностей требует отдельного подробного рассмотрения, которое, к сожалению, выходит за рамки настоящей статьи.

Путем калибровки модели по исходным данным были получены следующие значения параметров[30]:

$$c_0=0{,}000;\ c_1=1{,}000;\ \eta=0{,}612^{31};\ \mu=0{,}611.$$

При подстановке данных значений в исходное выражение получаем модель:

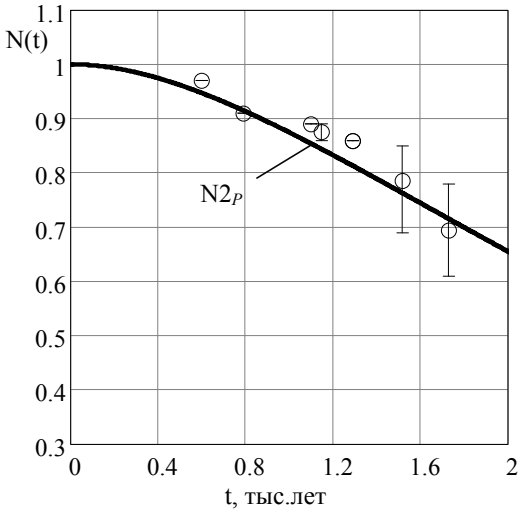
$$N2_P(t)=0{,}000+1{,}000\cdot\left(1297\cdot e^{-0{,}612t}-1296\cdot e^{-0{,}611t}\right),$$

которая в силу близости значений η и $\mu\approx0{,}61$ может быть представлена в упрощённой форме:

$$N2_P(t)=e^{-0{,}611t}(1+0{,}61t).$$

График полученной модели (рис. 7) наглядно демонстрирует хорошее совпадение расчетных значений с исходными данными на всем временном интервале[32].

Рисунок 7. Соответствие потоковой модели исходным данным: $N2_P(t)=e^{-0{,}611t}(1+0{,}61t)$; ε=0,54.

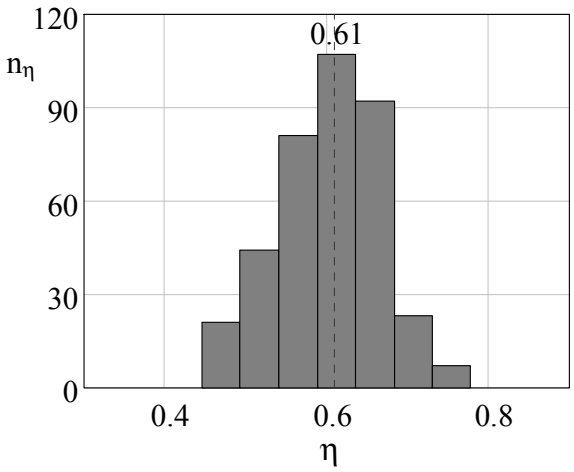


[30] Напомним, что калибровка всех моделей производилась типовым методом наименьших квадратов отклонений, см. также сноску 18.

[31] Распределение значений коэффициента η, полученных в ходе калибровки, представлено на рис. 8. Отметим, что найденное значение η (0,61) не совпадает с полученным ранее (0,45) на другом материале, включающем данные нескольких языковых семей (см. Васильев, Милитарев 2008: 529). Несоответствие полученных значений лишний раз подтверждает, что параметры глоттохронологический моделей в значительной мере определяются набором и качеством исходных данных, используемых при их калибровке. Поэтому при поиске параметров модели особенно важно привлечение максимального широкого языкового материала, как с точки зрения географического охвата, так и временной глубины.

[32] Видимая неравномерность распределения опорных точек относительно кривой связана с неравномерностью распределения исходных данных. Как уже отмечалось выше, наибольшим весом обладают две крайние опорные точки №№7 и 8 (см. табл. 1 и ссылку 23), поэтому в результате калибровки методом наименьших отклонений именно они главным образом определяют значения параметров модели. При этом легко убедиться, что указанные точки расположились симметрично по обе стороны расчётной кривой $N2_P$.

Рисунок 8. Распределение значений коэффициента η, рассчитанных по опорным точкам (табл. 1) с помощью потоковой модели. Полученное математическое ожидание величины η составляет 0,61; значение среднего квадратического отклонения σ_η=0,065.



Завершив подробное рассмотрение каждой из глоттохронологических моделей, перейдем теперь к сравнению и анализу полученных результатов.

2.4. Сравнение полученных моделей и их оценка

Сопоставление полученных моделей будем производить с помощью графиков, представленных на рис. 9 (а, б, в), а также их числовых параметров, приведенных в табл. 2.

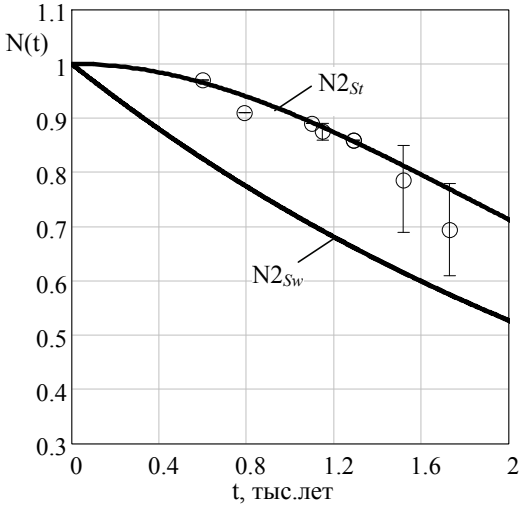
В первую очередь отметим, что калибровка формулы М. Сводеша ($N2_{Sw}$) как по данным общей, так и относительной дивергенции не приводит к существенному увеличению точности расчетных датировок, что, как уже говорилось выше, вызвано несоответствием экспоненциальной зависимости характеру процесса замен при расхождении двух языков-потомков. Так, при использовании модели с калиброванным коэффициентом λ_{SwC}=0,09 мы получаем правдоподобные даты разделения для интервала 1500–1700 лет и сильно заниженные (или наоборот — завышенные) значения за его пределами.

Гораздо лучшие результаты демонстрирует модель С. А. Старостина ($N2_{St}$), которая, после отказа от двух постулатов Сводеша и внесения соответствующих поправок в классическое уравнение глоттохронологии, позволила добиться хорошего совпадения с опорными точками на всем рассматриваемом временном отрезке. Причём оптимальное значение коэффициента λ_{StC} (0,07), найденное в ходе подбора параметров модели по фактическим данным, оказалось очень близко к исходному (0,05) — см. рис. 9а,в и табл. 2. При этом данное значение λ_{StC} заметно отличается от полученного при калибровке модели Старостина по тем же романским данным, но для одного языка (λ'_{StC}=0,11) (рис. 9б). Обнаруженное несоответствие свидетельствует о том, что процесс дивергенции между родственными языками не может быть смоделирован на основе двух независимых процессов развития каждого из них[33] и, следовательно, указывает на несостоятельность постулата Сводеша о независимом развитии языков-потомков после их разделения. Отказ от принципа независимости создает предпосылку для перехода к модели, которая могла бы учитывать согласованность процесса лексических замен в разделившихся идиомах.

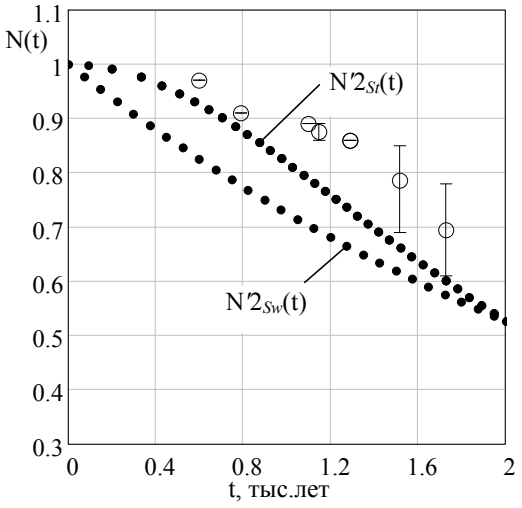
[33] Что подразумевается методикой Сводеша и Старостина при получении модели дивергенции из общей формулы глоттохронологии.

Данный подход был реализован при построении потоковой модели ($N2_P$), эффективность использования которой при датировании процесса дивергенции подтверждается как графически (рис. 9в), так и численно — наименьшей (по сравнению с остальными моделями[34]) величиной суммарного отклонения ε_P=0,54 (см. табл. 2).

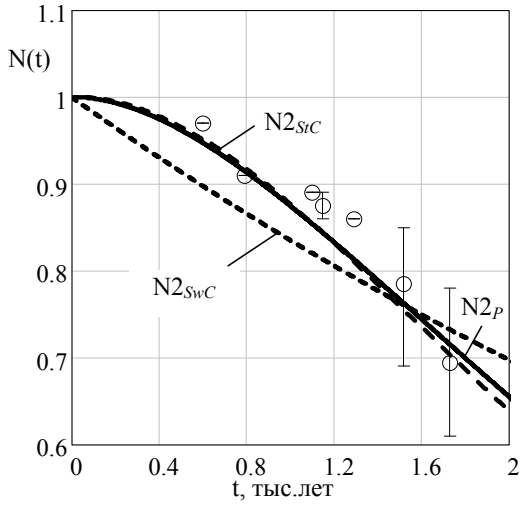
Рисунок 9. Сравнение исходных и калиброванных моделей дивергенции М. Сводеша, С. А. Старостина и потоковой.



а) Исходные модели М. Сводеша ($N2_{Sw}$) и С. А. Старостина ($N2_{St}$)



б) Модели Сводеша ($N'2_{Sw}$) и Старостина ($N'2_{St}$) с коэффициентом λ, калиброванным по данным процесса замен в одном языке



в) Калиброванные по фактическим данным модели Сводеша ($N2_{SwC}$), Старостина ($N2_{StC}$) и потоковая модель ($N2_P$).

[34] Тем не менее, калиброванная модель Старостина, несмотря на выявленные методические недостатки, численно даёт значения, почти идентичные потоковой модели на большей части временного интервала (ср. кривые $N2_{SwC}$ и $N2_P$ на рис.9в), что позволяет использовать её для датирования дивергенции (по крайней мере — в рамках указанного интервала времени).

Таблица 2. Сравнение параметров исходных и калиброванных моделей

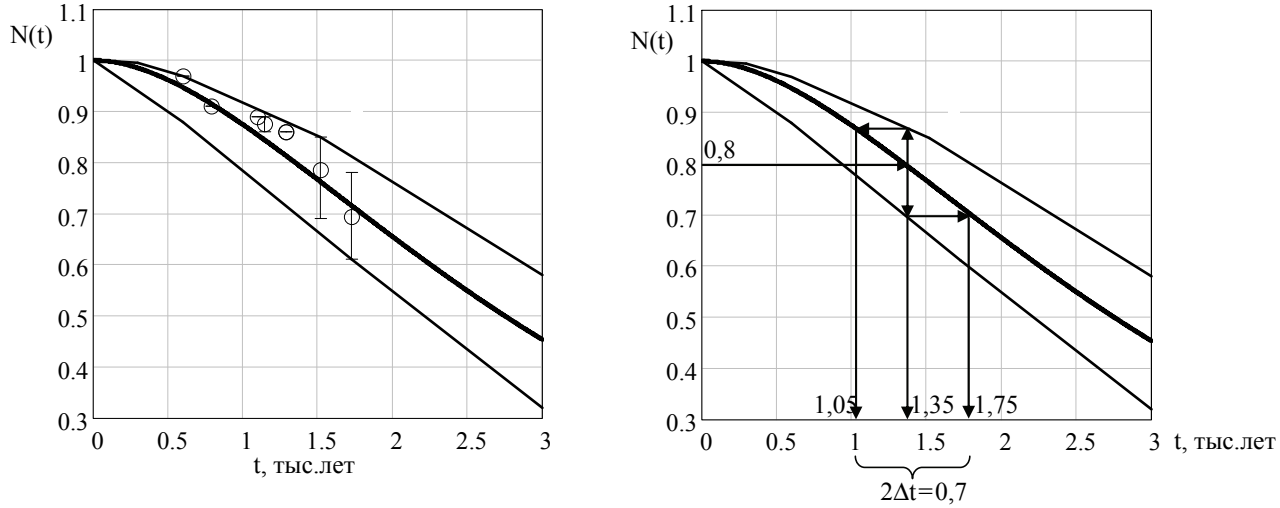
Название и общий вид модели		Исходные параметры модели	Параметры, калиброванные по данным процесса замен в одном языке	Параметры, калиброванные по фактическим данным процесса дивергенции
Модель М. Сводеша	$N2_{Sw}(t) = e^{-2\cdot\lambda\cdot t}$	λ_{Sw}= 0,16 (ε_{Sw}=8,46)	λ'_{Sw}= 0,16 (ε_{Sw}=8,46)	λ_{SwC}=0,09 (ε_{Sw}= 0,76)
Модель С. А. Старостина	$N2_{St}(t) = e^{-2\cdot\lambda\cdot\sqrt{N2_{St}}\cdot t^2}$	λ_{StC}= 0,05 (ε_{St}=1,97)	λ'_{StC}= 0,11 (ε_{St}=7,07)	λ_{StC}=0,07 (ε_{St}=0,72)
Потоковая модель	$N2_{P}(t) = e^{-\eta t}(1+\eta t)$	—		η=0,61 (ε_{P}=0,54)

Завершив сравнение существующих моделей и установив их основные особенности, мы можем перейти к вопросу о теоретической и практической погрешности глоттохронологических датировок, а также их статистической достоверности.

3. Погрешности и достоверность глоттохронологических датировок

Определение погрешностей, возникающих при глоттохронологических расчетах, начнём с оценки фактического разброса значений в исходных данных, которые очевидным образом и будут определять минимальную погрешность наших расчетов (Васильев, 2010: 538; Васильев, Коган: 2013: 156–159). Для этого воспользуемся данными из табл. 1, а также диаграммой (рис. 10а), на которой верхняя и нижняя кривые соединяют соответственно максимальные и минимальные значения процентов совпадений, известные для каждой опорной точки, а средняя линия отражает расчетные значения потоковой модели ($N2_P$). Например, в соответствии с табл. 1, процент совпадений между списками языков, разделившихся 1520 лет назад (точка 7), варьируется в пределах от 69 до 85% (ΔN=16%). Еще больший разброс долей совпадений (ΔN=17%) мы наблюдаем для даты разделения 1730 лет назад (точка 8) — от 61 до 78%.

Рисунок 10



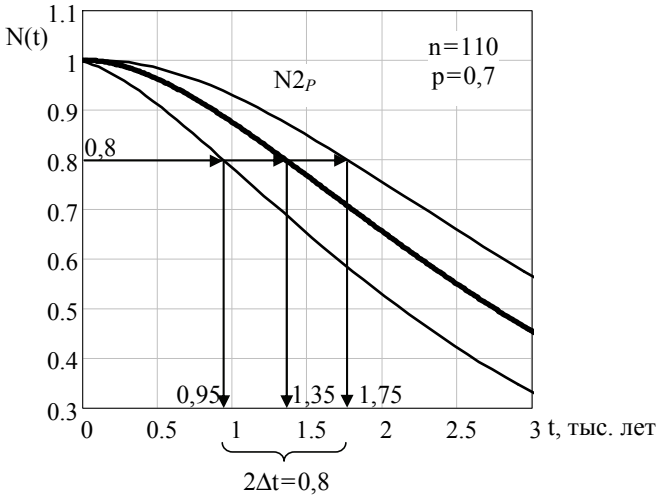
а) иллюстрация разброса фактических долей совпадений по отношению к расчётным значениям N(t), полученным по модели $N2_P$

б) иллюстрация разброса фактических датировок по отношению к расчетным значениям t, вычисленным по модели $N2_P$

Аналогично с помощью диаграммы на рис. 10б можно оценить разброс датировок, вычисленных с помощью потоковой модели для выбранного значения процента совпадений (N). Так, подставляя в формулу $N2_P(t)$ значение N=0,8 (80%), получаем расчетную датировку t=1350 лет назад. При этом, как видно на рисунке, фактическое разделение идиомов могло произойти в диапазоне времени от 1050 до 1750 года — т. е. с разницей в 700 лет. Это означает, что на практике дата разделения рассматриваемых идиомов не может быть определена точнее, чем в диапазоне 1350±350 лет. Таким образом, фактический разброс исходных данных (вне зависимости от используемых моделей) вносит неизбежную и существенную погрешность в результаты любых глоттохронологических расчетов[35].

Величину установленной *фактической* погрешности, связанной со статистическим характером процесса замен, полезно сопоставить с погрешностью *теоретической*, обусловленной особенностями самих используемых моделей. Мерой этой погрешности, применительно к потоковой модели, является величина *доверительного интервала*, описанная нами ранее в первой части исследования (Васильев, Саенко 2016: 274–275)[36]. В частности, доверительный интервал, вычисленный для некоторого известного процента совпадений, позволяет определить временной диапазон, в который с заданной вероятностью укладывается расчетная датировка.

Рисунок 11. Доверительный интервал модели $N2_P(t) = e^{-0,61t}(1+0,61t)$, рассчитанный для 110-словного списка с заданной вероятностью p=0,7.



Например, для доли совпадений 80% и соответствующей ему расчетной датировки 1350 лет назад, теоретическая величина доверительного интервала составляет 800 лет (рис. 11) — т. е. искомая датировка с вероятностью 0,7[37] будет располагаться в диапазоне 1350±400 лет назад.

Как видно на рис. 11 и 12, с увеличением временной дистанции доверительный интервал также увеличивается, однако в процентном отношении его значение убывает по

[35] Причиной такого разброса, как уже говорилось выше, является, с одной стороны случайная природа рассматриваемого процесса дивергенции (см. Васильев, Г. Старостин, 2014: 60), а с другой — невозможность абсолютно достоверного датирования опорных точек по известным историческим событиям.

[36] Методика расчета доверительных интервалов для потоковой модели дивергенции основана на вычислении плотности распределения вероятностей первых замен в списках каждого из языков-потомков (Вентцель, Овчаров 1969: 235–237).

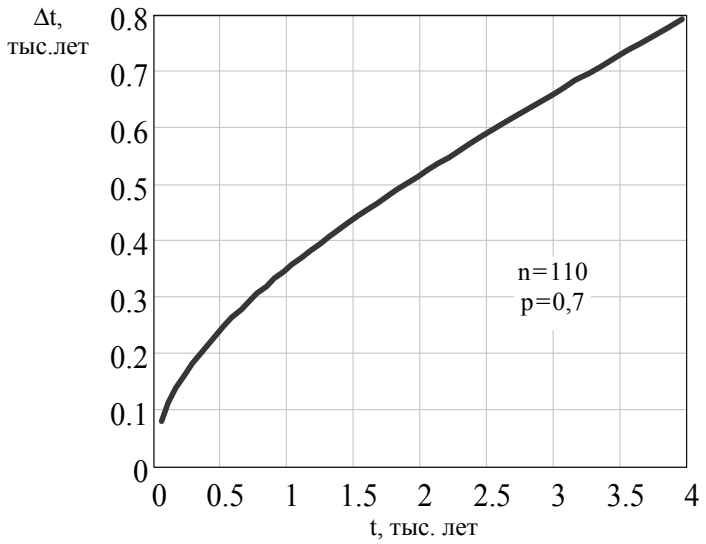
[37] Т. е. в 70 случаях из 100.

мере углубления датировок (табл. 3). Например, в соответствии с таблицей 3, при расчетной дате разделения 1000 лет назад доверительный интервал составляет ±350 лет (т. е. начало дивергенции с вероятностью 0,7 может датироваться от 650 до 1350 лет назад). Аналогично для времени разделения 3500 лет назад получаем доверительный интервал ±730 лет. Таким образом, абсолютная величина доверительного интервала выросла более чем в два раза (от ±350 до ±730 лет), в то время как его относительное значение снизилось с 35% до 21% (табл. 3). Это означает, что, несмотря на уменьшение абсолютной точности, практическая ценность глоттохронологических датировок будет заметно выше при бóльших временных интервалах.

Таблица 3. Значения доверительного интервала, рассчитанные для потоковой модели $N2_P$ с заданной вероятностью 0,7[38]

t, *лет*	200	400	600	800	1000	1200	1400	1600	1800	2000	2500	3000	3500	4000
Δt, *лет*	±150	±220	±270	±310	±350	±390	±420	±450	±480	±510	±590	±660	±730	±790
100Δt/t	75%	55%	45%	39%	35%	32%	30%	28%	27%	26%	24%	22%	21%	20%

Рисунок 12. Изменение величины 70-процентного доверительного интервала (Δt) в зависимости от времени (t) для 110-словного списка

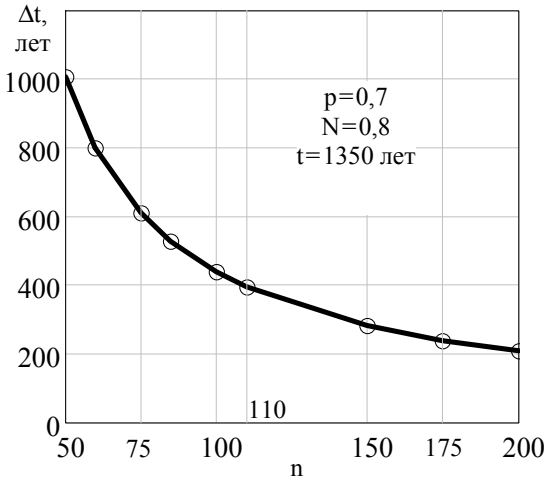


Пользуясь установленными свойствами потоковой модели, мы можем также определить зависимость величины доверительного интервала от количества значений в списках сравниваемых языков. Результаты проведенных расчетов представлены в виде графика на рис. 13.

Так, при использовании двухсотсловных списков для датирования дивергенции двух идиомов с долей совпадений 0,8 мы получаем дату разделения 1350 лет назад с доверительным интервалом ±200 лет (погрешность 15%). При уменьшении размера списка до 110 слов доверительный интервал увеличивается до ±400 лет (30%), а в случае с пятидесятисловным списком достигает значения ±1000 лет (74%!).

[38] Величина доверительного интервала очевидным образом зависит также от выбранной вероятности. Например, при использовании вероятности 0,95, расчетные значения доверительного интервала увеличатся примерно в два раза.

Рисунок 13. Изменение ширины доверительного интервала (Δt) в зависимости от числа лексических значений (n) в списках сравниваемых языков (для времени дивергенции t=1350 лет)



Полученная зависимость Δt(n) указывает на бесперспективность использования для глоттохронологического анализа коротких списков, что, однако, не умаляет полезности этих списков при установлении генеалогических связей между языками.

Сопоставляя между собой рис. 10 и 11, а также полученные нами расчетные значения, несложно убедиться в том, что величина доверительного интервала (при выбранной вероятности 0,7) лишь незначительно превышает фактический разброс исходных данных на рассматриваемом интервале времени. Следовательно, мы можем предположить, что решающее значение при оценке общей точности глоттохронологических датировок будет иметь именно эта объективная погрешность. Проверим справедливость нашего предположения на конкретных примерах дивергенции между языками романской группы, а также некоторыми другими языками.

В первой части табл. 4 приведены данные для нескольких пар идиомов с предположительной датой разделения 480 г. н.э. При этом средний процент совпадений между их списками варьируется от 75% (между португальским и галло-романскими) до 87% (между фриульским и лигурийскими). Расчетные датировки, полученные для этих значений с помощью потоковой модели составляют 410 и 960 г. соответственно. Таким образом, диапазон разброса фактических значений для всей группы из 14-ти романских языков составил 550 лет (или ±275 лет), что с запасом «укладывается» в теоретический доверительный интервал ±415 лет, вычисленный для среднего значения совпадений N=78,5 (см. табл. 1) и вероятности p=0,7. При рассмотрении отдельных пар языков (см. выделенные строки табл. 4) в 3-х случаях из 14-ти, (т. е. в 22% случаев) отклонение фактических дат распада от расчётных значений выходит за рамки 70-процентного доверительного интервала, что также согласуется с теоретической оценкой его статистической значимости. Так, большинство пар с участием фриульского дают сильно завышенные проценты совпадений, что приводит к «омоложению» расчетных датировок почти в два раза по сравнению с предполагаемой датой разделения[39]. Например, для фриульского и лигурийского с долей совпадения 86,7% получаем дату 960 г. с доверительным интервалом ±310 лет, в который очевидным образом не укладывается фактическое значение 480 г. В то же время для большинства остальных пар величина доверительного интервала оказывается избыточной, а отклонение расчетных датировок от фактической

[39] Заметим, впрочем, что подобный «подскок» значений может также объясняться более поздним отделением фриульского от сравниваемых с ним идиомов.

Таблица 4. Даты дивергенции языков, а также их доверительные интервалы, рассчитанные на основе потоковой модели $N2_P$.

Сравниваемые языки	Средний % совпадений	Фактическая датировка (лет)	Расчётная датировка (лет)	Доверительный интервал, p=0,7 (лет)
Руманшские — лигурийские	78,8	480	590	±410
Руманшские — сицилийские	80,7	480	680	±380
Руманшские — португальский/галисийский	78,3	480	570	±420
Руманшские — галло-романские	77,1	480	520	±430
Лигурийские — португальский/галисийский	80,3	480	660	±400
Лигурийские — галло-романские	77,0	480	510	±430
Сицилийские — португальский/галисийский	81,5	480	720	±370
Сицилийские — галло-романские	80,2	480	660	±400
Португальский/галисийский — галло-романские	74,7	480	410	±460
Фриульский — руманшские	84,3	480	850	±350
Фриульский — лигурийские	86,7	480	960	±310
Фриульский — сицилийские	86,8	480	960	±310
Фриульский — португальский/галисийский	83,0	480	790	±360
Фриульский — галло-романские	82,0	480	740	±370
Южнославянские — восточнославянские[40]	77,0	480	510	±430
Путунхуа — миньские идиомы[41]	63,5	–110	–90	±580
Балкано-романские — основной массив романских	69,6	271	180	±520

не превышает 200 лет, (см., например, рис. 14), что подтверждает адекватность используемой модели и её параметров рассматриваемому процессу дивергенции.

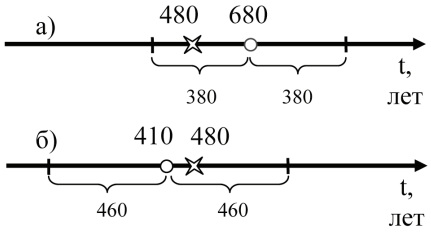
Безусловно, более показательной (и методически корректной) была бы апробация модели на другом языковом материале (который ранее не использовался при её калибровке) и на других интервалах времени. В качестве примера можно привести результаты датирования дивергенции китайских, славянских, а также балкано-романских языков (последние три строки табл. 4). Во всех трех случаях полученные датировки оказались очень близки к предполагаемой фактической дате разделения. Подобные примеры сви-

[40] В сравнении участвовали списки орбаничского чакавского, градищанского кайкавского и люблянского словенского, с одной стороны, с туровским белорусским и деулинским русским — с другой.

[41] Использовались списки путунхуа, цзяньоу и хайнаньского, составленные Г. С. Старостиным и Е. А. Кузьминой. Проценты совпадений для обоих примеров приводятся по данным из «Глобальной лексикостатистической базы данных», представленным на сайте проекта http://starling.rinet.ru/new100/main.htm (по состоянию на 01.05.2017) .

детельствуют о том, что эффективное использование полученной модели в теории не ограничено только романскими языками или определенным временным отрезком.

Рисунок 14. Иллюстрация взаимного расположения некоторых фактических и расчётных дат распада относительно доверительных интервалов: а) руманшские — сицилийские; б) португальский/галисийский — галло-романские.



Таким образом, точность глоттохронологических расчетов определяется в первую очередь не свойствами моделей, а случайным характером процесса лексических замен, который проявляется в существенном разбросе фактических долей совпадений, полученных для пар языков с одинаковыми интервалами распада. Величина этого разброса вносит основной вклад в конечную погрешность получаемых датировок.

Заключение

Подводя итоги, сформулируем основные результаты проведенного исследования в виде нескольких обобщающих выводов и положений:

1. Сравнительный анализ существующих глоттохронологических методов показывает, что наилучшие результаты при датировании процесса дивергенции достигаются при использовании модели С. А. Старостина и потоковой модели (после их предварительной калибровки). При этом попытки построения моделей дивергенции на основе данных общего распада, как это подразумевается методикой М. Сводеша и С. А. Старостина, приводит к абсурдным результатам и указывает на несостоятельность используемого в них постулата Сводеша о независимом развитии языков-потомков после разделения. Таким образом, моделирование процессов диверегенции должно учитывать возможность согласованного изменения в лексике родственных языков, при котором в списках разделившихся идиомов происходят замены одних и тех же значений.
2. Калибровка рассмотренных моделей по исходным данным позволила добиться хорошего численного совпадения расчетных и фактических датировок. При этом отдельные примеры показывают, что калиброванные модели могут эффективно применяться для датирования языковой дивергенции в других языковых семьях и на различных временных глубинах.
3. Точность глоттохронологических расчетов определяется в первую очередь не свойствами моделей, а вероятностным характером процесса лексических замен, который выражается в существенном разбросе фактических значений, величина которого и вносит основной вклад в конечную погрешность получаемых датировок. В силу случайного характера лексических замен определение времени разделения языков возможно только в пределах некоторого доверительного интервала с заранее выбранной вероятностью попадания фактической даты в этот интервал. Таким образом, корректная датировка дивергенции двух идиомов должна представлять собой не конкретное значение, а интервал значений с соответствующей величиной вероятности. Например, вместо «1000 лет назад», следует указывать «1000±350 лет назад с вероятностью 70%».

4. Теоретическая оценка доверительных интервалов, полученная на основе моделирования процесса дивергенции в виде потока лексических замен, позволила установить, что по мере увеличения времени дивергенции относительное значение этого интервала уменьшается и стремится к некоторому постоянному значению. Например, для заданной вероятности p=0,7 и периоде дивергенции 500 лет доверительный интервал составляет ±50% от этого периода, при 2000 лет — 26%, а к 4000 лет приближается к 20-процентному уровню.
5. Сравнение теоретических погрешностей моделей с фактическим разбросом известных данных, полученных для романских языков, свидетельствует о том, что на временном интервале до 2 тыс. лет погрешности датировок, вызванные случайным характером замен, являются доминирующими и носят объективный характер — т. е. не могут быть существенно снижены (в статистическом смысле) за счет дальнейшего уточнения стословных списков или привлечения дополнительных данных.
6. Установленная зависимость ширины доверительного интервала от числа лексических значений в списках сравниваемых языков показывает, что при расширении списка значение доверительного интервала пропорционально уменьшается. Например, при периоде дивергенции 1350 лет величина доверительного интервала для 200-словного списка в два раза меньше, чем для 110-словного. Таким образом, увеличение размера списков в теории позволяет существенно повысить точность глоттохронологических расчётов.
7. Дальнейшее повышение теоретической точности и надёжности глоттохронологических моделей возможно в первую очередь за счет привлечения дополнительных данных (опорных точек) для калибровки моделей на материале различных языковых семей на разных временных глубинах.

Литература

Алексеев, А. Н. 2013. Ранние кочевники в Якутии. Вестник Северо-Восточного федерального университета им. М. К. Аммосова 5(10): 62—69.

Арапов, М. В., М. М. Херц. 1974. Математические методы в исторической лингвистике. Москва: Наука.

Багаев, М. Х. 2015. К вопросу об этнокультурной общности на северо-восточном Кавказе в VIII–IV тыс. до н.э. В.: Н. Ярычев (ред.). 4-я ежегодная итоговая конференция профессорско-преподавательского состава Чеченского государственного университета. 28 февраля 2015 г.: 126–128. Грозный: Чеченский государственный университет.

Васильев, М. Е. 2010. Об использовании лексического критерия для построения генеалогической классификации. В: З. М. Шаляпина (ред.). Востоковедные чтения 2008. Бюллетень обшества востоковедов РАН. Тɒvды межинститутской научной конференции. Выпуск 17: 530–572. Москва: Институт востоковедения РАН.

Васильев, М. Е., А. И. Коган. 2013. К вопросу о восточнодардской языковой общности. Вестник РГГУ. Серия: Филология. Вопросы языкового родства. № 16 (117): 149—177.

Васильев, М. Е., А. Ю. Милитарев. 2008. Глоттохронология в сравнительно-историческом языкознании. Модели дивергенции языков. Orientalia et Classica: Труды Института восточных культур и античности 19: 509—536.

Васильев, М. Е., М. Н. Саенко. К вопросу о точности глоттохронологии: датирование процесса лексических замен по данным романских языков. Вестник РГГУ. Серия: Филология. Вопросы языкового родства. 14(4): 259–278.

Васильев, М. Е., Г. С. Старостин. 2014. Лексикостатистическая классификация нубийских языков: к вопросу о нильско-нубийской языковой общности. Вестник РГГУ. Серия: Филология. Вопросы языкового родства. № 16 (138): 51—72.

Вентцель, Е. С., Л. А. Овчаров. 1969. Теория вероятностей. Москва: Наука.

Иллич-Свитыч, В. М. 1966. Мнимые и действительные возможности лексикостатистики. В: Основные проблемы эволюции языка: 160–162. Самарканд: Фан.

Клейн, Л. С. 2010. Время кентавров. Степная прародина греков и ариев. С.–Петербург: Евразия.
Нарумов, Б. П. 2001. Арумынский язык/диалект. В: И. Челышева (ред.). Языки мира. Романские языки: 636–656. Москва: Academia.
Сводеш, М. 1960. Лексикостатистическое датирование доисторических этнических контактов. Новое в лингвистике 1: 23–52.

References

Alekseev, A. N. 2013. Ranniye kochevniki v Yakutii. Vestnik Severo-Vostochnogo federal'nogo universiteta im. M. K. Ammosova 5(10): 62—69.
Arapov, M. V., M. M. Herz. 1974. Matematicheskiye metody v istoricheskoy lingvistike. Moskva: Nauka.
Bagaev, M. Kh. 2015. K voprosu ob etnokul'turnoy obshchnosti na severo-vostochnom Kavkaze v VIII–IV tys. do n.e. In: N. Yarychev (ed.). 4 ezhegodnaya itogovaya konferenciya professorsko-prepodavatel'skogo sostava Chechenskogo gosudarstvennogo universiteta 28 fevral'a 2015 goda: 126–128. Grozniy: Chechenskiy gosudarstvennyj universitet.
Illich-Svitych, V. M. 1966. Mnimye i deystvitel'nye vozmozhnosti leksikostatistiki. In: Osnovnye problemy evolyutsii yazyka: 160–162. Samarkand: Fan.
Klein, L. S. 2010. Vremya kentavrov. Stepnaya prarodina grekov i ariev. S.-Petersburg: Evraziya.
Narumov, B. P. 2001. Arumynskiy yazyk/dialekt. In: I. Chelysheva (ed.). Yazyki mira. Romanskie yazyki: 636–656. Moskva: Academia.
Starostin, S. 2000. Comparative-historical linguistics and lexicostatistics. In: Colin Renfrew et al. (eds.). Time Depth in Historical Linguistics: 233–259. Cambridge: McDonald Institute for Archaeological Research.
Swadesh, M. 1960. Leksikostatisticheskoye datirovaniye doistoricheskikh etnicheskikh kontaktov. Novoye v lingvistike 1: 23—52.
Vasilyev, M. E. 2010. Ob ispol'zovanii leksicheskogo kriteriya dlya postroyeniya genealogicheskoy klassifikatsii. In: Z. Shalyapina (ed.). Vostokovednyye chteniya 2008. Bvulleten' Obshchestva vostokovedov RAN. Trudy mezhinstitutskoy nauchnoy konferentsii 17: 530-572. Moskva: Institut vostokovedeniya RAN.
Vasilyev, M. E., A. Yu. Militaryov. 2008. Glottokhronologiya v sravnitel'no-istoricheskom yazykoznanii. Modeli divergentsii yazykov. Orientalia et Classica: Trudy Instituta vostochnykh kultur i antichnosti 19: 509—536.
Vasilyev, M. E., A. I. Kogan. 2013. K voprosu o vostochnodardskoy yazykovoy obshchnosti. Journal of Language Relationship 10: 149–177.
Vasilyev, M. E., G. S. Starostin. 2014. Leksikostatisticheskaya klassifikatsiya nubiyskikh yazykov: k voprosu o nil'sko-nubiyskoy yazykovoy obshchnosti. Journal of Language Relationship 12: 51–72.
Vasilyev, M. E., M. N. Saenko. K voprosu o tochnosti glottokhronologii: datirovaniye protsessa leksicheskikh zamen po dannym romanskikh yazykov. Journal of Language Relationship 14(4): 259–278.
Venttsel, E. S., L. A. Ovcharov. 1969. Teoriya veroyatnostey. Moskva: Nauka.

Mikhail Vasilyev, Mikhail Saenko. How accurate can glottochronology be? Dating language divergence on the basis of Romance data.

The paper is a sequel to an earlier study by the authors, in which they discussed the accuracy of linguistic datings arrived at by the glottochronological method on the basis of data from 110-item wordlists for Romance languages. The object of this second part of the study is the dating of linguistic divergence, i.e. determining the separation dates for two or more modern languages. In this paper, we compare several traditional as well as newly offered models for the glottochronological process, with special attention paid to the margin of error and reliability of glottochronological calculations on different time depths. The results of the study allow for a realistic assessment of the degree of accuracy in the glottochronological dating of the divergence of Romance languages and lead to a number of practical conclusions that will be useful for the application of glottochronology to any other linguistic material.

Keywords: glottochronology, lexicostatistics, Swadesh wordlist, Romance languages.

Book Reviews / Рецензии

Luka Repanšek
Department of Comparative and General Linguistics, Faculty of Arts, University of Ljubljana; luka.repansek@ff.uni-lj.si

Blanca María Prósper.
The Indo-European Names of Central Hispania.
A Study in Continental Celtic and Latin Word Formation.

Innsbrucker Beiträge zur Sprachwissenschaft (herausgegeben von Wolfgang Meid), Bd. 152.
Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck (Bereich Sprachwissenschaft), 2016. 237 pp.

Over the last two decades or so increasing interest has been noted in the study of the onomastic linguistic remains of Indo-European Europe and Asia Minor. Paleolinguistic data of otherwise poorly attested Indo-European languages (or, for that matter, linguistic systems that are known solely through the names and naming systems that have left an imprint on a given *Namenlandschaft*) has of course always been studied but never has the opportunity been greater to approach this ultimately uncompromising and extremely sensitive set of data with the quickly expanding knowledge that contemporary Indo-European comparative linguistics has to offer. Not only is it now becoming possible to refine and substantiate old etymologies, refute the old ideas and promote alternative, methodologically decidedly less reproachable solutions, or find convincing linguistic explanations for the here-to unetymologisable linguistic data, but also — and this is all the more important — correct the word-formational patterns projected back into the proto-language *on the precious evidence of just such fragmentary pieces of evidence*, sometimes even contributing to the established set of lexical items reconstructed for the parent language (such opportunities are of course comparatively rare and about ninety per cent of the onomastic material will as a rule be explicable on grounds of what we already know or hold for well-established on the basis of the comparative data (mostly appellative) offered by the Indo-European corpus languages). On a smaller scale, however, the onomastic material of a given linguistic system is able to provide valuable missing puzzles in the understanding of the historical development of a particular language family or one of its individual daughters, especially given the fact that 1) the onomastic systems are repositories of often residual linguistic features, and 2) being essentially generated by the non-onomastic sphere of language use, names are fundamentally words (a fact that is perhaps too often unrespected) and as such reflect in all the details the phonological, morphological, word-formational, syntactic, and lexical peculiarities of a given language. Names therefore demand a careful and well-balanced etymological explanation that will assign the correct coordinates on all the relevant levels of linguistic expression. One must never neglect the crucially important fact, however, that the onomastic system of a given language, while it does indeed feed upon the appellative sphere of use, obeys its own rules in terms of the productivity on the level of the word-formational patterns and to a minor extent lexis (and rarely even morphology). The generally unavoidable and natural analogical processes will consequently chose different trajectories and affect different segments of language.

B. M. Prósper's monograph is an attempt at extracting as much information as possible from the selected corpus of anthroponymy to, first and foremost, provide the scholarship with a glimpse into the linguistic diversification of North-Western Hispania that is otherwise difficult or, in parts, impossible to track on the basis of the existing epichoric Celtiberian textual documents (mostly, of course, because these are rather earlier than the data embedded in the Latin inscriptions), and, second, to detect potential residual features of Hispano-Celtic that may offer an interesting insight into the word-formational make-up of the older layers of Celtic. The author is to be especially commended on her conscientious application of the premises that were pointed out above, subjecting every studied piece of evidence to multi-faceted etymological analysis. Whenever a particularly tentative suggestion is made to account for a given (mostly sporadic) sound change that otherwise receives no or very little back-up from the immediately relatable language material (due to sheer scarcity of the relevant data), the solution is supported and exemplified by typologically comparable instances from better documented languages, which is an extremely welcome and methodologically indeed necessitated increment. Several of the proposed etymologies are

rather convincing and well-grounded in the framework offered in each individual case by the internal and external comparative data. In the majority of cases, of course, even the likeliest interpretation will still remain rather tentative and ultimately purely provisional as is to be expected in any work dealing with etymological onomastics, but a solid starting point is a good stepping stone towards future refinement (this usually becomes possible when more data comes to light or old data receives an updated explanation).

The book is divided into two sizable chapters which contain a condensed and contextualised version of the ideas previously expressed and elaborated by the author in a number of separate studies, supplemented by several novel suggestions and discussions. *The Names of the Celtic Cantabri* (pp. 11–122) studies the anthroponymical heritage of the Celtic population to the west of the region dominated by the Celtiberian inscriptions, while *The Names of Western Celtiberia* (pp. 123–198) seeks to pinpoint the individualistic traits and/or dialectal differences potentially mirrored in the personal names of the belt between Burgos, Segovia, Soria, Guadalajara, and Cuenca. The second part of the book is organised as a lexicon of alphabetically arranged names that have been conveniently grouped together according to the place of their concentration or, if assignable, the appertaining ethnic (p. 124: *Pelendones* and *Turmogi*, p. 128: *Autrigones*, p. 144: *Arevaci*, and p. 180: names concentrated around Cuenca). The first chapter is more significantly structured and studies selected personal, and to a lesser extent also ethnic and divine names (most notably *Cabuniaeginus* and *Erudinus*, pp. 118–119), on the basis of a particular feature — be it on the level of morphology, word formation, or historical phonology — that the author considers important to isolate and expose to closer scrutiny. In this way the book is able to provide a number of neatly integrated (however miscellaneous) specialist studies on several important aspects of Hispano-Celtic, or generally Celtic, historical development: the (older layers of the) Celtic numeral system (pp. 15–21), the idiosyncrasies of the Celtic comparative and superlative formations (pp. 96–100), the still somewhat problematic question of the specifically Celtic continuation of PIE **k'er-(H₂)(-u-)* 'horn & c.' (pp. 21–26), some overlooked cases (the autor's choice of the word "neglected" here is perhaps less appropriate) of departicipial formations (pp. 26–33) and obscured compound names (pp. 51–58), the survivors of the PIE category of holokinetic *t*-stems (pp. 58–65), the history of the verbal adjectives in **-eto-* (pp. 71–87) and the surviving instances of possessives containing the Hoffmann suffix (**-$H_{1/3}$en-*, pp. 87–96), secondary formations based on nasal (*passim*) and sigmatic stems (pp. 111–115), dissimilation of geminates (such as, e.g., **nn* > **nd*, pp. 65–71), metathesis in *LVP* clusters (pp. 101–102), anaptyxis in *VRPRV* sequences (pp. 102–104), and a few scattered bits and pieces of provisional but insightful comments on various heterogeneous problems of historical phonology and/or morphology. Although the title of the monograph indicates that specific problems of Latin word formation will also be addressed somewhat extensively, this is not in fact among the central foci of the monograph. Much Italic comparanda is, admittedly, adduced in support of the author's claims on a particular etymological interpretation, but the book is as much a study of Latin (or Italic, for that matter) word-formational patterns as it is more generally a contribution to the understanding of the somewhat still problematic points of PIE word formation. This is of course the expected side-effect of the study of that side of the language that does not normally take part in the process of reconstruction and may therefore have the noteworthy value of being able to refine or even correct what has been projected back and reconstructed for the parent solely on the basis of the appellative data. What needs to be called to the reader's attention in this respect, however, is the relatively long and extremely interesting and insightful excursus on the history of the type of *-ilo-* adjectives in Celtic and here, specifically, Latin.

Both chapters conclude with synoptic sections on synchronically productive word-formational patterns (i.e. predominantly suffixes and suffixal chains) of the surveyed names as well as their "phonetic" peculiarities. This, however, is perhaps the most problematic part of the monograph. Even though conditioned sound changes are expected to sporadically occur (a good example is, perhaps, *Murce*, p. 143, if from **morko-*, with conditioned raising of **-o-* immediately comparable with the equally sporadic ven. *mu̲rtuvo.i.* < **morto-*) in the attested names (when such phenomena are not in fact just a by-product of the rendition of epichoric, native sound sequences in the Latin script), these are more or less as a rule assigned systemic value by Prósper, who tends to take them at face value (occasionally perhaps somewhat too uncritically) and parallels the proposed developments and their results with rather individualist views of the sound changes underwent by the language(s) displayed in the native Celtic text documents conducted in the Iberian (and, to a smaller extent, Latin) script. Too often, perhaps, a particular sound change is proposed to have occurred on the basis of the author's own view of an etymological source behind a name/group of names. Highly

speculative is in my view the proposed reduction *#u̯eR- > #uR- (pp. 111 and *passim*), which by the way is an old idea, based on the names such as *Vrcaloco*, *Vurovio* and *Avlgigun* and supported by CIb. *urantiom* as if from **up-ero-*, none of which can be irreproachably claimed to actually contain the addressed sequence (note that the etymological connection between Lusitanian *Uramus* and CIb. *Veramos* cannot be proved in any significant way). It is a staple fact of historical comparative linguistics that etymology of a given word in any given language is the bedrock foundation upon which a set of regular (and conditioned) sound-changes can be observed and established (combinatorially, of course, and using forward as well as backward reconstruction). This customary, although demanding procedure logically receives a methodological caveat: the etymological connection has to be irrefutable for the results to obtain. There is an immediate problem with the onomastic data, however. Regardless of the progress made in the direction of successfully approaching the fragmentary linguistic evidence, names still often prove to be ultimately difficult if not momentarily impossible to subject to exact interpretation, which is simply due to their general opacity, brought about by too many points of contact with potentially promising formal correspondences and simultaneous lack of purely synchronic transparency (this is more often than not a rule for onomastic languages but not uncommon in the case of fragmentarily or otherwise poorly attested systems). Whenever an individually observed sound change is supported by a comparandum with ultimately uncertain etymology, the reader should have been warned that the author is basing her views on her own individual interpretation of a particular piece of data and not in fact on a substantiated piece of evidence as seems to be the impression. Several of author's points on such proposed phonological developments should therefore perhaps be understood as very tentative and provisional. I remain very sceptical towards several of the suggestions, especially towards the proposed voicing of stops preceded by nasals (cf. the Old Irish type *-*ant-* > *-*ænd-* > *-*ẽd-*) in the likes of *Pi(n)ganco*, *Letondo*, *Plandica* etc. (pp. 185–190 and *passim*). If voicing were a late systemic sound-change, it should affect all instances of such sequences, which it clearly does not, exempting the *-nt-*participles (unless, as it is argued, obscured participial formations) and the productive suffix *-*Vnko-*. As far as I can see, there is not a single incontestable and unambiguous case of a *-nP-*sequence in the material adduced in favour of the sound law and neither would I be too eager to recognise the numeral five in the likes of *Pi(n)ganco* & co. That such regular voicing would be hindered by the "palatalising effect" of the following *i̯ (as, interparadigmatically, in *stenionte* and *gente*, as is suggested) and that names built around **arganto-* that never show voicing do not in fact go back to the commonly accepted thematisation of the present participle seems like special pleading. In light of the unproblematic fact that at least Celtiberian attests to the process of phonetic lenition of voiced stops I wonder if the (surely telling) spelling of etymological medial *-*g-* in *Dahae* and perhaps *Saihli* as >h< does not rather simply encode the voiced velar fricative rather than its secondary devoicing as suggested by Prósper (pp. 139, 184). A rather strong case is also made in favour of gemination as a direct systemic consequence of a phonological process. This is likely in case of *-*Ri̯-* clusters, where it is even typologically expected, even though the data forces one to simultaneously accept the somewhat suspicious (because strangely sporadic) accompanying glide absorption (note, however, the potentially interesting case of subsequent dissimilation to *-rd-* < *-*rr-* < *-*ri̯-*, pp. 70–71, 120), but I cannot see a convincing reason to favour gemination as a purely phonological process over hypocoristic gemination in cases such as *Accua*, *Pedaccianus*, *Boddi* etc. It is moreover rather difficult to accept the idea behind the proposed development of *-*Vpn-* > *-*Vϕun-* > *-*Vβun-* (pp. 105, 118), since PCelt. *ϕ is otherwise never voiced intervocalically and is normally lost without trace. I am undecided on the late change of the inherited voiceless labiovelar into *p (*passim*) as potentially mirrored by *Petraioci*, *Pentius* & co. (the textual documents of course clearly attest to the preservation of *k^u). These names are extremely likely to go back to the obvious numerals that they contain in the derivational base, but how sure can one be that they are ultimately Celtic? Note that the divine name *Vailico ~ Vaelico* (p. 182) is said to preserve the otherwise regularly monophthongised inherited diphthong *ai̯ on account of its being of onomatopoeic origin (cf. OIr. *fáel* 'wolf'). I cannot see, however, how an inherited lexical item, regardless of its etymological source, would be able to resist a regular sound change. It remains unclear what the author's views are on the probable simplification of *-χt- cluster in the seemingly popular name built on *Ambato-* < PCelt. **amb-aχto-*. On p. 125 it is described as regular and expected, whereas in ft. 65 (p. 73) the development is said to be surprising. The supposed metathesis in *Crastunon-* (p. 160) is despite a good appertaining discussion left unaccounted for in the end.

There are an additional few minorly problematic points I would like to draw attention to. The PN *Carauanca*, if it is indeed related to the PIE word for

'horn' & c. (p. 23ff.), which seems more than likely, could equally well reflect a possessive *-u̯o- derivative, so *k'er=H$_2$-u̯o-, cf. Gr. κερᾱός < *k'er=H$_2$=s-(u̯)ó-. The mountain ridge Καρουαγκα(ς) is certainly non Celtic (p. 25). I have recently explained it (Repanšek 2016c: 187–188) as reflecting *(s)kor=u̯n̥-ko- (cf. OIr. *lië* < *lēH$_2$=u̯n̥-ko-) to *(s)ker- 'split' (for the secondary semantic shift towards a nomen rei actae cf. PSl. *skala* 'rock'). Balto-Slavic *kárṷā (p. 25–26) is undoubtedly a *vr̥ddhi* formation (and as such a formal substantivisation of the underlying possessive adjective) but only in as much as it copies the naturally co-occurring metatony in the inherited type (cf. the Slavic type *u̯ýdra to *ud=r-ó-); *kōru̯ā is therefore an unjustified projection and does not as such "fail to account for the Celtiberian form" (p. 26). Hittite *makkiešš-* 'become big' can hardly be convincingly traced back to *m^{e}ǵ=H$_2$-eH$_1$sH$_1$- (p. 27) — a projection that strives to account for the exclusive geminate spelling of the intervocalic -šš-. Such fientives are synchronically most probably based on the established model *palḫ-ešš- (adj.) (*cf. *palḫ=ešš-^{ar}/$_{n}$-*) → *palḫ-ešš- (fient.) and ultimately reflect simple conversions. In terms of word formation, the type continued by Latin *senēsce/o-* etc. < *-e-H$_1$- + *-sk'e/o- (*ibid.*) is of course completely unrelated. Vedic *mahī-yá-^{te}* 'to be/feel big' is a deadjectival denominative verb and as such goes back directly to a straightforward *m^{e}ǵ=H$_2$-i̯é/ó- (with regular and morphophonetically conditioned lengthening of the reflex of *schwa primum* before the suffix) rather than indirectly reflect an "older *-ei̯/i-" (p. 35). There is absolutely no reason to uphold Hamp's view that the PN *Brigetio* is of deverbative origin (p. 52; see Repanšek 2016a: 248). The sequence *-ǵn̥n-i̯o- < *-ǵn̥H$_1$-i̯o- would certainly not have had a different outcome (purportedly *-gn̥i̯o- > *-ganiio-, p. 54) than the ubiquitous type *-ǵn̥n-o- (< *-ǵn̥H$_1$-o-) > *-gn-o- in Italo-Celtic; an inherited *-ǵn̥H$_1$-i̯o- (cf. OIr. *búachaill* < *-i̯o-) that would preserve the laryngeal intact, on the other hand, would indeed produce PCelt. *-ganio- (via laryngeal loss by what is descriptively known as Pinault's rule), logically matching the simplex. The PN *Adnamatia* in Pannonia is formally a substantivised adjective of appurtenance to *Adnamato-* and could under no etymological approach to the root in question come to mean "the frightened city" (p. 85). Incidentally, the PN *Adnomatus* from Ig (sic!) should be properly said to *indirectly* reflect the length of the *ā in its Gaulish donor, given that this was phonetically most probably realised as a low rounded */ɒ/, cf. such spellings as Gaul. *Blotu-rix* for *blātu-* (see Sims-Williams 2003: 56). I do not share Olsen's views on the origin of the Hoffmann suffix and I do not find the proposed semantic relationship between the derived and underived versions at all convincing (pp. 87–96) — we must rather simply be dealing with a complex suffix with the basic function of deriving from the nominal base a possessive adjective (liable to subsequent formal substantivisation). The alleged cases of *-H$_{1/3}$n-o- should, however, probably be segmented differently (specifically *-H$_1$-no-), as has already been proposed. I am cautious to accept admittedly interesting cases of *-āno- as reflecting the old, basically unshortened version of *-o-H$_{1/3}$n-o-, because this seems to significantly complicate the traditional (and in my view rather convincing) explanation for the "normal" and ubiquitous type in *-ono-, especially since cases such as Gaul. *Toutanno-* could easily be secondarily built on the inherited *ā*-stem and thus represent a younger parallel to the inherited *Toutono-* < *-o-H$_{1/3}$no- (with regular laryngeal loss by Dybo's law) ← *-e-H$_2$-. I am not convinced that pairs such as *Aiu* (PN) vs. *Aiankum* (family name) can in fact reflect an old relationship *-ō(n)* : *-n̥-ko-, since *Aiu* is cleary an *u*-stem, cf. the Gaul. hypocoristic *Aiiuca* (see Meid 2005: 213). Latin *patrōnus* and its oppositional derivative *matrōna* (p. 92) are almost definitely not old inherited formations, neither is the apparent thematic base of Av. *vīsān-* (as per Olsen 2010: 160–161), which simply copies the model established by the predominance of the *puϑrān-* type. *Cormerton-*, if it indeed goes back to *kom-merton-, is hardly a case of a Hoffmann-derivative, *merto-n- (most notably in Av. *mar^{ə}tan-*, which only means 'mortal'; there is no conclusive piece of evidence that would point to a homonymous *mar^{ə}tan-* with the meaning of 'chief of men' in the GāϑāS) being a clear case of an individualisation. Note that the divine name *Vidasus* is certainly Pannonian rather than Celtic (p. 113), i.e. Gaulish. PCelt. *ulk^{u}o- for PIE *u̯lk^{u}o- 'wolf' (OIr. *olc*, perhaps = Lepontic *Ulkos*) is in my view a case of resyllabification rather than a final stage of the proposed developmental stage *u̯ulk^{u}o- (p. 115), cf. Old Albanian *ulk* and Pannonian *ulko- (in *Ulcisia*), going towards the same end as *luk^{u}o- with full metathesis of the *daḱru- type. Consequently, I find it extremely unlikely that the PN *Vlibagi* could conceal the expected PCelt. reflex *u̯lik^{u}o- (> *u̯lipo- > *u̯libo-). The PN *Voltisemae* should not be simply called Italic (p. 154, ft. 125) as the relationship between the reflex of the sequence *-m̥H$_2$o- that the name attests to, namely *-am-* (with expected, even though sporadically marked vowel weakening in an unaccented syllable), exactly matching the sequence *-am-* >am<, >em< amply attested in Ig (there the PN *Decomon-* is not autochthonous), vs. PItal. *-om-, for which consider Ven. *dekomo-* '10th', points to the fact that things are

significantly more complicated. The phenomenon actually seems to reflect an important isogloss that brings in further (and rather welcome) internal diversification within the Northern-Adriatic language continuum (see Repanšek 2016b: 337 and a much updated view in id. 2017). I have trouble accepting the claim that the data seems to point towards the "reconstruction of a single Celtic and Italic Suffix *-ed(i)i̯o-*" (p. 164); this would leave **-o-di̯o-* (the latter morphemic segmentation is dictated by deadverbials such as Gaul. **uχsedi̯o-* & co., in my view also by the PN *Remetodia* < **-eto-di̯o-*, for which consider ουερ-ετο-μαρε[ο]υι), the widespread variant found in Gaulish (matching OIr. *-(ai)de*, W *-eid*), completely unaccounted for. OIr. *búachaill*, MW *bugeil* do not represent transfer forms to the *i*-stems (p. 166), but regularly and unproblematically reflect old, inherited agent nouns in plain monosyllabic **-i̯o-* (cf. Uhlich 1993). Gaul. *neððamo-* (*/tˢ/) = OIr. *nessam* reflect the expected deadverbial superlative **nesd-tm̥H$_2$o-* (cf. Indo-Iranian **nazd-*) rather than **ned$^{(h)}$-to-* or **ned-samo-* (p. 171 with ft. 135).

There are very few typographical errors. I notice *for* for *form* (p. 98), *already* (p. 99), *postdating* (p. 119 under 8.), *the a* (p. 125 s.v. **argamo-*), *means* for *menas* (p. 130), a dot instead of a comma before *It ...* (p. 136), a missing *on* (p. 143 s.v. ?*morko-*), *and* (p. 146 s.v. **kou̯no-*), the adjective *unknown* on p. 170 is likely to be unsuitable (does the author mean "unclear"?). The author's English is generally very good, but several non-nativisms occur throughout the text. This may occasionally pose a problem in as much as it can at times, although very rarely, obscure the idea behind the formulation to the point that it is rather difficult to be sure what exactly the author is trying to convey to the reader. The Table of Contents is not entirely synchronised with the actual pagination and "*1. Introduction*" in the head of p. 117 is misplaced. The monograph is equipped with a comprehensive and generous (in particular by assigning the individual entries a linguistic affiliation) index (pp. 219–237), that leads the reader to the onomastic and the appellative language material (be it reconstructed or factual). One perhaps misses more elucidation on different sets of exposed phenomena (mostly of phonological nature) that the author has dealt with at length elsewhere. At least the main points of argumentation should be given at the relevant sections.

In summary this is a fine and very capable addition to paleohispanic linguistics, comparative philology of Celtic languages (contributing importantly to every level of linguistic expression), and a refinement of several difficult aspects of the comparative grammar of Indo-European languages in general.

References

Meid, Wolfgang. 2005. *Keltische Personennamen in Pannonien*. Budapest: Archaeolingua.

Olsen, Birgit Anette. 2010. *Derivation and Composition. Two Studies in Indo-European Word Formation*. Innsbruck: Institut für Sprachen und Literaturen der Universität Innsbruck.

Repanšek, Luka. 2016a. Review of Alexander Falileyev: In Search of the Eastern Celts. Journal of Celtic Linguistics 17: 239–280.

Repanšek, Luka. 2016b. *Quiemonis* and the epichoric anthroponymy of Ig. *Arheološki vestnik* 67: 321–257.

Repanšek, Luka. 2016c. Keltska dediščina v toponimiji jugovzhodnega alpskega prostora. Ljubljana: ZRC SAZU, Založba ZRC.

Repanšek, Luka. 2017. New Onomastic Evidence from Ig. *Historische Sprachforschung* 130, *forthcoming*.

Sims-Williams, Patrick. 2003. The Celtic Inscriptions of Britain: Phonology and Chronology, c. 400–1200. London: Blackwell.

Uhlich, Jürgen. 2003. Die Reflexe der keltischen Suffix-Varianten **-i̯o-* und **-ii̯o-* im Altirischen. In: Rockel, M., Zimmer, S. (Hrsg.), *Akten des ersten Symposiums deutschsprachiger Keltologen, Gosen bei Berlin, 8.–10. April 1992*: 353–370. Tübingen: Max Niemeyer Verlag.

И. М. Егоров†, *А. С. Безлепкин*‡
† Российский государственный гуманитарный университет (Москва); i.m.jegorow@gmail.com
‡ Российский государственный гуманитарный университет (Москва); a.s.bezlepkin@gmail.com

XII традиционные чтения памяти С. А. Старостина
Москва, РГГУ, 23–24 марта 2017 года

23–24 марта 2017 года в Москве под эгидой Института восточных культур и античности РГГУ состоялись XII чтения памяти Сергея Анатольевича Старостина. Первый день был посвящен вопросам исторического изучения индоевропейских и других ностратических языков; во второй день рассматривались также проблемы изучения других языков Евразии, Африки и Центральной Америки. По традиции конференцию открыл директор Института восточных культур и античности Илья Смирнов, затем выступил Георгий Старостин с обзором новейших достижений в области сравнительно-исторического языкознания. Тезисы докладов и презентации доступны на сайте проекта «Вавилонская башня»[1].

На утренней сессии первого дня чтений Артем Трофимов (Лаборатория востоковедения ШАГИ РАНХиГС) представил доклад «Особенности составления 110-словных списков Сводеша для ведийского и авестийского языков». Докладчик отметил, что списки для «Ригведы» и «Атхарваведы» могут быть рассмотрены независимо, при этом «Атхарваведа» по сравнению с «Ригведой» имеет 5 лексических замен и 3 морфологические. Для составления авестийского списка необходимо привлечение данных «Младшей Авесты», включая самые поздние источники, такие как «Видевдат» и авестийско-пехлевийский словарь. Некоторые вхождения остаются сомнительными, т. к. встречаются только в композитах. Проблему близких синонимов в ведийских и в авестийских текстах А. А. Трофимов предлагает решать, выбирая наиболее свободную от мифологических коннотаций лексему. Так, для значения ‘person’ выбирается *jána-*, т. к. *mānuṣa-* обозначает человека, противопоставленного богам, а *pŭruṣa-* — человека, противопоставленного животным, и мифологического персонажа. Специфическая проблема, связанная с авестийской соматической лексикой, состоит в существовании синонимов, обозначающих части тел ахуров и дэвов. В силу большей частотности и более прозрачной этимологии выбирается «ахуровская» лексика. В трех случаях противопоставление является трехчленным. Один из них — ‘belly’: *udara-* (немаркированное), *uruθuuan-* (об ахурах), *maršu-* (о дэвах). В этом и в двух аналогичных случаях (‘head’ и ‘ear’), несмотря на меньшую частотность, основной считается немаркированная лексема. Одна из значительных проблем — супплетивизм ведийских глаголов. А. А. Трофимову удалось выявить специфику значения отдельных основ: для глагола со значением ‘eat’ как основные выбраны *ad-* и *aś-*, т. к. для основы *ghas-* в «Ригведе» усматривается значение ‘есть (о животных)’; для ‘say’ основные — *brav-*/*āh-*, а *vak-*/*vac-* — ‘(про)возглашать (в ритуальных контекстах)’.

Усовершенствованию глоттохронологической модели распада двух языков посвящен доклад Михаила Васильева и Михаила Саенко (Институт славяноведения РАН) «Точность глоттохронологического метода при датировании языковой дивергенции: данные романских языков». Авторы сравнили модель М. Сводеша, С. А. Старостина и потоковую модель. По их оценке, процесс дивергенции наиболее точно описывается моделями Старостина и потоковой моделью после их предварительной калибровки. Это привело их к выводу, что большая относительная величина доверительного интервала на небольших временных дистанциях ставит под сомнение практическую ценность глоттохронологии при датировании дивергенции языков, разделившихся менее 1,5 тыс. лет назад. Теоретически глоттохронологические расчеты гораздо лучше применимы на больших временных глубинах, т. к. относительная величина доверительного интервала уменьшается. Несколько сужает дове-

[1] http://starling.rinet.ru/confer/confer2017.htm.

рительный интервал рассмотрение данных не двух, а трех идиомов. По мнению авторов, существенное повышение точности глоттохронологических расчетов возможно в первую очередь за счет расширения используемых списков, а также привлечения дополнительных данных для уточнения моделей на материале различных языковых семей и на разных временных интервалах.

Антон Коган (Институт востоковедения РАН) в докладе «Еще раз о генетическом положении языка документов из Ния (Восточный Туркестан)» пересмотрел выдвинутую Томасом Барроу гипотезу о принадлежности к дардской группе языка документов из оазиса Ния (III в. н. э.) на территории современного Синьцзян-Уйгурского автономного района КНР, написанных письмом кхароштхи. Т. Барроу постулирует особую близость этого языка с языком торвали кохистанской подгруппы восточнодардских языков. По данным глоттохронологии, распад кохистанской общности датируется III в. н. э., следовательно язык документов из Ния должен отражать пракохистанское состояние. Единственная изоглосса, которую приводит Т. Барроу как аргумент в пользу близости с торвали — *sv* > *śv*, *śv* > *śp*, *sm* > *m*. Однако, как отметил докладчик, эти переходы имеют типологические параллели в разных индоиранских языках, а внутри кохистанской группы наблюдаются только в торвали. Это обстоятельство не позволяет выводить их на пракохистанский уровень. А. И. Коган рассмотрел изоглоссы, разделяющие дардские и индоарийские языки: рефлекс **r̥* и наличие «дардской метатезы»; рефлексы звонких придыхательных; соответствия др.-инд. *kṣ*. Все они объединяют язык документов из Ния с индоарийскими, но не с дардскими языками. Две лексические изоглоссы, общие для рассматриваемого идиома и дардских языков (*patama* ‘сзади’, *jaṃdu-* ‘змея (?)’), могут быть интерпретированы как контактные явления. В итоге был сделан вывод, что язык документов из Ния, возможно, был соседом дардских языков, но его принадлежность к этой подгруппе исключена.

В докладе «“Путь — дорога” (№ 68 из списка Сводеша): уточнение семантики и попытка выявления типологических и этимологических универсалий» Татьяна Михайлова (Институт языкознания РАН) рассмотрела параллельные семантические развития слов для обозначения данного концепта в кельтских, германских, славянских и балтийских языках. Она подчеркнула, что не следует искусственно удревнять список. Так, для немецкого языка в значении ‘road’ необходимо использовать раннее заимствование из народной латыни *Straße*, а не старое *Weg*. Существует цепочка типичных семантических переходов: ‘дорога’ > ‘путь (в высоком смысле)’ > ‘способ/образ действия’. Общеславянское **pǫtь* сохранилось в значении ‘road’ в македонском, болгарском и сербохорватском, ср. рус. *путь* и с аналогичным русскому значением чеш. *pout’*. Фигуральное значение сохранили англ. *way*, а также др.-ирл. *sét* < **sentu-*. Т. А. Михайлова выделила два типа лексических замен для концепта ‘road’: «по линии maker» (рус. *дорога*, чеш. *cesta*, связанные с идеей расчистки пути) и «по линии user» (лит. *kelias*, связанное с идеей хождения). Общекельт. **sentu-* в значении ‘road’ вытеснено в древнеирландском словом *slige* — изначально ‘просека’ (ср. славянское развитие), *sét* сохраняет значения ‘тропа’, ‘путь (высокое)’, ‘судьба’. Затем *slige* повторило судьбу *sét*, сохранившись в значении ‘путь (высокое)’, ‘направление’. Основное же значение ‘road’ в ирландском приобрела лексема *bothar* — изначально ‘коровья тропа’. Сопоставление семантического развития в древнеирландском и шотландском подтверждает теорию автора о гойдельском ветвлении. В валлийском *hynt* в значении ‘road’ вытеснено заимствованием *ffordd* со старым значением ‘брод’.

Актуальным вопросам компьютерных методов в лингвистике был посвящен доклад Александры Евдокимовой (Институт языкознания РАН) «Проблема членения текста в византийских надписях и в ряде языков со слитным написанием при компьютерном анализе». В докладе были подробно разобраны трудности, возникающие при членении строки из византийской посвятительной надписи. Речь шла об основной проблеме автоматического парсеринга слитных текстов — о так называемом «комбинаторном взрыве». Проблема состоит в огромном количестве гипотетических членений текста, которые может построить компьютер. Опираясь на опыт работы с арабским языком, А. А. Евдокимова предложила задавать формулу структуры словоформы, что отчасти помогло бы правильно анализировать неизвестные компьютеру слова. Также она отметила, что членение от конца оказывается более четким и дает меньше вариантов. Была рассмотрена возможность представления структуры словоформы по аналогии с агглютинативными языками в виде дерева. По заключению автора, на данный момент избежать «комбинаторного взрыва» и тем самым решить проблему автоматического парсеринга слитных текстов все-таки не удается.

Илья Грунтов (Институт языкознания РАН) и Ольга Мазо (Институт восточных культур и античности РГГУ) в докладе «Оценка гипотезы о заимст-

вовании личных местоимений в монгольских и тунгусских языках из булгарско-тюркского: комментарии к статье Александра Вовина» разобрали аргументы, приводимые автором статьи (Vovin 2011) в пользу данной гипотезы. Для тюркского чередования *b*/*m*, которое А. Вовин считает признаком булгарского заимствования, И. А. Грунтов и О. М. Мазо находят аналогию в монгольском в виде свободного варьирования, привлекая сравнительный материал бурятского языка: монг. *büyile* ‘десны’ — бур. *мүйлэ*/*мылэ*, ст.-монг. *qobur* ‘редкий’ — бур. *хомор*; *намша*/*набша* ‘листья’; *qabar*/*qamar* ‘нос’; монг. *metü* ‘как’ — бур. *мэтэ*, *бэтэ*, *бүтэ*. Опровергая тезис А. Вовина о том, что последовательность *min* встречается только в рассматриваемых местоимениях и в заимствованиях, докладчики привели бур. *минж* ‘бобр’. Они отметили, что гипотеза А. Вовина не объясняет монгольский и тунгусский аблаут. По их мнению, остается непонятным, почему при том, что большие пласты базовой лексики заимствованы в тунгусский через монгольский, это неверно для такого базового слова, как местоимение 2-го лица. Авторы доклада указывают на наличие в тунгусском и монгольском противопоставления инклюзивных и эксклюзивных местоимений 1-го лица множественного числа, отсутствующее в тюркских языках. Гипотеза заимствования предполагала бы перестройку под него всей системы. В докладе рассмотрен и ряд других проблем аргументации А. Вовина.

Тему тюркских заимствований в монгольских языках продолжил Михаил Живлов (Институт восточных культур и античности РГГУ; Лаборатория востоковедения ШАГИ РАНХиГС) докладом «Об одной группе тюрко-монгольских лексических параллелей». Он выделил группу слов, принадлежащих к культурной лексике, с особой системой соответствий гласных: **o — u** ПТю. **Koč* ‘баран’ ~ ПМо. **kuča* ‘id.’, ПТю. **Koŕɨ* ‘ягненок’ ~ ПМо. **kurigan* ‘id.’, ПТю. **tokli* ‘ягненок до полугода’ ~ ПМо. **tugul* ‘теленок’, ПТю. **bodo-* ‘красить’, **bodo-g* ‘краска’ ~ ПМо. **budu-* ‘id.’, **budug* ‘id.’, ПТю. **Koruŕ* ‘вид музыкального (струнного) инструмента’ ~ ПМо. **kuhur* ‘id.’; **ö — ü** ПТю. **(h)öküŕ* ‘крупный рогатый скот’ ~ ПМо. **hüker* ‘id.’; **u — i** ПТю. **buŕagu* ‘теленок’ ~ ПМо. **birahu* ‘теленок (1 года)’, ПТю. **jular* ‘недоуздок’ ~ ПМо. **ǯiluha* ‘недоуздок, вожжи’ (?), ПТю. **jApaku* ‘свалявшаяся шерсть’ ~ ПМо. **dahaki* ‘линька, свалявшаяся шерсть’; **ɨ — a** ПТю. **Kɨptu* ‘ножницы’ ~ ПМо. **kajiči* (= **kahiči*) ‘id.’, ПТю. **Kɨs(ɨ)rak* ‘яловая кобыла’ ~ ПМо. **kasirag* ‘трехлетняя корова, телка’ (?), ПТю. **Atɨŕ* ‘орошаемое поле, межа’ ~ ПМо. **(h)atar* ‘необработанная земля’, ПТю. **jAgɨr* ‘нагнет от седла’ ~ ПМо. **dahari* ‘id.’; **ạ — u** ПТю. **ạguŕ* ‘молозиво’ ~ ПМо. **uhurag* ‘id.’, ПТю. **Kạjɨr* ‘солончак’ ~ ПМо. **kužir* ‘id.’. М. А. Живлов предположил, что эти лексемы были заимствованы в прамонгольский из некоторого несохранившегося паратюркского языка.

Второй день конференции открылся докладом Сергея Кулланды (Институт востоковедения РАН) «К этимологии некоторых северокавказских заимствований в русском языке». Одно из слов, о которых шла речь — *хинкали*, название блюда из теста с мясной начинкой, заимствованное через грузинский из аварского и происходящее из формы аварского множественного числа от *hinḳ* ‘пельмень, хинкали’. Докладчик высказал предположение, что в аварском оно из пралезгинского **χ:imḳ* ‘кулак’ (по форме пирожков). Основой для еще одной предложенной этимологии послужил анализ происхождения др.-инд. *gandharvá-* и авест. *gaṇdərəβa-*. Несоответствие анлаутного согласного индийской и иранской форм, с одной стороны, и сопоставляемого обычно с ними греч. Κένταυροι (мифические полулюди-полукони), с другой стороны, объясняется заимствованием из какого-то субстрата. Для его идентификации, по мнению С. В. Кулланды, следует обратить внимание на то, что гандхарвы — мифологическое воплощение социовозрастных мужских союзов. Подобно молодежи индийских племен Нового и Новейшего времени, обитавшей в мужских домах, гандхарвы занимались пением, музыкой и танцами, совершали набеги на соседей (как в известном мифе о Пуруравасе и Урваши), в том числе для похищения невест («брака по обычаю гандхарвов»). Как и инициируемые юноши, они считались духами, пребывавшими в промежутке между двумя рождениями. Между тем название, сходное с первой частью слова *gandharvá-*, носят канты — персонажи осетинского нартского эпоса; в несколько иной форме это слово встречается у адыгов как название враждебного нартам народа. Все это дает возможность предположить, что источником послужило нахск. **ḳan(a)t* (ср. чеч. и инг. *ḳant*, бацб. *ḳnat*) ‘мальчик, юноша, молодец, удалец’ — такое значение вполне соответствует представлению о молодых людях, которые по своему возрастному положению обязаны проявлять удаль. Вторая часть слова *gandharvá-* и родственных ему композитов, возможно, происходит от пранахск. **ʁāri* ‘шайка, ватага’ (чеч. *кӏера*, инг. *кӏар*, бацб. *ʁajrĭ*), начинавшегося на звонкий увулярный фрикативный, что способствовало озвончению предшествовавшего *t* в индоиранских языках. Как считает С. В. Кулланда, к тому же *ḳant* может восходить русское жаргонное *кент*.

Алексей Касьян (Институт языкознания РАН; Лаборатория востоковедения ШАГИ РАНХиГС) и Георгий Старостин (Институт восточных культур и античности РГГУ; Лаборатория востоковедения ШАГИ РАНХиГС) представили доклад под названием «Автоматическое сравнение енисейско-буришской базисной лексики и вероятностная оценка схождений». В нем были предложены доводы в пользу давнего предположения о родстве между енисейской семьей и языком-изолятом бурушаски. Материалом исследования послужили 110-словные списки реконструированной базисной лексики для праенисейского и прабурушаски, составленные Г. С. Старостиным в рамках проекта «Global Lexicostatistical Database». Они были сопоставлены друг с другом методом консонантных классов с применением перестановочного теста[2]. Компьютерная программа проводила сравнение дважды — с использованием более общего и более дробного набора консонантных классов (каждый из которых обозначается прописной буквой — далее такое обозначение использовано при упрощенной записи этимонов, приводимой в квадратных скобках). В обоих случаях обнаружены одни и те же 6 совпадений: (1) ен. *qɔɢ- [KK] = бур. *qaq- [KK] ‘dry’; (2) ен. *si:- [SH], бур. *ʂi [SH] ‘to eat’; (3) ен. *=o [HH], бур. *=u- [HH] ‘to give’; (4) ен. *xe:y [KH], бур. *=s=ʁa- [KH] ‘to kill’; (5) ен. *ʔiɢ [HK], бур. *ek [HK] ‘name’; (6) ен. *ʔu, *ʔa [HH], бур. *i- [HH] ‘that’. Перестановочный тест показал, что вероятность получить минимум 6 схождений составляет 0.048 и 0.013 в зависимости от используемого набора классов. Эти величины статистически значимы при пороге 0.05. Таким образом, фонетические сходства между праенисейским и прабурушаски 110-словниками не случайны, и результат исследования можно рассматривать как сигнал дальнего родства между двумя семьями. Интересная особенность использованного метода состоит в том, что из 6 полученных с его помощью схождений 4 подтверждаются сравнительно-историческим методом, а 2 — явно ошибочны: ‘to kill’ и ‘that’. При этом 5 этимологий, предложенных С. А. Старостиным с опорой на сравнительно-исторический метод, пропущены алгоритмом, поскольку фонетическое сходство между соответствующими корнями слишком затемнено: (1) ен. *de-s ‘eye’, бур. *=l-ɕi ‘eye’; (2) ен. *ʔaʒ ‘I’, бур. *ʒa ‘I’; (3) ен. *yɔ:pe ‘leaf’, бур. *ʎap ‘leaf’; (4) ен. *ci:ǯ ‘root’, бур. *cʰereʂ ‘root’; (5) ен. *ʔaw ‘thou’, бур. *un ‘thou’. В заключение Г. С. Старостин обратил внимание на то, что вывод о родстве между енисейскими языками и бурушаски сохраняет силу независимо от того, принимается ли при этом гипотеза о существовании сино-кавказской макросемьи, предположительно объединяющей, кроме названных, еще несколько языковых семей.

В докладе Йоханна-Маттиса Листа (Институт Макса Планка, Йена) «Сетевой подход к анализу древнекитайских рифм» был показан пример применения методов, используемых при изучении социальных сетей, в реконструкции древнекитайской фонологии. Проведенный автором анализ графа, вершинами которого являются иероглифы «Книги песен» («Шицзин»), подтверждает гипотезу С. А. Старостина о существовании в древнекитайском языке коды *-r*, позднее слившейся с кодой *-n*. Об этом свидетельствует разделение графа на группы рифмующихся иероглифов, аналогичное выявлению сетевых сообществ.

Анастасия Крылова (Институт восточных культур и античности РГГУ) в докладе «Языки мунда: лексикостатистика по новым данным» рассказала о результатах экспедиции в индийский штат Орисса, в ходе которой были собраны 100-словные списки для четырех языков мунда: сора, хо, бонда и мундари. Полученные сведения позволили уточнить генеалогическую классификацию этих языков методами лексикостатистики. В целом получила подтверждение точка зрения исследователя австроазиатских языков Ильи Пейроса, хотя между списками, собранными экспедицией, и 100-словниками, представленными в его работах, есть расхождения (в случае бонда весьма существенные). Проведенная работа показывает, что распад протомунда на северную и южную группы датируется первой половиной II тысячелетия до нашей эры, вскоре после этого разделились бонда и сора (южная группа), а в середине I тысячелетия нашей эры — мундари и хо (северная группа).

Евгения Коровина (Институт языкознания РАН; Институт восточных культур и античности РГГУ) выступила с докладом «Классификация языков майя: методы и результаты». В первой его части она показала, как сложились существующие представления о классификации языков майя. Во второй — продемонстрировала результаты построения генеалогии майя по 100- и 200-словным спискам Сводеша при помощи различных иерархических и филогенетических методов, а также классификацию, полученную средствами этимостатистики. Кроме того, были рассмотрены причины некоторых расхождений между различными вариантами генеалогического дерева этих языков.

[2] Об этом способе сопоставления можно прочитать в публикации Kassian et al. 2015.

Виктор Порхомовский (Институт языкознания РАН) в докладе «Диахроническая типология в семитском историческом языкознании» изложил идеи, которые легли в основу разработанной им четырехступенчатой модели, описывающей эволюцию видо-временной системы семитских языков. Эта модель противопоставлена концепции И. М. Дьяконова, согласно которой семито-хамитские языки делятся на языки древней, средней и новой ступени. Стадиально-типологическая классификация языка в целом, по мнению автора доклада, затруднена тем, что различные уровни языковой системы эволюционируют неравномерно — например, среди семитских языков есть такие, которые архаичны с точки зрения фонологии, но морфологически сильно изменились по сравнению с прасемитским состоянием. Между тем применительно к глагольной системе такое распределение языков по разным этапам развития, как считает В. Я. Порхомовский, возможно. Все семитские языки он относит к какой-нибудь из четырех ступеней, предполагаемых его моделью, в зависимости от того, какая ситуация в них сложилась в рамках аспектной оппозиции перфектива и имперфектива: на каждом этапе один из членов оппозиции «сильный», а другой «слабый» (употребляемый там, где нет необходимости в употреблении сильного); это приводит к появлению наряду со слабой формой глагольного вида его сильной формы, со временем вытесняющей слабую; привативная оппозиция превращается в эквиполентную; на новой ступени порядок изменений повторяется.

Ольга Попова (Институт языкознания РАН) посвятила доклад «Орфография царского имени Артаксеркс в нововавилонских клинописных источниках как критерий датировки текстов» проблеме различения имен трех правивших в Вавилонии ахеменидских царей: Артаксеркса I (465–424 гг. до н. э.), Артаксеркса II (405–359 гг. до н. э.) и Артаксеркса III (359–338 гг. до н. э.). Рассматривая написание второй части имени, О. В. Попова установила три варианта передачи др.-перс. *ç* (вероятно, обозначавшего /ts/): šá-**as**-**su**; šá-**su**; **šat**-**su**. На основании корпусного анализа текстов, чья датировка не вызывает сомнения, докладчице удалось установить, что для записи имени Артаксеркса III употреблялось только написание со знаком šat. Это связано с общей тенденцией распространения знаков структуры CVC и не имеет идеологических причин.

Ольга Столбова (Институт востоковедения РАН) в докладе «50-словный список базовой лексики чадских языков и внутренняя классификация афразийской семьи» представила результаты поиска лексических изоглосс, выявление которых могло бы помочь уточнению положения чадских языков на афразийском генеалогическом дереве. В дискуссии о том, на какие части первоначально разделился праафразийский язык, чадская ветвь занимает центральное место, поскольку есть по крайней мере три гипотезы о том, с какой другой ветвью она связана наиболее близким родством: 1) с берберской, 2) с египетской, 3) с кушитской (или, возможно, кушитско-омотской). Для проверки этих предположений О. В. Столбова рассмотрела те случаи, когда у чадских корней из 50-словного списка нет общесемитских параллелей. Проведенный анализ не позволил сделать вывода о предпочтительности какой-либо из гипотез, но способствовал постановке вопросов о том, какую роль в решении проблемы могли бы сыграть реконструкции разного уровня и разной репрезентативности (например, опирающиеся только на материал отдельных ответвлений внутри чадской семьи).

Проблема определения степени близости языковых семей Африки была затронута и в заключительном докладе конференции, с которым выступил Георгий Старостин (Институт восточных культур и античности РГУ; Лаборатория востоковедения ШАГИ РАНХиГС), — «К вопросу об ареально-генетических “клубках” в базисной лексике на примере этимона ʾлунаʾ в макросуданском регионе». Докладчик проанализировал соотношения между несколькими этимонами с общим значением ‘луна’, реконструируемыми для языков гипотетической «нило-сахарской» макросемьи (нилотских и центральносуданских), а также для некоторых других языков той же части континента, в частности афразийских (кушитских и омотских) и нигер-конголезских (убангийских). Было показано, что, несмотря на очевидное фонетическое сходство между рассмотренными нилотскими и центральносуданскими формами, они не могут служить доводами в пользу нило-сахарского родства. По мнению Г. С. Старостина, наиболее логичное и экономное объяснение их распределения по таксонам — ареальная диффузия, затронувшая «нило-сахарские» и афразийские языки региона. Первоначальный ее источник установить трудно, но вряд ли это прямой предок какой-либо из упомянутых групп. Приблизительная исходная форма может быть восстановлена как **ɬape* или **ɬapay*. Она эволюционировала двумя основными путями: **ɬapay* → **ɬafay* → **ɬahay* (отсюда произошло большинство кушитских и южнонилотских форм) и **ɬapay* → **apay*

→ **pay* (отсюда — большая часть нилотских, центральносуданских и убангийских форм). Аналогичные сценарии, как считает автор доклада, не исключены и для других базисных лексем, что важно учитывать в будущих исследованиях генетических связей между африканскими языковыми семьями.

Литература

Kassian, Alexei, Mikhail Zhivlov, George Starostin. 2015. Proto-Indo-European-Uralic Comparison from the Probabilistic Point of View. *Journal of Indo-European Studies* 43: 301–347.

Vovin, Alexander. 2011. First and second person singular pronouns: a pillar or a pillory of the 'Altaic' hypothesis? *Türk Dilleri Araştırmaları* 21(2): 251–278.