

Российский государственный гуманитарный университет  
Russian State University for the Humanities



Russian State University for the Humanities  
Institute of Linguistics of the Russian Academy of Sciences

# Journal of Language Relationship

International Scientific Periodical

N°19/4



Moscow 2021

Российский государственный гуманитарный университет  
Институт языкознания Российской Академии наук

# Вопросы языкового родства

Международный научный журнал

№ 19/4



Москва 2021

*Advisory Board:*

H. EICHNER (Vienna) / Chairman  
W. BAXTER (Ann Arbor, Michigan)  
V. BLAŽEK (Brno)  
L. HYMAN (Berkeley)  
F. KORTLANDT (Leiden)  
A. LUBOTSKY (Leiden)  
J. P. MALLORY (Belfast)  
A. YU. MILITAREV (Moscow)  
V. F. VYDRIN (Paris)

Editorial Staff:

V. A. DYBO (Editor-in-Chief)  
G. S. STAROSTIN (Managing Editor)  
T. A. MIKHAILOVA (Editorial Secretary)  
A. V. DYBO  
M. A. MOLINA  
M. N. SAENKO  
I. S. YAKUBOVICH

Founded by Kirill BABAEV

*Редакционный совет:*

Х. АЙХНЕР (Вена) / председатель

В. БЛАЖЕК (Брно)

У. БЭКСТЕР (Анн Арбор)

В. Ф. ВЫДРИН (Париж)

Ф. КОРГЛАНДТ (Лейден)

А. ЛУБОЦКИЙ (Лейден)

Дж. МЭЛЛОРИ (Белфаст)

А. Ю. МИЛИТАРЕВ (Москва)

Л. ХАЙМАН (Беркли)

*Редакционная коллегия:*

В. А. ДЫБО (главный редактор)

Г. С. СТАРОСТИН (заместитель главного редактора)

Т. А. МИХАЙЛОВА (ответственный секретарь)

А. В. ДЫБО

М. А. МОЛИНА

М. Н. САЕНКО

И. С. ЯКУБОВИЧ

Журнал основан К. В. БАБАЕВЫМ

Вопросы языкового родства: Международный научный журнал / Рос. гос. гуманитар. ун-т; Рос. акад. наук. Ин-т языкознания; под ред. В. А. Дыбо. — М., 2021. — № 19/4. — x + 80 с.

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, 2021. — No. 19/4. — x + 80 p.

ISSN 2219-3820

<http://www.jolr.ru/>  
[gstarst1@gmail.com](mailto:gstarst1@gmail.com)

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

Подписано в печать 28.12.2021. Формат 60×90/8.  
Усл. печ. л. 11,3. Уч.-изд. л. 7,5. Заказ № 1446.  
Тираж 1050 экз.

Издательский центр  
Российского государственного гуманитарного университета  
125047, Москва, Миусская пл., 6  
[www.rggi.ru](http://www.rggi.ru)

# Table of Contents / Содержание

Table of Contents / Содержание . . . . .	vii
Contributors / Сведения об авторах . . . . .	viii
Note for Contributors / Будущим авторам . . . . .	ix

## Articles / Статьи

<i>Alexander Militarev, Sergei Nikolaev</i> . Proto-Afrasian names of non-ungulate animals in light of the Proto-Afrasian homeland issue . . . . .	233
[А. Ю. Милитарев, С. Л. Николаев. Праафразийские названия некопытных животных в свете проблемы афразийской прародины]	
<i>Georg Orlandi</i> . Once again on the history and validity of the Sino-Tibetan bifurcate model . . . . .	263
[Георг Орланди. Еще раз к вопросу об истории и степени обоснованности бинарной модели классификации сино-тибетских языков]	
<i>С. А. Бурлак</i> . Устойчивость и частотность: есть ли корреляция? . . . . .	293
[Svetlana Burlak. Stability and frequency: is there a correlation?]	

## Reviews / Рецензии

<i>Juan Luis García Alonso</i> . ALEJANDRO G. SINNER, JAVIER VELAZA (eds.). Palaeohispanic Languages and Epigraphies. Oxford: Oxford University Press, 2019. — 483 + xx p. . . . .	308
--	-----

## Contributors

*Juan Luis García Alonso* – professor, Department of Classic and Indo-European Philology, University of Salamanca (Spain), jlga@usal.es

*Svetlana Burlak* – doctor of sciences (Philology), senior researcher, Institute of Oriental Studies of the Russian Academy of Sciences (Moscow), svetlana.burlak@bk.ru

*Alexander Militarev* – doctor of sciences (Philology), professor, Institute for Oriental and Classical Studies, RSUH (Moscow), amilitarev@gmail.com

*Sergei Nikolaev* – doctor of sciences (Philology), lead researcher, Department of Slavic linguistics, Institute of Slavic studies, Russian Academy of Sciences (Moscow/Novosibirsk), sergenicko@mail.ru

*Georg Orlandi* – postdoctoral researcher, Osaka University; resident researcher, Italian School of East Asian Studies, Kyoto, thegaygenius@gmail.com

## Сведения об авторах

*Алонсо, Хуан Луис Гарсиа* – профессор отделения классической и индоевропейской филологии Саламанкского университета (Испания), jlga@usal.es

*Бурлак, Светлана Анатольевна* – доктор филол. наук, профессор РАН, ведущий науч. сотрудник Института востоковедения РАН (Москва), svetlana.burlak@bk.ru

*Милитарев, Александр Юрьевич* – доктор филол. наук, проф. ИВКА РГГУ (Москва), amilitarev@gmail.com

*Николаев, Сергей Львович* – доктор филол. наук, вед. науч. сотрудник Отдела славянского языкознания Института славяноведения РАН (Москва), sergenicko@mail.ru

*Орланди, Георг* – постдокторант Осакского университета (Осака, Япония); научный сотрудник Итальянской школы восточноазиатских исследований (Киото, Япония), thegaygenius@gmail.com



## 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. Articles are published preferably in English or Russian, although publication of texts in other major European languages (such as French or German) may be considered by the Editorial Board.

All submissions should be uploaded electronically in MS Word and PDF format, using the online *Manuscript Submission Form* at the official website of the Journal (<http://jolr.ru>). Each article should be accompanied with information about the author(s) (names, affiliations, contact information), an abstract (not exceeding 300 words) and relevant keywords.

For more detailed guidelines on article submission and editorial policies, please see our website (<http://jolr.ru>) or address the Editorial Board directly at [gstarst1@gmail.com](mailto:gstarst1@gmail.com). Inquiries may also be sent by regular mail to the official address of the Journal:

Journal of Language Relationship  
Institute for Oriental and Classical Studies  
Russian State University for the Humanities  
125047 Moscow, Russia  
Miusskaya Square, 6

## 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. Предпочтительные языки публикации — английский или русский, хотя по согласованию с редакционной коллегией возможна также публикация статей на других крупных европейских языках (французский, немецкий и т. п.).

Подача материалов для публикации (в форматах MS Word и PDF) осуществляется через *Электронную форму подачи рукописей* на официальном сайте журнала (<http://jolr.ru>). К каждой статье обязательно прилагаются сведения об авторах (имена, аффилиации, контактная информация), краткое резюме (не более 300 слов) и список подходящих 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 непосредственно от редакционной коллегии по электронной почте ([gstarst1@gmail.com](mailto:gstarst1@gmail.com)). По различным вопросам с редакцией журнала можно также связаться по обычной почте:

125047 Москва  
Миусская площадь, д. 6  
Российский государственный гуманитарный университет  
Институт восточных культур и античности  
В редакцию журнала «Вопросы языкового родства»



## Proto-Afrasian names of non-ungulate animals in light of the Proto-Afrasian homeland issue

In this paper, we present the second section<sup>3</sup> of a relatively comprehensive thesaurus of Proto-Afrasian zoonyms, compiled and reconstructed by the authors. The list contains more than sixty terms, including monkeys, canines and hyenas, felines, rodents and other mammals, birds, reptiles, amphibia and fishes complementing and completing a wholesome (at the current level of our knowledge) picture of all relevant species of fauna in the period preceding the split of Proto-Afrasian (ca. 11th millennium BCE, according to glottochronology) within the area presumably inhabited by speakers of Proto-Afrasian. Results of the reconstruction may be ambiguously interpreted in favor of either one of the two different points of view on the localization of the Proto-Afrasian homeland, namely, the Levant vs. East Africa.

*Keywords:* Proto-Afrasian homeland; Afrasian (Afro-Asiatic) languages; etymology of zoonyms; reconstruction of cultural lexicon.

### Introduction

In this paper, we present comparative evidence for a large number of Proto-Afro-Asiatic (PAA) “non-ungulate” zoonyms, including monkeys (4 terms), predators (22 terms), other mammals (a selection of 8 terms), birds (selection of 11 terms), reptiles (selection of 9 terms), water reptiles and amphibia (selection of 4 terms), and fishes (selection of 5 terms). Within the “predator” subset, we differentiate between 8 reconstructed terms for canines, including ‘dog’ (\**kawit-*, \**k<sup>w</sup>ihan-*, \**KV3/žim-*), ‘k. of wild canine’ (\**ʔaway-*, \**ʔaw/ys-*, \**baʔiḥ-*, \**kur(-ay-)*, \**wanš-*), 4 undifferentiated terms ‘k. of canine or hyena’ (\**ʔa/usk-ay-*, \**baʔVš*, \**gVd-*, \**wahr-*), 9 terms for felines including ‘lion’ (\**labiʔ-*), ‘leopard or lion’ (\**baʔy-*), ‘k. of (wild) feline’ (\**ʔariw-*, \**baʔis-*, \**daʔ/y)m-*, \**giwar-*, \**layč-*, \**mary-*, \**sawr-*), and one undifferentiated term ‘k. of feline, hyena or viverra’ (\**ʔažur-*).

In the opinion of one of the paper’s authors (S. Nikolaev), this particular array of predators in the context of terms for monkeys, ungulates and large herbivores should stimulate a search for the PAA homeland in the Eastern part of the Sahel. Although some of these zoonyms are universal and thus “neutral” in regards to the debatable issue of the Afrasian *Urheimat*, others at least make it possible to add an important observation (which, however, could hardly shed any more light on the problem, since it can be interpreted both ways): while

---

<sup>1</sup> Lexical material, reconstruction of zoonyms and presentation of the Levant *Urheimat* hypothesis for Proto-Afrasians, including their identification with Natufians. Militarev’s research is supported by The Russian Science Foundation (Project № 20-18-00159); the financing organization is The Institute of Linguistics, Russian Academy of Sciences.

<sup>2</sup> Evaluation of PAA zoonyms, their paleozoological interpretation and argumentation in favor of the African (most likely East Sudanic) *Urheimat* of Proto-Afrasian, as well as commentary on the semantic reconstruction of zoonyms.

<sup>3</sup> The first section, “Proto-Afrasian names of ungulates in light of the Proto-Afrasian homeland issue”, was published in 2020 (*Journal of Language Relationship* 18/3–4: 199–226). The list in the second section continues the numeration begun in the first one (1. Bovids and pigs; 2. Camel and equids; 3. Largest herbivores).

there are terms in various Semitic (not only Ethiopian Semitic) languages that denote African animals such as the ones represented in the current paper (namely, monkeys and crocodiles), not a single one of those has a common Afrasian or even common Semitic etymology. This evidently implies a late origin of the corresponding Semitic zoonyms and, hence, late acquaintance of the already linguistically separated Semitic-speaking groups with the corresponding African animals.

A special commentary is needed for the issue of semantic reconstruction of zoonyms. It is quite rare that the Proto-Afrasian meaning may be reconstructed unequivocally on the basis of perfectly matching semantics, as in the following cases:

5.9. *\*k<sup>w</sup>ihan-* ‘dog [*Canis familiaris*]’.

Chad. *\*k<sup>w</sup>iHan-* ‘dog’: W. *\*k<sup>w</sup>iHan-*: Fyer *k<sup>w</sup>éerj* || E. *\*kany-* ~ *\*kayan-*: W.Dangla *kàny-à* | E.Dangla *kāny-à* | Migama *kānny-à* | Mabire *kany* | Jegu *kány* | Birgit *káyànj*.

Cush. E. *\*k<sup>w</sup>ihan-*: Yaaku *kwehen* ‘dog’.

Omot. *\*k<sup>w</sup>iHan-* ‘dog’: N. *\*ku/iHan-*: Wolaita, Gamu, Dawro, Zala, Malo, Dache, Koyra, Zayse, Gimira (Bench) *kyan* | Yemsa, Bworo *kan-a* | Kafa, Mocha *kun-ano* | Dizi (Sheko) *kean-u* | Mao (Hozo) *kan-a*, etc. || S. *\*kan-*: Dime *kén-ε*, Galila *kan-i*.

6.7. *\*labi?-* ‘lion [*Panthera leo*]’

Sem. *\*labi?-*: Akk. *labb-u* (*lab?-u*, *lāb-u*) || Ugr. *lb?-u* ‘lion’ [*Panthera leo*] || Hbr. *lābī(?)* ‘lioness’ || Sab. *lb?* ‘lion, lioness’ || Arab. *luba?-at-*, *labu?-at-*, *libw-at-*, etc. ‘lionne’.

Egyp. (OK) *rw ʒbw*, inscription under the image of a lion | Copt.: Bohairic, Sahidic *laboy* ‘lion; bear’ [*P. leo*; *Ursus arctos*].

Chad. *\*la/ib-*: W. *\*lab-*: Tangale *lab-ata* (f.) ‘lioness’ [*P. leo*] || C. *\*(ʔa-)lib-ar-* ‘lion’ [*P. leo*]: Hwona *lifār-i* | Margi *ḥa-livār-i* | Chibak *ʔa-lvār-i* | Kilba *lēvār-i* | Kapsiki *nivèr-i* (dissim.) | Higi-Nkafa *livèr-i* | Gude *livyār-á*.

Cush. E. *\*lu/ib-* ‘lion’ [*P. leo*]: SA *\*lub-ak-*: Saho, Afar *lubaak* | LEC *\*lib-aḥ-*: Somali *libaḥ* | Jiddu *libááh-u*.

In the vast majority of cases, semantic reconstruction is difficult for four main reasons:

1) intermediate proto-languages, except for the relatively young Proto-Berber, are characterized by a considerable time depth (Proto-Chadic, Proto-Cushitic, Proto-Omot, somewhat younger Proto-Semitic), which makes diverse semantic shifts inevitable;

2) meanings of zoonyms in dictionaries of most Afrasian languages are often quoted inaccurately — in particular, specific species of animals are given as approximate or generic terms: “antelope”, “deer”, “duiker”, “wild cat”, etc., although in reality these zoonyms probably denote more concrete species;

3) according to modern scientific data (though somewhat debatable), in the Proto-Afrasian period there were no domestic animals (except for the dog [*Canis familiaris*]); names of wild animals were transferred to domestic ones (and *vice versa*) in the history of individual Afrasian families and languages; if so, “domestic” zoonyms do not contribute to an accurate reconstruction of the Proto-Afrasian semantics of the animal names;

4) zoonyms in various Afrasian languages can denote entire zoological groups (“small ungulates”, “largest animals”, etc.) or sex and age characteristics of certain animals (‘bull’, ‘ram’, ‘kid’, ‘lamb’, etc.); to reconstruct terms with such semantics for Proto-Afrasian is particularly difficult because of their lack of semantic stability.

Below we list several examples in which specifically “wild” semantic reflexes are indicated in semi-bold type. The remaining reflexes either refer to domestic animals and age / sex

groups, or to unspecified wild species. With such semantically chaotic material, semantic reconstruction for Proto-Afrasian is possible only within the framework of generalizing to large zoological groups – “lesser bovids”, “larger bovids”, “equids”, “canines”, “felines”, etc., even though in Proto-Afrasian all or most of these terms may have had more specific meanings.

### 1.2. \*ʔayl- ‘k. of lesser bovid’

Sem. \*ʔayil-: (?) Akk. *al-u* (*el-u*) ‘a fine breed of sheep’ (reading arguable) || Ugr. *ʔal*, *ʔil* ‘ram of superior quality’ | Hbr. *ʔayil* ‘ram’ | ESA: Sab. *ʔyl* ‘**mountain goat, ibex**’ [C. *ibex*], Min. *ʔyl* ‘**bélier, bouquetin**’ [C. *ibex*] | Arab. *ʔiyyal*-, *ʔuyyal*- ‘**bouquetin, bouc de montagnes**’ [C. *ibex*].

Berb. \**ti-halay* (< \**-ʔayal*, met.?) ‘sheep’: Nefusa *t-ili* | Zenaga *t-iʒi* | Ahaggar *té-helé*, Ghat *či-hali*, Ayr *t-ele*, E. Wlm. *te-həle*, etc.

Cush. \*ʔayl-: N. \*ʔall-i: Beja *all-i*, pl. *ill-i* ‘long-haired sheep’ || E.: SA \*ʔill-: Saho *ill-e*, Afar *ill-i* ‘small cattle’ | LEC \*ʔel- (< \*ʔil- or \*ʔayl-): Somali *eel-o* ‘**tipo de gazella (antilope giraffa)**’ [*Litocranius walleri*], Arbore *ʔell-ém*, Elmolo *ʔél-em* ‘ram’ || S. \*ʔayl-: Gorowa *elete-mo* ‘**bushbuck**’ [*Tragelaphus sylvaticus*] | Ma’a *iʔal-é* ‘ram’, *iʔal-ú* (met. < \*ʔayl-?) ‘sheep’ | Dahalo *ʔèl-e* ‘**hartebeest**’ [*Alcelaphus buselaphus*].

### 1.13a. \*guday- ‘k. of larger bovid’

Sem. \**gaday*- ‘kid’: Ugr. *gdy* ‘kid’ | Phoen. *gdʔ* ‘goat’, Hbr. *gādī* ‘kid (of goat or sheep)’ | Aram.: Anc. *gdh* ‘goat’, Syr. *gady-ā*, Mand. *gadi-a* | Arab. *ʒady*- ‘kid’, *ʒadāy-at*- ‘**gazelle; petit de gazelle**’ [*Gazella gazella*].

(?) Berb. N. \**-gVnd-uz* ‘bull, calf’: Rif *a-yenduz* ‘taureau’, Snus *a-yenduz*, Shenwa, Qabyle *a-genduz*, Sened *a-gendus*, etc. ‘calf’.

Chad. \**ga/uday*-: W. \**ga/ud*- ‘k. of bovid’: Hausa *gàd-áa* ‘antelope, duiker sp.’, Kariya *gud-am*, Miya *gud-ən-zāku*, Pa’a *gud-an-cəka* ‘**Western kob**’ [*Kobus kob*], Ngizim *gád-ùwà* ‘duiker’ || C. \**gVday*-: Zime-Batna *gódày* ‘buck’.

Cush. \**gad*- and \**gund*- ‘k. of larger bovid’: E. \**gad-am*-: LEC \**gad-am*-: Oromo *gad-am-sa* ‘**greater kudu**’ [*Tragelaphus strepsiceros*], Dirayta *gad-an-sa*, *gad-am-sa* ‘antelope’ | HEC \**gud*-: Sidamo *god-a* ‘deer, gazelle’ || S. \**gwand*-: Iraqw *gwand-a*, Alagwa *gwand-o* ‘ram’, Burunge *gond-i* ‘old ram’.

Omot. N. \**gayd*- (met.): Zaisse *gaaid-é* ‘cattle’, *gaid-é-endo* ‘**buffalo**’ [*Syncerus caffer*].

### 1.14a. \*garw- ‘k. of bovid’

Sem. \*(ʔa-)gurr-: Akk. (MA) *gurr-atu*, *agurr-atu* ‘ewe’.

Egyp. (MK) *ḏr* ‘calf’ (if < \**gVr*).

Berb. \**-gur*- ‘small cattle’: Ghadames *a-ʒur* ‘bouc’ (Nait-Zerrad 2002: 859), Zenaga *ǝ-grərħ* ‘bélier’ (Nicolas)

Chad. (a) \*(ʔa-)garw/y- ‘kind of bovid (*Redunca*, *Oryx*, *Ourebia*, etc.); animal, game’: W. \*(ʔa-)gary-: Hausa *āgārē* ‘**a big, male red-fronted gazelle**’ [*Eudorcas rufifrons*], Tsagu *gāre* ‘**reedbuck**’ [*Redunca redunca*], Mburku *gāri* ‘**oryx**’ [*Oryx leucoryx* ?], Ngizim *a-gare* ‘gazelle’ || C. (ʔa-n-)gary-: Logone *gari-a*, Makeri *ingərii* ‘antelope’, Buduma *ɲgəri* ‘gazelle’ || E. \**gVrw*- and \**gVwVr*- (met.): Tumak *gəru*, Kwang *gowor-to* ‘antelope’, Ndam *gèrù* ‘**ourebi**’ [*Ourebia ourebi*]; (b) \**garaw*- ‘herd of cattle’: W. \**gar(V)w*-: Hausa *gār-kei* ‘a herd, flock’, Bolewa *gaarùw-à* ‘pack ox’, Karekare *gaarùw-à* ‘bull’, Tangale *káarw-a* ‘cattle’ || C. \**garaw/y*-: Bana *gàrəw-à* ‘troupeau de bœufs’, Gude *gára-nə* ‘herd of cattle’, *mà-gára* ‘shepherd, wachtsman’, Musgu *gari*, *gári* ‘Stier, Laststier’, Mandara *gāri* ‘bull’, Hursa *goragor-a* (redupl.), Uẓam *gwar-a*, Matakam *ɲ-gwur* ‘ram’ || E. \**garaw*-: Bidiya *gaaruw-o* ‘animal’, Migama *gáaráw* ‘bête, bétail’; (c) C. \**g<sup>w</sup>ar*- and \**g<sup>w</sup>arg<sup>w</sup>ar*- (redupl.) ‘ram’: Ouldem *g<sup>w</sup>àr-à*, Mbuko *g<sup>w</sup>àrg<sup>w</sup>ár-á*, Merrey *g<sup>w</sup>aràg<sup>w</sup>àr-a*, Muyang *g<sup>w</sup>òròg<sup>w</sup>òr-à*,

Mada *gurg<sup>w</sup>ar-a* || E. \**gagar-* (redupl.) ‘small cattle’: Sok *gaáger-o* ‘Schaf’, Mawa *gagar* ‘mouton’, Mubi *wegr-i* (met.) ‘chèvre’.

Cush. \**garaw-* and \**gwira?*- (met.) ‘k. of larger bovid [*Taurotragus*, *Alcelaphus*, etc.]’: N. \**garuw-*: Beja *garuw-a* ‘**male antelope; eland**’ [*Taurotragus oryx*] || C. \**gar-* ‘calf’: Bilin, Qwara, Qemant *gär*, Aungi *gara* || E.: HEC \**gur(r)-*: Sidamo *gur-um* *ĩččo* ‘gazelle’, *garr-ančo* ‘**Agazen antelope**’ [*T. buxtoni*] || S. \**g<sup>w</sup>ara?*:- Iraqw *gwara?-ai* ‘**Hartebeest**’ [*Alcelaphus caama*], Burunge *gera?-i* ‘**Grant’s gazelle**’ [*Nanger granti*].

Omot. N. \**gaHar-* ‘**antelope dekula**’ [*Tragelaphus decula*]: Wolayta, Dawro *gaar-aa*.

### 1.18. \**lawi?*- ‘k. of large (?) bovid’

Sem. \**lawi?*- and \**lawli?*- (redupl.) ‘k of. large bovid [*Bos*, *Syncerus*, *Connochaetes*, etc.]’: Akk. (OB on) *litt-u* (*lit-u*) ‘cow’, *lalû* (*lali?-u*, *lala?-u*) ‘kid’, *lulû-mu* ‘**red deer, stag**’ [*Cervus elaphus*] || Ebl. *lí-a-núm*, *lí-a-nu-um* ‘cow’ | Ugr. *ll?-u* ‘lamb, kid’ | Hbr. *lê?-ā* ‘wild cow’ (only as a pers. name) | Arab. *lā?a* ‘**taureau sauvage, buffle**’ [*Syncerus caffer*], *lu?lu?*- ‘antelope’ | Tgr. *lul-it*, pl. *läwäll-it* ‘(cow) with long, crooked horns’, Amh. *lil(lə)wil* ‘**gnu**’ [*Connochaetes taurinus*] || Mhr. *lahay-tən* ‘cows’, Hbt. *leé*, pl. *lháyta*, Jib. *le?*, pl. *lhó-ti*, Soq. *?e-lh-eh* ‘cow’, *lúl-oh* ‘brébis’.

Egyp. (Pyr.) *iw?* ‘bull’ (if < \**IVwV?*-).

Berb. \**w/yalaH-* (met.): Izayan *ta-wala* ‘troupeau de bœufs, sangliers’, Ahaggar *élah-ei* ‘mouton à laine’.

Chad. \**laway-* ‘k. of larger bovid (*Tragelaphus*, etc.)’: W.: Dera *la-à* ‘cow’ || C.: Gude *la* ‘cow’ (Jungraithmayr, Ibriszimow), Kapsiki, Higi-Nkafa *ε*, Higi-Baza *lol-ε* (redupl.) ‘gazelle’, Higi-Ghye *le*, Bachama *liyey* ‘duiker’ (St. 2005 #57), Masa *lúway* ‘troupeau, le betail’ (CED #549).

Cush. \**lawi?*- ‘cattle’: C. \**luway-*: Bilin *luwī*, Khamir *luwā*, Awngi *luwā* || E. \**la?*- and \**la?la?*- (redupl.): SA \**la(?)*:- Saho, Afar *lā* ‘cow, cattle’ | LEC \**lo?(lo?)*- ‘cows (coll.)’: Somali *lo?* ‘cows (coll.)’, Rendille (pl.) *loóly-o*, Oromo *loo-ni* ‘cows (coll.)’, *lal-eesa* ‘she-goat’, Konso *low-aa* ‘cows’, Dasenech *lal-o*, pl. *lal-i* ‘cow’, etc. | HEC \**lal-*: Sidamo *lal-o* ‘cows, cattle’, *laat-to* (f.) ‘young sheep, lamb’ | Dullay \**lo?*-, pl. \**le?*- ‘cow’: Harso, Dipina, Gollango *ló?-o*, Tsamay *lō?-ō-*, pl. *lē?-ē*, etc. || S. \**li?*:- Iraqw *le?-i* | Gorowa *lee?-i* ‘goat’ | Qwadza *le?-amuko* ‘bull’.

### 5.1. \**?away-* or \**yawa?*- (met.) ‘k. of canine’

Sem. \**?away-*: Hbr. \**ʔī* (pl. *?iyyīm*) ‘**jackal**’ [*C. aureus*] | Syr. *bənāt ?away* ‘**thoes, canes aurei**’ [*Canis aureus*] || Arab. *?ibnu-l-?āwan* ‘animal regardé comme un mélange né d’un chien et d’un renard’ || Tgr. *?aw* ‘eatable wild animals’ | Amh. *yäyi* ‘hyena’ [*Hyaenidae sp.*], *awu* ‘hyena’s cry; hyena’ [*Hyaenidae sp.*] | Gur.: Chaha, Gyeto, Ennemor, Muher *awi* ‘wild animal, beast’.

Egyp. (MK) *iw* ‘dog’ (met. < \**Vyw?*?).

Chad. W. \*(?)*iy-*: Warji *iyà-nà* | Kariya *íi* | Miya *i* ‘dog’.

Cush. E. \**yawa?*:- LEC \**yaw/ya?*:- Somali *éy* ‘dog’, *yéy* ‘**wild dog**’ [*Lycaon pictus*] | Boni *óy?* ‘dog’, *yey?* ‘**jackal**’ [*C. aureus*] | Rendille *yááy* ‘**wild dog**’ [*L. pictus*] | Oromo *yeyy-ii* ‘**wolf; wild dog**’ [*L. pictus*] | Konso *yoy-ta* ‘**hunting dog**’ [*L. pictus*] | HEC \**yayy-* ‘**hunting dog**’ [*L. pictus*]: Sidamo *iyäy* | Hadiya *yoyy-akko* | Burji *yeyy-ée*.

### 6.9. \**mary-* or \**mayr-* ‘k. of feline’

Sem. \**namir-* (< \**na-mir-* or \**na-mayr-* with the fossilized prefix) ‘**leopard**’ [*Panthera pardus*]: Akk. *nimr-* (*nammar-*) ‘**panther**’ || Hbr. *nāmēr* ‘**leopard, panther**’ || Aram.: Anc., Off. *nmr* | Bibl. *nəmar* ‘**panther**’ | Mand. *namar* (*nimar*, *namr*) ‘**leopard**’ | Sab. *nmr* ‘**leopard, panther**’ | Hdr. *nmr* ‘**panthère**’ || Arab. *nimr-* ‘**panthère**’, *namir-* ‘**léopard, panthère**’ || Gz. *namr* | Tna., Tgr. *nābri* ‘**leopard**’ | Amh. *nām̄r*, *nābər* ‘**lynx**’ [*Caracal caracal*] | Sel. *nāwər* | Wol. *nāwr*.

Egyp. (OK) *m?y* ‘**lion**’ [*Panthera leo*] (if < \**mVry*).

Chad. *\*mary-* and *\*mumur-/murum-* (redupl.): W. *\*mumur-*: Dera *mūmur-u* ‘cheetah’ [*Acinonyx jubatus*] | Polchi *mur* ‘jackal’ [*Canis aureus*] | Bokkos *mūrûm* | Sha, Daffo-Butura *mūrûm* ‘hyena’ [*Crocuta crocuta*] || C. *\*mary-*: Mafa *mariy-žélé* ‘lion’ [*Panthera leo*] | Lame *mēr* ‘serval’ [*Carracal serval*], *méri-án* ‘wild cat’ [*Felis sp.*] | Peve *méri-àn* ‘Id.’ | Zime-Dari *mīēr* ‘genette’ [*Genetta sp.*].

Cush. (a) *\*mary-*, *\*mawir-* and *\*marir-* (redupl.): E.: LEC *\*mawr-*: Oromo *mōr-ē* ‘civet(-cat)’ [*Civettictis civetta*] | Dasenech *mor* ‘lion, leopard’ || S. *\*mir-* and *\*marir-* (redupl.): Gorowa *marir-ika* ‘leopard’ [*P. pardus*] | Alagwa *mariy-amo* ‘wild cat’ | Asa *mer-ok* ‘lion’ [*P. leo*]; (b) N. *\*mir-al-*: Beja *miral-ai* ‘cheetah’ [*Acinonyx jubatus*] | E.: Dullay *\*mir-l-*: Tsamai *mirle* ‘leopard’ [*P. pardus*].

## Data

Below we present the entire material in numerical order.

### 4. Monkeys <sup>4</sup>

#### 4.1. *\*dang<sup>(w)</sup>Vl/r-* ‘monkey; dwarf’

(?) Sem. *\*duggal-*: Arab. *dužžal-* ‘vil, bas, de basses classes (hommes)’.

Egypt. (OK) *dng*, *dʒg* (< *\*dlg*), *dʒng* (< *\*dlnɡ*, met. of *\*dVngVl-*?) ‘dwarf’.

Berb. *\*da(n)gil*: Ghat *a-dažel* ‘singe’, *ta-dažel-t* ‘guenon (*Cercopithecus*)’ [*Cercopithecus sp.*] | Sokna *dágel* ‘monkey’ | Nefusa *a-dənžal* ‘dwarf’.

Chad. *\*da(n)g<sup>(w)</sup>ir/l-*: W.: Hausa *gand-i* (met. < *\*dang-*) ‘pigmy’ || C. *\*da(n)g<sup>w</sup>il/r-*: W. Margi *dagil* | Chibak *dakìl* (devoicing of *\*-g?*) ‘monkey’ | Gude *dánɡwár-á* ‘adult male of an *erd* monkey’ | Dghwede *gándàw-à* (met. < *\*dangw-*) | Buduma *dāgel* ‘monkey’ || E. *\*di(n)g<sup>w</sup>-Vr-*: Ndam *gágēm dōgré* | Dangla *gùdìny-à* (met. < *\*dingw-*) ‘monkey’ (Jungraithmayr, Ibrizimow 1994).

Cush. E. *\*da(n)gir-* ‘monkey’:<sup>5</sup> LEC *\*da(n)gir-*: Somali (var. dialects) *daŋer*, *danžar* | Boni *dašer* (< *\*dašer* < *\*dagir*) | HEC *\*dagir-*: Hadiya *dagier-aa* | Kambatta *dagier-aa*.

▲ Cf. Kordofan Nubian: Dair *tiḡil*, Koldegi *tingel*, Midob *tanni* ‘monkey’ (Blažek Ms. apud Meinhof) and Saharan: Tubu *deḡal* ‘ape’ (Lukas 1953: 183).

► AADB 38, 392.

#### 4.2. *\*guray-* ‘k. of monkey’

Chad. W. *\*garay-*: Polchi *gari* | Dwot *gàri* ‘monkey’.

Cush. E. *\*gVray-*: Yaaku *kərɔi*<sup>6</sup> ‘monkey Colobus’ [*Colobus sp.*].

Omot. *\*gayur-* (met.): N. *\*goyr-*: Malo *gor-o* | Basketo *goyr-a* | Chara *gor-za* | Gimira *gorr*, *gwor* ‘monkey’ | Dizi-Sheko *gyer-u* ‘baboon’ [*Papio sp.*] || S. *\*gayur-*: Ari *goyr-a* | Hamar *guro* | Dime *guur-u* ‘baboon’ [*Papio sp.*].

► AADB 4251.

<sup>4</sup> Some forms are quoted after Blažek Ms., though our understanding of consonantal correspondences and, hence, distribution of the forms among different entries is, in most cases, distinct from this very useful collection of data.

<sup>5</sup> Cf. what looks like a variant root – *\*dink-* ‘dwarf’: LEC: Oromo *dink-i* (borrowed in Mod. Eth *\*dink-* id.) | HEC: Kambatta *dənk-a*, Tembaro *dink-a* (borrowed in Omot. N.: Kafa *dikk-ō* id.). Various authors reconstruct for this root E. Cush. *\*ž-* instead of *\*d-* (Dolgopolsky 1973: 107–8; Blažek Ms.) or *\*z-* (Sasse 1976: 140), probably because of C. Cush. *\*ž-* (see 4.4). The only form probably tipping the scale towards *\*ž-* instead of *\*d-* is Jiddu *yāšire* ‘monkey’: that AA *\*ž-* > Jiddu *y* is a fact, but whether Jid. *-š-* can continue AA *\*-g-* is not clear: if it can, the E. Cush. root should be compared to 4.4.

<sup>6</sup> Yaaku *k* can continue both PAA *\*k* and *\*g*.

## 4.3. \*ka(ʔa)yr- ‘k. of monkey’

Chad. W. \*kayr-: Sura *kaar* ‘red monkey’ [*Ptilocolobus sp.*] | Angas *ker* ‘monkey’, *nker* ‘baboon’ [*Papio sp.*] (Kraft 1981) | Wangday *kàr-é* | Fyer *kìir* | Bokkos *kyer-āḥ* ‘red monkey’ [*Ptilocolobus sp.*]<sup>7</sup>

Cush. E.: LEC *ka(ʔa)r-*: Arbore *kaarr-an* ‘vervet monkey’ [*Chlorocebus pygerythrus*].

Omot. \*ka(ʔa)r- ‘monkey’: N.: Basketo *kar-a* | Male *kaar-a* ‘vervet’ [*Ch. pygerythrus*] | Gemu *kaar-é* | Gofa, Dorze *kar-e* | Oyda, Zasse *kaar-e* ‘baboon’ [*Papio sp.*] | Kafa (argo) *ker-iyo* ‘monkey (Cercopithecus)’ [*Cercopithecus sp.*] || S. *ka(ʔa)r-*: Ari-Bako *kar-ā* | Hamar *kaár* | Dime *kar-e* ‘monkey’.

► AADB 2208. Cf. HSED 1569.

4.4. \*ʒag<sup>(w)</sup>-il/r- ‘k. of monkey’

Chad. \*ʒi(n)gw-al-: C. \*ʒi(n)gway(-l)-: Hildi *dziláḡu* (\**zilang*, met. < \**zingal-*) | Mbuko *zāgày* | Zime-Batna *zēḡū* (or *žēḡu*) ‘k. of monkey’ || E. \*zugul-: Birgit *zúgúl-i* ‘monkey’ (Jungraithmayr, Ibriszimow 1994).

Cush. C. \*ʒag(g)ir- ‘monkey’:<sup>8</sup> Bilin *žäggir-a* (< \**žangir-?*) | Khamir *sažer-a* | Khamta *ziäger-aa* | Qwara *žägir-aa* | Qemant *žegər-aa* || Aungi *zagr-i* | Damot *zagr-ee*.

► AADB 398.

## 5. Canines and hyenas

5.1. \*ʔaway- or \*yawaʔ- (met.) ‘k. of canine’<sup>9</sup>

Sem. \*ʔaway-: Hbr. \*ʔī (pl. *ʔiyyīm*) ‘jackal’ [*Canis aureus*] || Syr. *bənāt ʔaway* ‘thoes, canes aurei’ [*C. aureus*] || Arab. *ʔibnu-l-ʔāwan* ‘animal regardé comme un mélange né d’un chien et d’un renard’ || Tgr. *ʔaw* ‘eatable wild animals’ | Amh. *yäyi* ‘hyena’ [*Hyaenidae sp.*], *awu* ‘hyena’s cry; hyena’ [*Hyaenidae sp.*] | Gur.: Chaha, Gyeto, Ennemor, Muher *awi* ‘wild animal, beast’.

Egyp. (MK) *ʔw* ‘dog’ (met. < \**ʔVyw ?*).

Chad. W. \*(ʔ)iy-: Warji *iyà-nà* | Kariya *í* | Miya *i* ‘dog’.

Cush. E. \*yawaʔ-: LEC \*yaw/yaʔ-: Somali *éy* ‘dog’, *yéey* ‘wild dog’ [*Lyacon pictus*] | Boni *óy?* ‘dog’, *yeye?* ‘jackal’ [*C. aureus*] | Rendille *yááy* ‘wild dog’ [*L. pictus*] | Oromo *yeyy-ii* ‘wolf; wild dog’ [*L. pictus*]<sup>10</sup> | Konso *yoy-ta* ‘hunting dog’ [*L. pictus*] | HEC \*yayy- ‘hunting dog’ [*L. pictus*]: Sidamo *iyäy* | Hadiya *yoyy-akko* | Burji *yeyy-ée*.

► AADB 290. SED II No. 21.

## 5.2. \*ʔaw/ys- ‘k. of canine’

Sem. \*ʔaws-: Arab. *ʔaws-*, dim. *ʔuways-* ‘wolf’ [*Canis lupus*].<sup>11</sup>

Chad. \*ʔaws-: W. \*ʔas-: Mupun, Angas *ās* | Sura *as* | Montol *ʔas* | Tala *ass* || E. \*ʔus-: Bidiya *ʔüs-ú* ‘dog’.

Cush. \*yaʔs- (met.) ‘dog’: N. \*ya(ʔa)s-: Beja *yaas* || S. \*sVʔay (met.): Iraqw *seeʔay* | Gorowa *sooʔay*.

Omot. S. \*ʔaws-: Ongota *ʔóós-e* ‘African hunting dog; jackal’ [*Lyacon pictus*; *C. aureus*].

► AADB 2938. Cf. Blažek 2007 #21; Stolbova 2019 #29 \*ʔ[a]s/c- ‘dog’ (comp. to Arab. *ʔaws-*).

<sup>7</sup> *k-* may continue both PAA \**k-* and \**k̄*.

<sup>8</sup> Borrowed in S. Eth. as \**ʒ/zangi/ar-* ‘baboon’ (Leslau III, 1979: 711)

<sup>9</sup> Very likely descriptive.

<sup>10</sup> Cf. *iyyaa* ‘k. of wild cat’.

<sup>11</sup> Cf. Geez *ʔ/ʔawst, ʔawsənt* ‘eagle, bird of prey’.



## 5.3. \*ʔa/usk-ay- ‘k. of canine or hyena’

Berb. \*ʔusk-ay: Tazerwalt *uskai* | Izayan *usk-a* | Iznassen *uskai* | Ahaggar *osk-a* ‘chien lévrier’.

Chad. \*(Hu)sVk(-ay): W. \*sik-: Dera *šíká(-nánà)* ‘dog crying *nanna*, hyena’ || C. \*sak-ey: Bana *sáká-tá* ‘jackal’, Bachama *sakéy* | Gudu *sakay* ‘dog’ | Afade *skè* ‘hyena’ || E. \*(H)usk-ey: Sokoro *osk-ey* ‘hyena’ [*Crocuta crocuta*] (Stolbova 2009 #187 \*sVk- ‘dog, hyena, jackal’).

Cush. E. \*se(?)ek-: Yaaku *seek-a* ‘hyena’ [*C. crocuta*].

Omot. S. \*ʔask-i: Ari *ʔaksi* (met.) ‘dog’<sup>12</sup> | Ubamer *aksi* (met.) | Ongota *ʔaski* ‘dog’.

▲ Cf. Kuliak: Nyangi *sak* ‘dog’.

► AADB 2716.

## 5.4. \*baʔVš- ‘k. of canine or hyena’

Egyp.: Copt. (Boheiric) *boiši* (f.) ‘vulpes’ [*Vulpes sp.*], ‘a desert animal, peut-être hyéne’ [*Hyaena hyaena*] (Vycichl 1983).

Chad. W. \*bVž-am- ‘hyena’ [*Hyaenidae sp.*]: Ngizim *bžàm-ù* | Bade *ʔà-bžàm-án*.

Cush. N. \*baʔaš-: Beja *baʔaš-o* ‘Fuchs, Schakal’ [*Canis aureus*].

► AADB 4248; EDE II 148.

5.5. \*bawih- ‘k. of canine’<sup>13</sup>

Egyp. (BD) *bh-n* ‘k. of a dog’.

Chad. \*bay/wH-: W. \*bayH-: Tangale *bay* ‘dog’ || C. \*bawH-: Chibak *bouʔ-a, bùh-á* ‘Schakal’ [*Canis aureus*].

Cush. N. \*bayH-: Beja *báyh-o* ‘Schakal, Fuchs, *Canis vulpes*’ [*Canis aureus, Vulpes vulpes*] || E. \*bay-: LEC: Dasenech *bày-c*, pl. *bay-a* ‘jackal’ [*C. aureus*] || S. \*bawih-: Iraqw, Alagwa *bah-a* | Burunge *bay-mo*, pl. *baw-u* (< \*bah-u, acc. to KM) ‘hyena’ [*Crocuta crocuta*] | Dahalo *ḅwéh-a* ‘jackal’ [*Canis aureus*].<sup>14</sup>

Omot. N. \*biw-:<sup>15</sup> Yemsa *biw-á* ‘Schakal’ [*C. aureus*].

► AADB 3774; Blažek 2003: 241; EDE II 147, 282.

## 5.6. \*gVd- ‘k. of canine or hyena’

(?) Sem. \*gad-l- (with fossilized -l?): Arab. *žadlāʔ-* ‘chienne’ (BK 1 267).<sup>16</sup>

Chad. \*gVd-: W.: Hausa *góǝ-èè* ‘dog’ | Ngizim *gádá-múzái*<sup>17</sup> ‘hyena’ [*Hyaenidae sp.*] || C. \*gVd- ‘dog’: Dghwede *ɣəd-ì, gád-è* | Guduf *ɣád-à* | Mofu *gəd-éy* || E. \*gVd- ‘dog’: Ndam *gíy* (< \*gVyd-) | Somrai *dò-gád-à, pl. gárád-ē* | Mokilko *géd-è*, etc.<sup>18</sup>

<sup>12</sup> Cf. *wúksak* ‘fox, jackal’.

<sup>13</sup> Obviously related is the PAA verb \*bwh- ‘to bark (dog)’: Sem. \*n-bh: Akk. *nabāh-u*, Hbr. *nābah*, Arab. *nabh*, Gz. *nabha*, etc.; Eg. *bhn*; Chad. C: Bura *mbwuha*, E.: Kera *bèʔé*; Cush. S.: Ma'a *-boha* (AADB 3774; cf. EDE II 282). While the noun \*bawih- refers to various kinds of canine and even hyena, the PAA verb speaks in favor of a domesticated dog: it is unlikely that the PAA speakers had managed to preserve a special term for wild animals' howling and yowling as a significant notion for over ten millennia.

<sup>14</sup> Beja *h* may continue both AA \*h and \*h, \*-h- in *báyho* is confirmed by -h- in Tgr. *bäyhot*, pl. *bäyhi* with the same meaning obviously borrowed from Beja; in W. Rift and Dahalo, *h* continues only AA \*h, not \*h (acc. to Takács 2011: 115-16); however similar to the present root the S. Cush. forms may appear, they are to be regarded as a variant root \*bayh-; in Dasenech, both AA \*h and \*h yield Ø.

<sup>15</sup> Both PAA \*h and \*h yield Ø.

<sup>16</sup> In Lane II 392, it is quoted as fem. of \*ʔašdal- ‘the hawk or an epithet applied to the hawk’.

<sup>17</sup> *múzái* < Proto-Chad. \*mVz- ‘man, male’.

<sup>18</sup> In CED #196, rel. to *g(V)HVd-* ‘to bite’. Compared to Arab. *gaʔd-at-* ‘wolf’ referring to BK I 299 likely by mistake (there is no such meaning). In Lane Vol. 1: 11, there is *ʔabū ʔaʔd-at-* ‘the wolf’, but this is no more than just

Cush. E.: LEC \**ged-all-*: Oromo *gedall-o* ‘jackal’<sup>19</sup> [*Canis aureus*].

Omot. S. \**gud-r-*: Ari *gudr-i* ‘hyena’ [*Crocuta crocuta*].

► AADB 353. Cf. HSED №856; CED #196.

5.7. \**kur(-ay-)* ‘k. of canine’

Sem. Eth. \**karkur-* (redupl.): Tgr. *kurkur* ‘dog’, *karakur* ‘young dog’ | Tna. *kurkur* ‘puppy, lion cub, whelp’.<sup>20</sup>

Berb. S. \**a-ykar* (met.): Ahaggar, Ayr, E. Wlm. *a-ikar* ‘jeune chien’.

Chad. \**kuray-* ‘dog’: W. \**ku/aray*: Hausa *kàré-e* | Zaar *kùràà-y-í* | Guus *kàràŋ* | Bokus *kyàr-a* || C. \**ku/ir(ay)-*: Fali-Kiria *kár-í* | Higi *kur-é* | Hildi *kr-i* | Lamang *kīrr-ε* | Gisiga *kər-e* | Dugwor *kìr-á* | Sukun *kr-a* | Bura *kil-a* | Buduma *kəl-i* | Logone *kl-e*.<sup>21</sup>

Cush. \**kuray-*: E. \**kayr-* (met.): SA \**kar-*: Saho *kar-e*<sup>22</sup> | LEC \**kayr-*: Rendille *kar* | Baiso *kerr-e* | Arbore *kair, ker* | Dasenech *čir* | Dirayta *herr-aa*, etc. | Dullay \**kar-*: Gawwada *haar-o* | Harso *ħer-o* | Tsamay *kar-o, kar-ito* ‘dog’ || S. \**kuray-*: Alagwa *to-koray-mo* ‘bat-eared fox’ [*Otocyon megalotis*] | Burunge *ta-kuraa?-imo* ‘wild dog’ [*Lycaon pictus*] | Ma’a *kuri* ‘dog’.

▲ Cf. Kanuri *keri*, Teda *kedí* ‘dog’.

► AADB 375; CED #455 (comp. to part of the Cush. terms and Tgr.), HSED 1434.

5.8. \**kawit-* ‘dog’

Chad. W. \**kawt-*: Hausa *kút-íí* ‘dog’ (used only in one set phrase) | Bolewa *kut-ì* | Dwot *kat* | Wangday *kàt* ‘dog’ (Jungraithmayr-Ibriszimow 1994).

Cush. \**ku/it-* ‘dog’ E. SA \**kut-*: Afar *kut-a* | LEC \**kut-*: Konso *kút-a* | Mashile *ħút-a* || S. \**kit-*: Asa *kit-e*.

► AADB 2944.

5.9. \**k<sup>w</sup>ihan-* ‘dog’.<sup>23</sup>

Chad. \**k<sup>w</sup>iHan-* ‘dog’: W. \**k<sup>w</sup>iHan-*: Fyer *k<sup>w</sup>éerj* || E. \**kany-* ~ \**kayan-*: W.Dangla *kàny-à* | E.Dangla *kāny-à* | Migama *kānny-à* | Mabire *kany* | Jegu *kány* | Birgit *káyànj*.

Cush. E. \**k<sup>w</sup>ihan-*: Yaaku *kwehen* ‘dog’.

Omot. \**k<sup>w</sup>iHan-* ‘dog’: N. \**ku/iHan-*: Wolaita, Gamu, Dawro, Zala, Malo, Dache, Koyra, Zayse, Gimira (Bench) *kyan* | Yemsa, Bworo *kan-a* | Kafa, Mocha *kun-ano* | Dizi (Sheko) *kean-u* | Mao (Hozo) *kan-a*, etc. || S. \**kan-*: Dime *kən-ε*, Galila *kan-i*.

► AADB 2195. Cf. CED #441 (Chad. and Omot.), HSED #1425, 1511.

5.10. \**KVzim-* or \**KVžim-* ‘dog’

Egypt. (MK) *tzm* (< \**kVzVm-* or -ž-) ‘dog’.

Berb. \**a-Ku(r)zi/un*: Siwa *a-gurzəni, a-qurzini, əl-gurazən, lu-grāzén* | Audjila *a-γzin* ‘dog’ | Semlal *i-kzin* | Izayan *a-kzin* | Seghrushen *a-qqzin* | Rif, Snus, etc. *a-qzin* ‘young dog, puppy’ | Shawiya *a-γərzul* | Qabyle *a-qžun* ‘dog’.

(?) Chad. C. \**kazim*: Bura *kazim* ‘a male baboon, monkey, or dog’.

---

a figure of speech. Another comparison, with Qabyle *agdi* ‘dog’, is also wrong since it is from \**aydi*, the main Berb. term for ‘dog’.

<sup>19</sup> Cf. also Somali *godgoddó* ‘type of shark’.

<sup>20</sup> Cf. *wākaru, wākarya* ‘fox’ < Saho *wakari* ‘jackal’ [*Canis aureus*]?

<sup>21</sup> < \**kVr-* in CED #455.

<sup>22</sup> Cf. Saho *wakari*, Afar *wākri* ‘jackal’ [*Canis aureus*].

<sup>23</sup> Very likely related is Canarian \**kun-*: Gran Canaria *cuna* ‘dog’ | Tenerife *cancha* ‘little dogs’, *cuncha* ‘chien’ (-*cha* < \**-ta* is probably a diminutive suffix).

Cush. C. \**gizim-* and \**gišim-* ‘dog’: Bilin *gidin*, pl. *gižin* | Khamir *gizin* | Khamta *gózəŋ*, pl. *gəsəŋ* | Qwara *gezen* | Qemant *gəzəŋ* | Kailiña *gəzəŋ* | Waag *gizin* | Falasha *gəzín* || Aungi *gsén* | Kunfāl *kassəŋ*.

▲ Cf. Sandawe *gwece* ‘wild dog’.

► AADB 2699.

5.11. \**wahr-* ‘k. of canine; hyena’<sup>24</sup>

Egypt. (late) *whr.t* ‘Hündin’ (prop. name) | Copt. *ʔuhor* ‘dog’.

Berb. \**a-wwur* (<\**a-wHur*): Ahaggar *ǎ-ggur* | E. Tawllemmet *a-ggur* | Ayr *é-ggur* ‘chacal’ [C. *aureus*]<sup>25</sup>.

Chad. \**hayr-* ‘dog’: C. \**har-*: Zina *hàr-i*.<sup>26</sup>

Cush. \**warH-* ‘hyena’ [*Hyaenidae* sp.]: E. \**warH-ab-*: LEC \**warH-ab-*: Konso *oray-ta* | Somali *waraab-e* | Oromo *waraab-esa* | Rendille *warab-a* | Dirayta *waraab-e* | Dullay \**war(H)-ab-*: Gawwada *oraap-atte* | Harso *araap-ičče* | Gollango *oraap-atte* || S. \**war-*: Ma’a *war-é*.

▲ Cf. Sum. *ur* [dog] (PSD, uncertain), *ur*, *ur<sub>5</sub>*, *ur.ri* ‘hairy animal, dog, lion, etc.’ (EDS #2782), *ur* ‘chien’ (LSF).

► AADB 3189.

5.12. \**wanš-* ‘k. of canine’

Egypt. (OK) *wnš* (met.) ‘wolf’ [*Canis lupus*].<sup>27</sup>

Berb. \**wVššin* (met. with redupl.) ‘jackal [C. *aureus*]’: Ghadames *weššin* | Qabyle, Nefusa *uššen* | Semlal *uššen*, etc.

Cush. E. \**wašš-*: HEC \**wašš-* (met. < \**wanš*): Sidamo *woš-iččo* | Kambatta *woš-ičču* | Burji *wačč-o*, pl. *wašš-a* ‘dog’<sup>28</sup> | Dullay \**woš-*: Gollango *oš-e* ‘Hyänenhund; Lycaon apictus somalicus’ [*Lycaon pictus*].<sup>29</sup>

▲ Blažek 2008 cites NS parallels: Nara *wos*, Taman: Merarir *wiis* ‘dog’ which may be comparable with 4.3. \**aws-* ~ \**yaʔs-* ‘k. of canine’ as well.

► AADB 291. Cf. Blažek 2007 #21.

<sup>24</sup> Cf. NAA \**harnag-* ~ \**hirguan-* ~ \**gurhan-*: Sem. \**harnag-*: Syr. *harnāg-ā* ‘golden jackal’ [*Canis aureus*] (cf. also Arab. *žāriḥ-at-* < \**gariḥ-* ‘bête ou oiseau de chasse (chien, guépard, faucon)’); Canarian \**hir(g)wan-*: Palma *hirguan* ‘...demonio in figura d’uomo lanuto’, *irvene* ‘el demonio en figura de perro lanudo’, *irvene* ‘apparitions’, *yrvene* ‘el Diablo’ (Wölfel 1965: 484, entry “Der Hund”); Chad.: C. \**gurhan*: Hildi *gərhan*, ‘jackal’ [C. *aureus*], Musgu *hérge*, Munjuk *hirge* ‘dog’, Musgoy *gurnai*, *ɣurnai*, Masa *ɣurnaita*, Banana *gɔnira*, E. \**gurnV*: Kera *gòrnòy*, Kwang *gòrény*, Kabalai *gwòrrnày* ‘hyena’ [*Crocuta crocuta*] (Jungraithmayr, Ibriszimow 1994: 107, 205; CED #224a.). The entire bunch of synonyms with somewhat similar consonantal roots, or “root variants”, may point to an ancient tabooing of a term denoting a special kind of canine.

<sup>25</sup> In Nait-Zerrad 2002: 856, placed under \**gwr*, though in Foucauld I: 480, the grapheme **X** (usually transliterated as *gg*) is used; in Ahaggar *tifinav* it renders the voiced stop *g* originating from \**-gʷ* < \**-ww* < Proto-Berber and Proto-Afrasian \**w* and thus opposed to the voiced velar fricative *ǧ* (rendered by **ʔ**) < Proto-Berber and Proto-Afrasian \**g*. There are also such forms as the somewhat enigmatic Senhadja *i-uhar*, Ait Tuzin *a-whar*, Ait Warain *uhar*, etc. ‘fox’, formally well compatible with this root but containing an unusual laryngeal, normally lost in North Berber.

<sup>26</sup> In Stolbova 2019 #153, the Zina term is combined under \**hVr-* ‘jackal’ with Musgu *hérge*, Munjuk *hirge* ‘dog’ analyzed after H. Tourneux (personal communication) as \**hir-ge* ‘jackal+house’ (cf., however fn. 25 above) and compared to the Egyptian and Coptic terms and Berber \**-whar* ‘fox’ (see fn. 26) after EDE I: 144.

<sup>27</sup> Cf. also NE *iš* pl. ‘dogs pulling the ship of the Sun-god’.

<sup>28</sup> Cf. also Darasa *walčo*, *warša* with comments in Blažek 2007 #21: *-r/-l-* originating via dissimilation of the primary geminate.

<sup>29</sup> Omot. N. \**wayši*: Haruro *wayše*, Mao (Hozo) *wiši*, (Sezo) *wiš(š)i* ‘dog’ – probably borrowed from HEC.

## 6. Felines

## 6.1. \*ʔaʒur- ‘k. of feline, hyena or viverra’

Sem. \*ʔa(n)dar-: Akk. *azar-u* (*azzar-u*) ‘lynx’ [*Vulpes sp.*] || Gz. *ʔanzar* ‘wild cat’ [*Felis silvestris*].  
 Berb. S. \*-zur-Vy ‘hyena’ [*Hyaenidae sp.*]: Ahaggar *tă-hûr-i* | Ayr, E. Wlm. *tə-zor-ay*.

Chad. W. \*ʒaʔur- (met.): Galambu *zər-əm* | Warji *ʒara-waš* ‘lion’ [*Panthera leo*] | Zul *zaʔār-i* ‘like a hyena or civet’ || C. \*ʒVr-: Bachama *ʒār-á* | Bata *ʒír-e* | Buduma *zāzúr-mà* (redupl.) ‘leopard’ [*P. pardus*] || E. \*ʒur- ‘id.’: Gabri *ʒur* | Mubi *ʒyúr-úk*.

Cush. E.: LEC \*ʔadur(r)-: Oromo *adurr-ee* | Dasenech *adur-e* ‘viverra’ [*Viverra sp.*]<sup>30</sup> || (?) S. \*ʒiʔVr- (met.) ‘viverra’ [*Viverra sp.*]: Alagwa *ʒeʔir-a* | Burunge *ʒiʔerar-e*.<sup>31</sup>

Omot. N. \*ʒi(?)ar-: Kafa *yeer-oo*, *ʒār-o* ‘viverra’ [*Viverra sp.*].<sup>32</sup>

► SED II No. 9; AADB 399.

## 6.2. \*ʕariw- ~ \*ʕiraw- ‘k. of feline’

Sem. \*ʕariw-: Arab. *ʕurw-at-* ‘lion’ || Tna. *waʕro* (met.) ‘leonessa’ (Bassano1918: 653) || Mhr. *ʔáyr-i*, pl. *ʔarēw-at* | Jib. *ʕér-i* ‘tom-cat’.

Berb. \*Hawr- ‘lion’: Ghadames *abur* | Ahaggar, Ayr, E. Wlm. *a-har* ‘lion’ (Prasse 1969 #362).

Chad. C. \*Hiraw-: Musgu *à-hiràw* ‘leopard’ | Munjuk *a-hraw* ‘panther’.<sup>33</sup>

Cush. E.: LEC: Somali *ʕúrr-i* ‘tom-cat’.

▲ In spite of scarce representation, the relatively infrequent combination of radicals seems to confirm the comparison. Note the same meaning in MSA and Somali (a borrowing?).

► SED II No. 41; AADB 2658.

## 6.3. \*baʔay- ‘leopard or lion’

(?) Eg. (MK) *bʔ* (if <\*bVʔ-) ‘panther’ [*Panthera pardus*].

Chad. \*bay-: C.: Gude *bwáy-à* || E.: Mubi *boy-á* ‘leopard’ [*P. pardus*].

Cush. S. \*baʔ-: Dahalo *baʔ-i* ‘lion’ [*P. leo*].<sup>34</sup>

Omot. \*bay- ‘lion’ [*Panthera leo*]: N.: Sheko *bay-a* | Nao *bay-ê* || S.: Dime *biy-u*.

► Cf. EDE II 22.

6.4. \*baʔis- ‘k. of feline (*Felis silvestris*)’

Sem.: Arab *bass(-at)-* ‘domestic cat’.<sup>35</sup>

Egyp. (OK) *bʔs.t.t*, name of the cat-goddess Bastet.

(?) Berb. N.: Nefusa *bišiu* ‘gatto’ (isolated).

Chad. E. \*bi(?)s- ‘cat’: E. Dangla *biisi* | Migama *biisú*.

Cush. N.: Beja *bēs-a*, *biss-a* ‘cat’ || E.: HEC: Sidamo *bas-u*, *bas-o* ‘cat’<sup>36</sup> || S.: Qwadza *baʔas-iko* (pl.) ‘cheetah’.<sup>37</sup>

▲ The Nefusa and E. Chad. forms are likely inter-borrowings.

► AADB 4246; EDE II 86–7.

<sup>30</sup> Sidamo *adurre* ‘cat’, Dullay \*ʔadurr- (Dihina, Gollango *aturre* ‘wild cat’): the Sidamo and Dullay forms look like lws. from Oromo in spite of the difference in meaning.

<sup>31</sup> The two W. Rift forms absent from Kiessling, Mous 2003 are cited in Dolgopolskiy 1973: 300 after Fleming.

<sup>32</sup> A lw. < Cush. ?

<sup>33</sup> Acc. to CED 42, *h* in this language group can continue \*ʕ.

<sup>34</sup> Cf. C. \*ʔib- (metathesis?): Qwara *ib-ā*, Qemant *yiv-a* ‘leopard’ [*Panthera pardus*] — rather a metathesis < \*ʔib- than < \*l-b suggested in EDE II: 22.

<sup>35</sup> Cf. Amh. *biss* ‘interjection pour chasser un chat’.

<sup>36</sup> Cf. also *basurr-i* ‘cat’ (composed < *bas* and *ʕur-?* Cf. \*ʕur- ‘k. of feline’).

<sup>37</sup> Cf. *paʔas-iko* ‘cat’.

6.5. *\*da(?)y)m-* ‘k. of feline’

Sem. *\*dam-*, *\*dimm-* and *\*dumam-* (redupl.): Akk. (SB) *dumām-u* (*tumām-u*) ‘a wild animal’ (CAD d 179), ‘Gepard’ [*Acinonyx jubatus*] (AHw: 175) || Arab. *dam-*, *dimm-at-* ‘chat’<sup>38</sup> || Gz. *damm-at* | Tgr., Tna. *damm-u* | Amh. *damm-ät* ‘cat’.

Chad. *\*dam-*: W. *\*dam-*: Yergum *dam-uŋ* ‘leopard’ [*Panthera pardus*] and *\*(?Vn-)da/um-*: Sura *ndùmú*, Gerke *damu*, Ankwe *tumu* ‘hyena’ [*Crocuta crocuta*]<sup>39</sup> || E. *\*dam-* and *\*damd-* (redupl.): Lele *žem* (<*\*dem-*?) ‘caracal’ [*Caracal caracal*] | Sokoro *dámd-ē* (redupl.) ‘leopard’ [*P. pardus*].

Cush. S. *\*dužum-* ‘leopard’ [*P. pardus*]: Iraqw, Alagwa, Burunge *dužum-a* | Asa *dužum-ok* | Qwadza *dužum-ayi*.<sup>40</sup>

► SED II No. 70; AADB 350.

6.6. *\*giwar-* ‘k. of feline’

Sem. *\*girr-*, *\*gu/ary-* and *\*gigVr-* (redupl.): Akk. *girr-u* ‘lion’ [*Panthera leo*] || Phoen. *gr* ‘lion whelp’ | Hbr. *gūr* ‘cub (lion, jackal)’, *gōr* (pl. only) ‘lion's cub’ || Aram. (D. Alla) *gr* ‘whelp (of a fox)’ | Syr. *gury-ā* ‘catulus (leonis, canis, suis, serpentis)’ | Mand. *guri-a* ‘whelp, cur, young dog, pariah-dog’ || Arab. *žarw-*, *žirw-*, *žurw-* ‘petit de chien, de lion ou de toute autre bête carnassière’<sup>41</sup> || Tgr. *gərž-ən*, pl. *gärägg-ən* ‘cub of lions or of leopards’ | Harari *gärgōr-a* (also *gän-gōr-a*) ‘leopard’.

Chad. *\*giwar-*: W. *\*giwar-*: Dwoth *gewar* ‘an animal smaller than civet’ || C. *\*(?a-m-)gwar-*: Hwona *žəm-ŋwara*, *žə-ŋwara* ‘cat’ | Fali-Muchella *mùŋ-gùr-i* | Fali-Bwagira *mùŋ-gwùr-žin* ‘wild cat’.

Cush.<sup>42</sup> (a) E. *\*gir-* ‘cat’: LEC *\*gir-*: Dirayta *kiro-ta* | HEC *\*gir-*: Burji *giraaw-ee* (met. <*\*?a-giraw-*?)<sup>43</sup> and (b) *\*gar-m-* ‘lion’ [*Panthera leo*]: LEC: Konso *karm-aa* | Dirayta *karm* | HEC: Burji *gárm-i* | Dullay: Gawwada-Dalpena, Harso, Dihina, Gollango *kárm-o*.

Omot. (a) N. *\*gawar-* ‘cat’: Zala *gawar-ya* | Dawro, Zaysse *garaw-a* | Malo, Gamu, Dače *gawar-a* | Zaysse *garaw-a* | Koyra *giraaww-e?*<sup>44</sup> and (b) N. *\*gar-m-* ‘lion’ [*P. leo*]: Zaysse, Koyra *garm-a*.<sup>45</sup>

► AADB 2629; cf. SED II No. 82.

6.7. *\*labiž-* ‘lion’<sup>46</sup>

Sem. *\*labiž-*: Akk. *labb-u* (*labž-u*, *lāb-u*) || Ugr. *lbž-u* ‘lion’ [*Panthera leo*] || Hbr. *lābī(?)* ‘lioness’ || Sab. *lbž* ‘lion, lioness’ || Arab. *lubaž-at-*, *labuž-at-*, *libw-at-*, etc. ‘lionne’.

<sup>38</sup> Regarded by some authors as a lw. from Gz.

<sup>39</sup> Cf. also Hausa *dāmis-à* ‘leopard’ and Berb. Tuareg *daməs-a* ‘panther’ (obviously inter-borrowing) with non-etymological *-s-*.

<sup>40</sup> Beja *žímo*, *dimmo* ‘cat’ – perhaps < Eth. or Arab. Cf. Bilin *dammu* (pl.), Qwara *damyā*, Qemant *daməya* ‘cat’; whether the Eth. term is the source of the Agaw one or the latter is inherited is hard to decide; SA *\*dumm-*: Saho *dummu*, Afar *dumoo* ‘cat’ can be Ethiopisms or genuine terms. LEC: Somali *dummad* ‘cat’ is likely < Arab. or Amh. HEC: Hadiya *aduuna*, *adun-čo*, Kambatta *adan-ču*, *adani-ta* ‘cat’, both <*\*?a-dalun-t-*, likely <*\*?adum-t-*, are probably genuine.

<sup>41</sup> Cf. Arab. *žarwal-* ‘serwal, lynx’ [*Caracal serval*, *C. caracal*], likely related with fossilized *-l* suffixed.

<sup>42</sup> Cf. S. Cush.: Dahalo *ŋgūro* ‘small black longtailed rodent’ (<*\*?a-ngur-*?).

<sup>43</sup> Sasse (1982: 85) compares it with a completely identical Koyra word; one of the two words is undoubtedly borrowed, but which one is hard to decide. The ending in *-?* seems uncommon in both languages.

<sup>44</sup> See fn. 170.

<sup>45</sup> While the difference in stem patterns in E. Cush. *\*gir-* and N. Omot. *\*gawar-* ‘cat’ more than likely points to a common origin, full similarity in reflexes of *\*gar-m-* ‘lion’ likely implies borrowing (in either direction).

<sup>46</sup> Cf. metathetic (?) *\*ribal-* attested in Arab. and Chad., either to be reconstructed at the NAA level or a result of independent parallel processes, accounted for by a very low compatibility of *l* and *r* in most AA languages: Sem.: Arab. *ribāl-* ‘lion; féroce, rapace (loup)’ and Chad.: W.: Sura *reḫəl* ‘wild cat’, (?) Hausa *rābbī* (with the fallen *-l?*), C.: Zeghvana *šrval*, Glavda *ārðvar-e*, Gava *žurwvar-a*, Lamang *šrvar-é*, E.: Somrai *lārb-š* (met.) ‘leopard’ [*Panthera pardus*].

Egyp. (OK) *rw*<sup>47</sup> *ʒbw*, inscription under the image of a lion<sup>48</sup> | Copt.: Bohairic, Sahidic *laboy* ‘lion; bear’ [*P. leo*; *Ursus arctos*].

Chad. *\*la/ib-*: W. *\*lab-*: Tangale *lab-ata* (f.) ‘lioness’ [*P. leo*] || C. *\*(ʔa-)lib-ar-*<sup>49</sup> ‘lion’ [*P. leo*]: Hwona *lifār-i* | Margi *ḥa-livār-i* | Chibak *ʔa-lvār-i* | Kilba *lēvār-i* | Kapsiki *nivèr-i* (dissim.) | Higi-Nkafa *livèr-i* | Gude *livyàr-á*.

Cush. E. *\*lu/ib-* ‘lion’ [*P. leo*]: SA *\*lub-ak-*: Saho, Afar *lubaak* | LEC *\*lib-aḥ-*: Somali *libaḥ* | Jiddu *libááh-u*.

▲ Cf. PIE *\*liw-* ‘lion’.

► SED II No. 144; HSED 1636; EDE I 61; CED #555. AADB 2279.

#### 6.8. *\*layč-* ‘k. of large feline’

Sem. *\*layt-* ‘lion’ [*Panthera leo*]: Hbr. *layiš* || Aram.: Sam. *lyt* | Jud. *lēt-ā*, *layt-ā* || Arab. *layt-*.

Chad. *\*lič-*: W.: *\*liš-um*: Sura *lūšùm* | Chip *lišim* ‘leopard’ [*P. pardus*].<sup>50</sup>

Cush. N. *\*loliš* (redupl.): Beja *loliš*, *noliš* (dissim.) ‘cat’.

► AADB 2646; cf. SED II No. 147.

#### 6.9. *\*mary-* or *\*mayr-* ‘k. of feline’

Sem. *\*namir-* (<*\*na-mir-* or *\*na-mayr-* with the fossilized prefix) ‘leopard’ [*Panthera pardus*]: Akk. *nimr-* (*nammar-*) ‘panther’ || Hbr. *nāmēr* ‘leopard, panther’ || Aram.: Anc., Off. *nmr* | Bibl. *nəmar* ‘panther’ | Mand. *namar* (*nimar*, *namr*) ‘leopard’ | Sab. *nmr* ‘leopard, panther’ | Hdr. *nmr* ‘panthère’ || Arab. *nimr-* ‘panthère’, *namir-* ‘léopard, panthère’ || Gz. *namr* | Tna., Tgr. *nābri* ‘leopard’ | Amh. *nāmr*, *nābər*<sup>51</sup> ‘lynx’ [*Caracal caracal*] | Sel. *nāwər* | Wol. *nāwr*.<sup>52</sup>

Egyp. (OK) *mʒy* ‘lion’ [*Panthera leo*] (if < *\*mVry*).

Chad. *\*mary-* and *\*mumur-/mumur-* (redupl.): W. *\*mumur-*: Dera *mūmur-u* ‘cheetah’ [*Acinonyx jubatus*] | Polchi *mur* ‘jackal’ [*Canis aureus*] | Bokkos *mūrûm* | Sha, Daffo-Butura *mūrûm* ‘hyena’ [*Crocuta crocuta*] || C. *\*mary-*: Mafa *mariy-žélé* ‘lion’ [*Panthera leo*] | Lame *mēr* ‘serval’ [*Caracal serval*], *méri-án* ‘wild cat’ | Peve *méri-àn* ‘Id.’ | Zime-Dari *mīēr* ‘genette’ [*Genetta sp.*].

Cush. (a) *\*mary-*, *\*mawir-* and *\*marir-* (redupl.): E.: LEC *\*mawr-*: Oromo *mōr-ē* ‘civet(-cat)’ [*Civettictis civetta*] | Dasenech *mor* ‘lion, leopard’ || S. *\*mir-* and *\*marir-* (redupl.): Gorowa *marir-ika* ‘leopard’ [*P. pardus*] | Alagwa *mariy-amo* ‘wild cat’ | Asa *mer-ok* ‘lion’ [*P. leo*]; (b) N. *\*mir-al-*: Beja *miral-ai* ‘cheetah’ [*Acinonyx jubatus*] | E.: Dullay *\*mir-l-*: Tsamai *mirle* ‘leopard’ [*P. pardus*].<sup>53</sup>

▲ Cf. NS *\*mEr* ‘leopard, lion, cat’; Songhay *mar* id.; ESud.: Nandi *merindo*, Suk *meril* ‘leopard’; Kuliak: Nyangi *merihl* id.; SNil.: Tatoga *marir-d*, Suk *mèrīl*, Sogoo *melil-tɔ* (quoted in EDE III: 37). In spite of the striking similarity between these and some of the AA forms, all seem genuine in their families.

► AADB 2305; SED II No. 164. Cf. HSED 1760 and EDE III 37, 134-5.

#### 6.10. *\*sawr-* and *\*sarw-* ‘k. of feline’

Sem. *\*šur-ān-* ~ *\*šinnawr-* (met.) ‘wild cat’ [*Felis silvestris*]: Akk. *šurān-u* ‘wild cat’ || Aram.: Anc. *šrn* ‘wild cat’ | Jud. *šunnār-ā*, *šūrān-ā* ‘cat’ | Syr. *šūrān-ā* ‘felis; mustela, animal quod vorat

<sup>47</sup> Eg. *rw* (Pyr.) is also ‘lion’ (unless < *\*IVw*) < *\*(ʔa-)rVw-* ‘k. of large feline, lion’ not quoted here since it is rather a NA than a PAA root (with only one parallel in SA: Cush. E.: LEC: Somali *ār* ‘lion’ which is probably not enough for a reconstruction at the PAA level).

<sup>48</sup> Also (18 Dyn.) *ʒby* ‘panther’ [*Panthera pardus*] (if < *\*IVb(V)y-*).

<sup>49</sup> With the extension *-ar* – acc. to CED #555, a marker of harmful animals.

<sup>50</sup> Cf. metathetic E. *\*ʔa-čil-*: Bidiya *ʔāčil-o* ‘serval’ [*Caracal serval*].

<sup>51</sup> The parallel variant root *\*nabr-* in Mod. Eth. is likely a result of dissimilation by nasality.

<sup>52</sup> Mhr. *nāmer* ‘leopard, tiger’ and Hrs. *nemr* ‘leopard’ are likely Arab. lws.

<sup>53</sup> S. Omot.: Ongota *mirila*, *morle* ‘serval, genet’ is a lw. from Tsamai.

gallinas', *šūnār-ā* 'felis' | Mand. *šunar-a* 'cat' || Arab. *sunnār-*, *sinnawr-* 'chat' || Mhr. *sənnáwr-ət*, Hrs. *sennōr-eh* | Jib. *sínór-t* | Hbt. *sənnóor-ət* 'cat'.<sup>54</sup>

Chad. \**sawar-*: W. \**sa(w)r-* and \**sasar-*, \**sawawr-* (redupl.): Hausa *sar-a*, *sawawar-a* 'civet' [*Civettictis civetta*] (Cosper 1994 #264)<sup>55</sup> | Polchi *šišer* 'smaller than civet' | Daffo-Butura *šúwír* 'Wildkatze' [*Felis sp.*] || E. \**sawar* ~ \**sarar-*: Tumak *šàwàr* 'chat doré' [*Caracal aurata*] | Dangla *sàrààr-à* 'panthère femelle' [*Panthera pardus*].

Cush. E. \**sarw-* and \**warwars-* (redupl. and met.): Dullay \**sar-iko*: Gawwada, Gobeze, Harso, Gollango *sar-iiko* 'leopard' [*Panthera pardus*] || Yaaku *warwars-an* (met.) 'serval' [*Caracal serval*].

Omot. N. \**sawr-*: Ganjule *šuur-o* 'cat'.<sup>56</sup>

► AADB 2651; SED II No. 206.

There are two more terms for large felines reconstructed by Václav Blažek (2013) and qualifying as PAA according to the present paper's criteria (representation in both NAA and SAA branches): \**camik/y/h-* (Cush. C., E.; Omot. N.; Chad. C., E.) and \**zi(n)g-(um)-* (Arab.; Cush. E.; Chad. W., C.)

## 7. Other mammals (a selection)

### 7.1. \**?V(n)čaw-* 'k. of smaller carnivoran or rodent'<sup>57</sup>

Sem. \**?ay(n)šaw-*: Akk. (OB on) *ayāš-u* 'weasel' [*Mustela sp.*] || Gz. *ʔanšaw-ā*, *ʔanšew-ā*, *ʔanšow-ā* 'mouse'<sup>58</sup> and 'weasel' [*Mustela sp.*] | Tna *ʔančəw-a* 'rat, mouse' | Tgr. *ʔanšay*, *ʔanšay*, | Amh. *ay(ə)ʔ* 'mouse'.

(?) Berb. S. \**-dVway*: Ayr, E. Wlm. *e-ḍəwi* 'jerboa' [*Dipodidae sp.*].

Chad. W. \*(*Ha-*)*Čiy-*: Hausa *čiy-ō* 'field rat' | Sayanchi *ààcá* 'mouse' || C.: Musgu *ausi* 'Maus'.

Cush. C. \**?i(n)čaw-*: Bilin *inšuw-aa* | Khamir *iečuw-aa* | Khamta *ačuw-aa* | Qemant *yešw-aa* || Aungi *enč-aa* | Damot *inč-ii* 'mouse' || E.: LEC \**waawač-* (met.): Oromo *waawač-oo* 'mongoose' [*Herpestidae sp.*].

Omot. N. \**?i(n)čaw-* 'rat': Wolaita *uča-a* | Kafa *ičo-o* | Bworo *inčo-o*, *iinč-a* | Gimirra *uč/ç* | Dizi (Sheko) *iičč-o*.

► SED II No. 26; AADB 379.

### 7.2. \**?andaw-* 'mouse'

Chad. C. \**ma(n)dVw* (likely < coll. \**ma-ʔa(n)daw-*) 'rat': Mofu *m-àndùw-ánj* | Gisiga *m-onduw-anj* | Muktele *m-ádàw-á* | Matakam *m-ùdùw-à*.

Cush. E. \**?andaw-*: SA \**?andaw-*: Saho *andəw-a* | Afar *andaw-aa* 'mouse, rat'.<sup>59</sup>

► AADB 285.

### 7.3. \**?ar-* 'hare'

Sem. (a) \**?ar-an-*: Arab. (pl.) *ʔarān<sup>m</sup>* || Jib. *ʔern-í* 'hare' [*Lepus sp.*];<sup>60</sup> (b) \**?ar-n-ab-* 'hare' [*Lepus sp.*]: Akk. (Oakk. on) *arnab-u* (*annab-u*) || Hbr. *ʔarnāb-āt* | D.-Alla *ʔrnb-n* (pl.) | Syr. *ʔarnāb-* ||

<sup>54</sup> The MSA forms can be Arabisms.

<sup>55</sup> Boghom *sawawara* 'like a hyena or civet' is most likely < Hausa.

<sup>56</sup> Isolated in Omotic but, perhaps, genuine: borrowing from E.Cush. semantically and phonetically improbable.

<sup>57</sup> \**-č* instead of \**-ç* is reconstructed mainly based on the Berb. forms: if they belong here, *d* can continue AA \**č* (beside AA \**ʔ*) but not \**ç*. In Akk. and Eth., as well as in Oromo, reflexes of \**č* and \**ç* coincide; in Chad. they remain obscure in the intervocalic position (but coincide in Hausa); the Agaw and Omotic reflexes are contradictory.

<sup>58</sup> Also *hanšaw-ā*, *hanš-e* id. with the fossilized prefix *h-*.

<sup>59</sup> Cf. LEC \**?antu-*: Oromo *antu-ta* 'mouse'.

<sup>60</sup> Cf. also Aram.: Syr. *ʔarnāfā* 'mus magnus', perhaps related with a non-etymological -*f*.

Arab. *ʔarnab-* || Gz. *ʔarnāb* | Har. *ḥarbāññ-o* (met.) | Gur. *\*arbāññ-ä* (met.)<sup>61</sup> || Hbt. *ḥarniib* | Mhr. *ḥarnáyb* | Hrs. *ḥeynēb*.<sup>62</sup>

Egyp. (old) *wn* ‘hare?’<sup>63</sup>

Chad. (a) *\*ʔarn-*: W.: Hausa *ánná-kó* (assim.) ‘a variety of small hare’ [*Lepus sp.*];<sup>64</sup> (b) *\*ʔa(n)bar-* (met.) and *\*nabir-* (> *\*ribun-*): W. *\*ʔa(n)bar-*: Gerka *tu-bbar*<sup>65</sup> | *ti-ḥaar* (< *\*ti-baʔar* or *\*ti-ʔabar*) | Dera *búr-kiŋ* || C. *\*bir-*: Kapsiki *vir-a* ‘rabbit’ [*Lepus sp.*] | Podokwo *viŋur-a* ‘lapin’ [*Lepus sp.*], etc.; *\*nabir-* and *\*ribun-* (both met.) ‘hare’ [*Lepus sp.*]: Wandala *navir-e* | W. Margi *ta-ribuna* | Malgwa *navire* | Wandala *nàvir-e* || E.: Mobu *tì-bèr* ‘rabbit’ [*Lepus sp.*] | Mubi *hom-bur-o* ‘rabbit’ [*Lepus sp.*].

Cush. *\*ʔar-*: N. *\*ʔar-*: Beja *ar-aat* ‘young hare, leveret’ || C. *\*b/mäntäl/r-* (met. < *\*ta-rnab-*) ‘hare, rabbit’ [*Lepus sp.*] | Bilin *mäntälä-rä* | Khamtanga *bitl-a* | Qemant *bäntär-a* (Appleyard 2006: 79-80) || S.: Alagwa *ʔorh-óo* | Burunge *ʔoroʔ-oo* ‘big hare sp.’ [*Lepus sp.*].

▲ Cf. PKoman *\*warVn-* ‘hare’. Multiple assimilations and metatheses may be due to poor compatibility of *\*r* and *\*n* in the same root, while various root extensions may be accounted for by assuming tabooing of the hare’s name as an object of hunting or a totem.

► SED II No. 14, AADB 3477; EDE I, 44; Stolbova 2019 #26; CED #11.

#### 7.4. *\*bawr* - ‘k. of a large carnivorous animal’

Sem. *\*ba(r)bar-* (redupl.):<sup>66</sup> Akk. *barbar-u* ‘wolf’ [*Canis lupus*] || Arab. *babr-* (pl. *bubūr-*) ‘espèce de chacal qui conduit le lion sur la piste de la proie’ [*C. aureus*], *ʔal-barbār-* ‘lion’ [*Panthera leo*].<sup>67</sup>

Berb. *\*barr-an-*: Ahaggar *ǎ-barr-ân* ‘chat sauvage’ (not in Foucauld 1951-52) | Shilh *a-br-an* ‘wild cat’<sup>68</sup> | Shilh *a-bayrr-u* ‘fox’ (Naït-Zerrad 1998: 146).

Chad. *\*bawr-*: W. *\*bawr-*: Hausa *bār-ū* ‘hyena’ [*Crocota crocuta*] | Sura *m-bóór* | Angas *bwār* ‘lion’ [*P. leo*] | Mupun *pùpwór* | Tangale *par-i* ‘cheetah’ [*Acinonyx jubatus*] | Ngamo *bàràyr-iyà* (redupl.) ‘wild animal’ | Zul *m-bor-i* | Fyer *ʔm-bwaar* ‘lion’ [*P. leo*] || C. *\*ma-bawr-*: Bura *mo-bul-u* (*l* < *\*r*) | Mwulien *m̀búr-mò* | Bachama *mbur-uməy* ‘hyena’ [*C. crocuta*] | Mofu *má-bár-* ‘panther, lion’ [*P. pardus*, *P. leo*] | Gisiga *mo-bor* ‘lion’ [*P. leo*] || E. *\*bawr-*: Dangla *bùur-ì* ‘lion’ [*P. leo*] | Migama *bòdr-ú* | Bidiyo *bóor-e* | Mubi *bór-ì* ‘hyena’ [*C. crocuta*] | Birgit *bàr-á* | Toram *bar-a* ‘jackal’ [*Canis aureus*].

Cush. *\*bawr-*: S. *\*bawr-*: Dahalo *ḥór-a* ‘any dangerous animal’.<sup>69</sup>

▲ Cf. Sum. *urbara* [wolf], wr. *ur-bar-ra*; *ur-bar*; *ur-bar-ra<sup>sar</sup>*; *ur-ba-ra*; *ur-<sub>2</sub>-bar-ra* (PSD), *ur<sup>w</sup>bar.ra*, *ur.bar*, *ur.bar.ra* ‘wolf’ [*Canis lupus*] (EDS: 304), *ur-bar-ra-k*, *ur-bar-ra* s. ‘loup’ (LSF). Cf. P.Masai *\*-barie* ‘jackal’ [*C. aureus*].

► AADB 1863, 2621. Cf. HSED 246; EDE II 22, 148 and Stolbova 2021 #281 (compared to Berb.: Ghadames *a-bur*, Ghat *ahər*, Ahaggar *a-har*, etc. ‘lion’; comparison debatable).

<sup>61</sup> All of the above Eth. forms were considered Arabisms by Leslau (Leslau 1987: 38), though it seems correct only for Gz. Cush. E.: HEC: Burji (*h*)*irbáan-čoo* ‘rabbit’ [*Lepus sp.*] and Omot. N.: Wolaita *hirbaan-e* ‘hare’ [*Lepus sp.*] are Eth. loans.

<sup>62</sup> In MSA, *ḥ-* is an article before *\*ʔ-*. Cf. also Ugr. *ʔanhb* ‘hare?’ (more likely ‘snail’) and Arab. *ḥarnab-*, Gz. *ḥarnab* ‘hare’ obviously related with enigmatic *\*ḥ-*.

<sup>63</sup> The hieroglyph for *wn* depicts a hare; acc. to EDE I: 44, perhaps < *wʔn* < *\*wrn*.

<sup>64</sup> Acc. to CED #11, < *\*ʔarnau-k-* < *\*ʔarnab-k-*, assim.; however, there are no traces of *-u-*. Likely related is C.: Gude *ùrìŋá* (< *\*ʔurin-*) ‘type of mouse or rat’.

<sup>65</sup> Acc. to CED #11, < *\*tu-nbar-*, with metathesis and assimilation.

<sup>66</sup> Cf. Sem. *\*bVrbVr-* ‘to hawl, roar’.

<sup>67</sup> Cf. also hubayrat- ‘hyène’ [Hyaena hyaena].

<sup>68</sup> Both quoted in EDE II 21, not in Naït-Zerrad 1998.

<sup>69</sup> Yaaku *barié* ‘jackal’ [*Canis aureus*] is a lw. from Masai.



## 7.5. \*čap- ‘k. of smaller carnivoran or rodent’

Sem. \**tap-an-* ‘rock hyrax’ [*Procavia capensis*]: Hbr. *šāpān* ‘rock badger, hyrax, dassie’ [*P. capensis*] || Mhr. *tōfān*, Jib. *tōfun* ‘rock hyrax’ [*P. capensis*].

(?) Chad. W.: Boghom *sop*<sup>70</sup> ‘rabbit’ [*Lepus sp.*].

Cush. \**čap-*: E.: LEC: Oromo *šuff-ee*<sup>71</sup> ‘k. of badger, ratel’ [*Mellivora capensis*] || S. \**čap-*: Bu-runge *čaa-p-u* ‘ground squirrel’ [*Spermophilus sp.*] | Alagwa *cap-ù* ‘civet’ [*Civettictis civetta*].

Omot. N. \**šif-*: Kafa *šif-ō* ‘faina (beech marten)’<sup>72</sup> | Dizi (Sheko) *šimf-a*<sup>73</sup> ‘mouse’.

► AADB 3908; SED II No. 240.

## 7.6. \*dab- ‘k. of large animal’

Sem. \**dabb-* (redupl.) ‘bear’ [*Ursus arctos*]: Akk. *dab-u* (*dabb-u*) || Hbr. *dōb* || Aram.: Syr. *debb-ā* || Arab. *dubb-* || Gz., Tgr., Amh. *dabb*, etc.

Egyp. (OK) *db* ‘hyppopotamus’ [*Hippopotamus amphibius*].

Chad. \**dab(b)-*: W. \**dabb-*: Hausa *dább-à* | Ngizim *dább-à* ‘wild animal’<sup>74</sup> | Boghom *túp*<sup>75</sup> ‘hyppopotamus’ [*Hippopotamus amphibius*] | Pero *ḍeḅ-à* ‘hyena’ [*Crocota crocuta*] | Ngamo *dab-ən* ‘lion’ [*Panthera leo*] || C. \**dVb-*: Gudu *dōv-ən* ‘lion’ [*P. leo*], *dív-ā* ‘leopard’ [*P. pardus*].

Cush. S. \**dab-*: Dahalo *dabi* ‘animal, game’.

Omot. N. \**dab-*: Gimirra *dodba* (redupl.) | Ganza *dwob* ‘lion’ [*Panthera leo*] || S.: Ari *de(e)b-i* | Hamar *dab-i*, *dab-i* ‘wild animal’.

▲ Cf. Proto-Kuliak \**dob-* ‘rhinoceros’.

► AADB 282. SED II No. 65.

7.7. \*sa/ik<sup>w</sup>- ‘k. of lesser animal’

Sem. \**šikVk-* (redupl.): Akk. (OB on) *šikk-û* ‘mongoose’ [*Herpestidae sp.*] || Amh. *šakokk-o* ‘rock hyrax’ [*Procavia capensis*].

Chad. C. \**sikw-*: Pižimbi *šekw-a* ‘squirrel’ [*Sciuridae sp.*] | Logone *swék-a*, *súwék-a* (met.) ‘ground squirrel’ [*Spermophilus sp.*].

Cush. E.: Dullay \**šak-t-* ‘chameleon’ [*Chamaeleonidae sp.*]: Gawwata *šaḥ-to* | Gollango *šaḥ-to*.

Omot. N. \**šak-*:<sup>76</sup> Wolaita *šak-aanča* ‘chameleon’ [*Chamaeleonidae sp.*]<sup>77</sup> | Kafa *šakšak-o* (redupl.) ‘pipistrello’ (Cerulli 1951: 496) || S. \**sik-*: Dime *sik-u* ‘bat’ [*Microchiroptera sp.*].

► AADB 2653. Cf. SED II No. 193.

## 7.8. \*žaʔib- ‘k. of scavenger’

Sem. \**diʔb-*: Akk. *zīb-u* ‘jackal’ [*Canis aureus*]; ‘vulture’ [*Gyps sp.*] || Hbr. *zəʔēb* || Aram.: Off. *dʔb*, *dyb* | Jud. *dēb-ā*, *zīb-ā*, *zēb-ā* | Syr. *dēb-ā* ‘wolf’ [*C. lupus*] | Mand. *zab-a*, *dib-a* || Arab. *dīʔb*, *dīb-* ‘loup; chacal’ [*C. lupus*, *C. aureus*] || Gz. *zəʔb* | Tna. *zəbʔ-i*<sup>78</sup> | Amh. *žəb* ‘hyena’ [*Hyaenidae sp.*] || Jib. *dīb* | Soq. *dīb* ‘wolf’ [*C. lupus*].<sup>79</sup>

Egyp. (Pyr.) *zʔb* ‘jackal’ [*C. aureus*].

<sup>70</sup> s- may continue \*č.

<sup>71</sup> If the theory that \*č > Oromo *f* is correct, perhaps \*č > Oromo *š* before a labial, cf. “s did not become *f* before a labial” (PEC 33).

<sup>72</sup> An erroneous determination: *martens* (including beech-marten [*Martes foina*]) do not inhabit Africa.

<sup>73</sup> With insertion of *-n-*.

<sup>74</sup> These forms are arabisms, acc. to some authors; however, this form with the meaning ‘wild animal’ is not attested in Class. Arab.

<sup>75</sup> Regular devoicing.

<sup>76</sup> Cf. also Kachama *šikaale* ‘wild cat’.

<sup>77</sup> Borrowed from or into Dullay.

<sup>78</sup> Tgr. *zəbʔi* likely from Tna.

<sup>79</sup> It is not clear whether the MSA forms are genuine or Arab. lws.

Chad. \**žab-*: W. \**žabžib-* (redupl.): Hausa *žibžib-tà* ‘vulture’ [*Gyps sp.*]<sup>80</sup> || E. \**žabiy-* ‘hyena’ [*Crocuta crocuta*]: Migama *žábíy-á* | Bidiya *žèbèy-gì*.

Omot. \**zab-*: N. \**zab-*: Basketo *zobba, zop* (-*p* < \**?b-*?) | Male *zobi* ‘lion’ [*Panthera leo*] | Gimirra *e-zub-ay* ‘cat’ || S. \**zab-* ‘lion’ [*P. leo*]: Ari *zob, žobba* | Hamar *zob-o* | Dime *zof*.<sup>81</sup>

► AADB 284; HSED #2660; SED II No. 72.

## 8. Birds (a selection)

### 8.1. \**bar-* ‘k. of bird (pigeon, partridge)’

Sem. \**bar-* and \**barbur-* (redupl.): Hbr. *barbūr* ‘a bird fattened to be eaten by king Solomon’ || Tna. *bar-eto, bar-äto, bar-ito* ‘turtledove’ [*Streptopelia sp.*], *bar-o* ‘k. of turtledove’ | Amh. *bar-et, bar-eto* ‘pigeon, dove’ [*Columba sp., Streptopelia sp.*].

Egyp. (old) *b<sup>3</sup>* (if < \**bVr*) ‘kind of bird, black stork, jabiru’ [*Ciconia sp.*].

Berb. \**barr-*: E. Wlm. *ta-bærr-ut* ‘caille d’Europe’ [*Coturnix sp.*] | Senhadja, Rif *a-barr-an* ‘perdrix mâle’ [*Perdix sp.*] | Qabyle *burur-u* (redupl.) ‘owl’ [*Strigiformes sp.*].

Chad. \**bar-* and \**barbir-* (redupl.): W. \**?V-(m)bVr-* and \**bVrbVr-*: Hausa *bì/ùrabì-à* ‘quail’ (Bargery 1934) [*Coturnix sp.*], *bàruu* ‘pigeons’ [*Columba sp.*] | Kirfi *mbìr-ù* ‘dove’ [*Streptopelia sp.*] | Jimbin *a-bur* ‘crested crane’ [*Balearica regulorum*] | Ngizim *bàr-í-* ‘domestic pigeon’ [*Columba livia*] || C. \**birbVr-*: Bura *biber-wi* ‘k. of bird’ || E. \**barr-* and \**birbir-*: W. Dangla *bââr-è* (m) ‘perdrix (Francolin)’ [*Perdix sp.*] | E. Dangla *bàrr-é* (f.) ‘perdrix’ [*Perdix sp.*] | Bidiya *birbir-íny* (f.) ‘caille’ [*Coturnix sp.*].

Cush. \**bVr-*: E.: HEC \**bur-*: Burji *búr-o* ‘partridge, dove’ [*Perdix sp., Streptopelia sp.*] || S. \**m-bar-* and \**bambar-* (redupl.): Iraqw *bambar-é* ‘hoopoe’ [*Upupa sp.*] and *biir-íſ* ‘pytilia bird’ [*Pytilia sp.*] | Dahalo *mbār-e* ‘egret’ [*Egretta/Ardea sp.*].

▲ Cf. Sum. *buru* [bird], wr. *buru<sub>4</sub><sup>mušen</sup>, buru<sub>6</sub><sup>mušen</sup>* ‘crow; a bird of prey or a vulture’ (PSD), *buru<sub>4</sub>, bur<sub>4</sub>* ‘crow’ (EDS #385), *buru<sub>4</sub><sup>mušen</sup>* ‘corbeau’ (LSF).

► AADB 2228; SED II No. 61; cf. HSED № 293, EDE II 431-2, Stolbova 2021 #274, 274a (comp. to Eth. Burji, Senhadja, Rif).

### 8.2. \**çiraſ-* ‘k. of bird’<sup>82</sup>

Sem. \**ʕVššūr-* (met.):<sup>83</sup> Akk. *iššūr-u* || Ugr. *ſſr* [ʕuššūru] ‘bird (gen.); domestic fowl, poultry’.<sup>84</sup>

Chad. \**çVr-*: W. \**çury-*: Hausa *çíry-ã* ‘parakeet’ | Polchi *cúr-a* ‘hawk’ || E. \**sar-* and \**sarsir-* (redupl.): Dangla *sàrsir-ò* ‘bird of prey’ | Bidiya *sár-a* ‘épervier’ [*Accipiter sp.*].

Cush. \**çirr-* and \**çiçir-* (redupl.) ‘red-billed oxpecker (*Buphagus erythrorhynchus*)’: C \**çarrur-*: Bilin *çarrúr-ā* ‘Madenhacker’ [*B. erythrorhynchus*] || E.: SA \**çarraſ-*: Saho *çarãſ, çarrãſ* ‘Madenhacker, buphaga erythrorhynchus’ [*B. erythrorhynchus*] | LEC \**çirr-*: Oromo *çi/err-ii* ‘k. of bird,

<sup>80</sup> Supported by a similar meaning shift in Akk.

<sup>81</sup> Cush. N.: Beja *diib* ‘wolf’ is probably an Arabism. HEC: Sidamo *doob-iččo*, Kambatta *zob-ečču*, Tambaro *zob-eččo* ‘lion’ are likely to be borrowed from S. Omot.

<sup>82</sup> Egyp. (Pyr.) *dry.t* ‘kite’ compared (if < \**çry*) in SED II No. 43 is not related due to lack of ʕ whose presence in this root is confirmed by both Sem., Saho and S. Cush. forms.

<sup>83</sup> Cf. also Arab. *ʕušfūr-* ‘any small bird’, Malt. (*gh*)*asfūr* (< \**ʕašfūr-*), Hbt. *ašferóot* ‘small bird’, Jib. *ſešférót*, Soq. *ſašferoh* ‘bird’ (perhaps a result of contamination between \**ʕVššūr-* and \**šVp(p)Vr-* see SED II No. 212).

<sup>84</sup> Tgr. *çarraſ, çarnaſ* (met.) ‘a bird, Buphaga erythrorhynchus’ [*Buphagus erythrorhynchus*] and Tna. *çarraſ* ‘a bird a little larger than the thrush’ continue the common form \**çirraſ-* (with dissim. of -rr-), most likely borrowed from Saho *çarrãſ* with the same highly specific meaning *Buphagus erythrorhynchus* which can hardly be reconstructed at the PAA or even PCu. level. Hence, attestation of the corresponding term in C. Cush. (Bilin), E. Cush. and even S. Cush. (Dahalo) can only reflect a series of borrowings, probably with Saho as a primary source, eventually related to S. Cush. (W. Rift and Asa) and, metathetically, Sem. (Akk. and Ugr. only) forms.

sucks blood from cow; like hornbill' [*B. erythrorhynchus*] | HEC \**çirr-*: Sidamo *çirr-e* 'red-billed oxpecker' [*B. erythrorhynchus*] || S. \**çiraʕ-*: Iraqw, Gorowa *çirʕ-i* | Alagwa *ciraʕ-a* | Burunge *çiraʕ-a* | Asa *širaʔ-a* 'bird' (generic term) | Dahalo *çičir-a* 'tick bird' [*B. erythrorhynchus*], *çuár-e* 'sp. bird (Roller?)'.<sup>85</sup>

(?) Omot. N.: Wolaita *çer-aa* 'k. of bird' (very likely a Cushitism).

► AADB 320; cf. HSED #1095 and SED II No. 43.

### 8.3. \**dVr-* 'k. of bird'

Sem. \**durar-* (redupl.) 'k. of bird': Hbr. *dərōr* 'a kind of bird (swallow, dove?)' || Deir Alla *dr̄r* 'subst. indicating bird: swallow or dove?' || Arab. *durr-at-* 'perroquet' (BK 1, 682), 'la peruche à collier couleur de rose' (Dozy 1927: I 428) [*Psittacinae sp.*] || Amh. *där(r)-ay* 'aquatic, web-footed bird which has black or white plumage' | Masqan *dərr-i* 'k. of bird'.<sup>86</sup>

Chad. \**dawr-* (or \**daHur-*): W. \**du/ur-* and \**dVrdVr-* (redupl.): Hausa *dúr-wǎ* 'lark-quail (Ortyxelos Meiffrenii)' [*Ortyxelos meiffrenii*] | Dera *dárir-ù* 'type of swift nightbird' | Seya *dər-i* 'crown-bird' [*Balearica regulorum*] | Buli *dádár* 'weaver bird' [*Proceidae sp.*] || C. \*(*n-*)*dawr-*: Mbara *n-dòr-ò* 'green pigeon' [*Treron sp.*] | Gude *dár-á* 'type of bird' || E. \**dVHur-*: Somrai *dùr-ē* 'hen' | Tumak *door* 'turtle-dove' [*Streptopelia sp.*], *dèr-t<sup>87</sup>* 'oiseau (nom générique)'.

Cush. \**diruh-*: N. \**?an-dirh-*: Beja *andiirh-o* 'fowl' || C. \**dir(h)w-*: Bilin *diruw-a* | Qwara, Dembea *dirh<sup>w</sup>-ā* | Aungi *dir-ī*, etc. 'cock, hen' || E.: SA \**do/irh-*: Saho *door(o)h-o*, *diirh-o* 'Huhn, Henne' | Afar *dorrah-e* 'hens' | LEC \**du(H)r-*: Somali *door-o* | Jiddu *duur-i* 'chicken, hen' || S. \**di(H)r-*: Dahalo *déer-e* 'woodpecker' [*Picidae sp.*].

Omot. N. \**dVr-*: Zergula *dər-i* 'chicken', Chara *deer-aa* 'rooster',<sup>88</sup> Mao (Hozo) *dorŋ-a* 'bird'.

▲ Cf. Sum. *dar* [bird], wr. *dar<sup>mušen</sup>*, *dara* 'bird, black francolin' (PSD), *dar<sup>mušen</sup>* 'black francolin' [*Francolinus francolinus*] (EDS #437), *dar<sup>mušen</sup>* 'francolin' (LSF). Blažek (2003) also quotes NS: Kunama *doora*, *dorha* 'chicken' (obviously a lw. < Cush.) and, less convincingly, Nubian: Kenuzi *darbád* 'hen', etc. (probably a look-alike).

► AADB 1665; SED II No. 71. Cf. HSED № 748, Blažek 2003: 235 (the Agaw forms in *g* are erroneously mixed with the forms in *d* quoted above).

### 8.4. \**g<sup>w</sup>ay-* and \**g<sup>w</sup>ag<sup>w</sup>ay-* (redupl.) 'k. of bird'

Sem. Eth. \**g<sup>w</sup>ag<sup>w</sup>-*: (a) Gz. *g<sup>w</sup>āg<sup>w</sup>ā* | Amh. *gag-ano* 'swan, ibis' [*Cygnus sp.*, *Threskiornithinae sp.*]; (b) Gz. *gug-ā*, *g<sup>w</sup>ag-ā* 'owl, nighthawk, raven' | Tgr. *g<sup>w</sup>agg<sup>w</sup>-a* | Amh. *g<sup>w</sup>agg-a*, *g<sup>w</sup>ag<sup>w</sup>-ət* 'owl' [*Strigiformes sp.*] | Gur.: Gyeto *g<sup>w</sup>ag<sup>w</sup>-a*, *gug-a*, etc. 'night bird'.

(?) Egypt. *dwy.t* (sarc) 'kind of bird' (if < \**gwy-t*).

Chad. (a) \**g<sup>w</sup>ay-*: C. \*(*?V-*)*giy-*: Higi *gýē*, Mandara *giy-e*, etc. | Gude *?i-ggi-ná* 'bird', *gyagy-a* 'hen' || E. \**guy-*: Somrai *gui* 'partridge' [*Perdix sp.*]; (b) \*(*n-*)*gag/k-* (redupl.) 'crow' [*Corvus sp.*]: W. \*(*n-*)*gag/k-*: Gwandara *ŋgaŋk-ā* | Angas *ngak* (-*k* < \**-g*) | Pa'a *gáàg-a* || C. \**n-gag/k-*: Buduma *n-gag-é* || E. \**gag-*: Bidiya *gaag-a* | Mubi *gak* (dissim.?) | Mokilko *gaag-u*.

Cush. E. \**gug/k-*: SA \**gugg-* (redupl.): Saho, Afar *gugg-a* 'hornbill' [*B. erythrorhynchus*] | LEC \**gug-*: Oromo *gug-e* 'dove' [*Streptopelia sp.*] | HEC: \**gug-Vn-*: Darasa *guug-en-žo*, pl. *guug-*

<sup>85</sup> Both words probably borrowed, since \**ç* is expected to be preserved in Dahalo.

<sup>86</sup> Cf. Gz. *dorho*, *doroho* 'chicken', Tgr. *derho* 'chicken', Tna. *dārho* 'pollo, gallina', Amh. *doro* 'chicken', Gog. Sod. *žärä* 'hen, chicken' borrowed from Cush. The non-etymological "extra" *-h* is equally unexplainable in both Eth. and Cush. Cf. also Syr. *dardā* 'vultur' [*Gyps sp.*]

<sup>87</sup> From \**dVHVr-*, acc. to Caprile 1975: 123.

<sup>88</sup> Both may be Cushitic lws.

*enna* ‘dove’ [*Streptopelia sp.*]<sup>89</sup> || Yaaku *goigui* ‘bee-eater’ [*Meropidae sp.*] || S. \**n-gV*k- (dissim.): Dahalo *ngōk-o* ‘whydah bird’ [*Vidua macroura*].

▲ Cf. Kuliak \**gwak-* ‘bird’. Cf. also Sum. *uga* [raven], wr. *uga<sup>mušen</sup>*, *uga* (PSD), *uga<sup>mušen</sup>* ‘crow, raven, bird of pray’ (EDS #2693). Note special affinity between some Chad. and Dahalo forms.

► AADB 324, 2098.

### 8.5. \**g<sup>w</sup>am-* ‘k. of bird’

Sem. \**g<sup>w</sup>am-* and \**g<sup>w</sup>amg<sup>w</sup>am-* (redupl.): Akk. (SB) *gamgamm-u* ‘a bird’ || Gz. *gumgum-ā* ‘pelican’ [*Pelecanus sp.*] | Tna. *g<sup>w</sup>ag<sup>w</sup>amm-a* ‘bird which chirps when in numbers’, *gum-a* ‘large vulture the neck of which is featherless’ [*Gyps africanus*] | Tgr. *gumgum-a* ‘esp. d’oiseau’.

Chad. \**gam* and \**gamgim-* (redupl.): W. \**gam-*: Polchi *gam* ‘eagle’ [*Aquila sp.*] | Sayanchi *gām* ‘hawk’ [*Accipiter sp.*] || C.: Zulgo *gúwùm*, Gisiga *gigim* ‘owl’; E. \**gumgi/um-* ‘owl’: W. Dangla, Migama *gúgùm-ò* | Bidiya *gúugum-a* | Mokilko *gùugím-à*, Ubi *guugùm*, Mawa *gugum* (cf. Stolbova 2011 #630).

Cush. N. \*(?*a-*)*gam-*: Beja *agam-a* ‘seagull’ [*Larus sp.*] || E.: SA \**gum-*: Saho *gúm-ā* ‘der Adler’ [*Aquila sp.*] | Afar *gum-a* ‘vultures; white-backed vultures’ [*Gyps sp.*].

Omot. N.: Gimirra *gyam* ‘chicken’.

► AADB 2628; cf. SED II No. 78.

### 8.6. \*(?*ǝa-*)*gwir-* ‘k. of bird’

Sem. \*(?*ǝa-*)*gawir-*: Akk. (OB) *igir-û* ‘heron’ [*Ardea sp.*] || Hbr. *ǝāgūr* ‘short-footed thrush or swift or wryneck (?)’ || D.-Alla *ǝr* ‘subst. indicating certain kind of bird’ || Arab *ǝawraḵ-*<sup>90</sup> ‘*Struthiocamelus mas*’ (Freytag 1833) [*Struthio camelus*].

Egyp. (late) *gr-y* ‘poultry’.

Chad. \**gawir-* and \**girw-* (met.): W. \**gir-*: Zar *gèr-í* | Seya *gèr-i* ‘hen’ || C. \**girw-* and \**gagwir-* (redupl.): Matakam *gógwàr* ‘cock’ | Munjuk *gagray* ‘guinea-fowl’ [*Numida meleagris*], *yugúr-iy* ‘hen’ | Mandara, Padukwo *ǝírrw-e* | Uzam *ǝírw-e* ‘ostrich’ [*Struthio camelus*] || E. \**gawr-*: Kera *gúgur-i* (redupl.) | Somrai *gàray* ‘hen’ and *gwár-a* ‘heron’ [*Ardea sp.*] | Dangla *gár-è* ‘duck’ [*Anas sp.*].

Cush. E. \**gawr-ay-*: SA: Saho *gàry-a*, *gàraay* | Afar *goroyy-a* | LEC: Somali *goray* | Boni *korii*, *koree* ‘ostrich’ [*Struthio camelus*] | Oromo *gogorr-ii* (redupl.) ‘guinea fowl’ [*Numida meleagris*]<sup>91</sup> || S.: Dahalo *ngára* ‘crested crane’ [*Balearica regulorum*].

▲ Cf. Sum. *igira* [heron], wr. *igira<sub>2</sub><sup>mušen</sup>* (PSD), *igira*, *igirû* ‘stork’ (EDS 1233).

► AADB 1931; cf. HSED #1053; SED II No. 29.

### 8.7a. \**ɣura?* - ‘crow’

Sem. \**ɣa(r)ɣur-* (redupl.): Akk. (OB) *ḥaḥḥur-u* ‘a bird of the raven/crow family’.

Berb. S. \**ɣru-t*: Ayr, E.Wlm. *a-ɣrut* ‘corbeau’ [*Corvus corax*] (otherwise <\**k<sup>w</sup>ar-* ‘crow, raven’ 8.13).

Chad. \**ɣar-* ‘crow’:<sup>92</sup> W. \**ɣaHar-*: Sura *gòr-òò* ‘Krähe’ | Zul *ɣàr-átí* ‘crow’ || C.: Glavda *ɣáayàhór-a* (redupl.) ‘crow’ [*Corvus sp.*] | Sukun *ɣram-ak* ‘pied crow’ [*Corvus albus*] || E. \*?*ɣa-gr-* (met.): Ndam *ɣagr-a* ‘crow’ [*Corvus sp.*].

Cush. \**ɣura?* - ‘crow’ [*Corvus sp.*]: C.: Qemant *ɣoray*; S. \**ɣura?* - Iraqw *ɣwaʔari* (met.) | Bu-runge *ɣwarar-iya* (redupl.) | Alagwa *ɣohoraʔi* (redupl.) | Ma’a *i-h/ɣureʔ-a* (met.)

► AADB 2632; SED II No. 89; CED #300.

<sup>89</sup> Cf. Dullay \**ǝg-it-*: Warazi, Gobeze, Harso *aag-iče*, Tsamay *ʔaag’-* ‘bird (generic)’.

<sup>90</sup> Non-etymological -*ḵ* is probably due to contamination with Arab. *ɣirniḵ-* ‘grue’ (see in 8.7.3. \**ɣurayḵ-an-*).

<sup>91</sup> Omot. N.: Koyra *googgóra* ‘partridge, quail’ < Oromo.

<sup>92</sup> Our \**ɣ* corresponds to *h* in Stolbova’s notation.

8.7b. \**ɣurayb-* ‘crow, raven’

Sem. \*(*ʔa-*)*ɣārib-* and \**ɣurab-*: Akk. *ārib-u* (*ērib-u*, *hēreb-u*) ‘crow, raven’ [*Corvus sp.*] || Ebla *ḥa-ri-bu/ī*, [*h*]a-ri-bū-um, g[a-r]í-bū /*ɣārib-um*/ ‘corvo’ [*Corvus sp.*] || Hbr. *ṣōrēb* ‘raven’ [*C. corax*] || Aram.: Syr. *ṣūrāb-ā* ‘corvus’ [*C. corax*] | Mand. *ṣurb-a* ‘crow’ [*Corvus sp.*] | Arab. *ɣurāb-* ‘corbeau (tout noir); ‘corneille’ [*Corvus sp.*] || Mhr. *ɣə-ɣəráyɔb* ‘raven’ [*C. corax*] | Jib. *ʔa-ɣərəb* ‘raven, crow’ [*Corvus sp.*] | Soq. *ʔá-ṣreb* ‘crow’ [*C. corax*], ‘Egyptian vulture’ [*Neophron percnopterus*].

(?) Egyp. (Lit MK) *ṣbw* ‘kind of bird?’ (if < \**ṣbw* met. < \**ṣrb*).

Chad. E. \**ɣurb-*: Jegu *gurb-ak* ‘crow’ [*Corvus sp.*]<sup>93</sup>

Cush. E. \**ɣurrub-* (redupl.) ‘crow’ [*Corvus sp.*]: Dullay \**ḳurrub-*: Gawwata *ḳurruup-akko* | Harso *ḳurruup-ičče* | Galila *ḳurruup-itte*.<sup>94</sup>

► AADB 2632. Cf. SED II No. 89; HSED N<sup>o</sup>1015.

8.7c. \**ɣurayk-an-* ‘k. of bird (crane, raven)’

Sem. \**ɣurnayk-* (met.): Akk. (SB) *urnīk-u*, *ḥurnīk-u* ‘Kranich’ || Arab. *ɣirniḳ-*, *ɣurnūḳ-*, *ɣirnaḳ-*, *ɣurnayḳ-* ‘grue’.

Chad. \**n-gawrak-* (met.)<sup>95</sup> ‘crested crane (*Balearica regulorum*): W. \*(*n-*)*gaw(m)rak-*: Hausa *gaurak-a*, *gamrāk-a*, *garmāk-a* ‘(crested) crane’ [*Balearica regulorum*] | Karekare *ngùràak-àa* ‘crowned crane’ [*B. regulorum*]<sup>96</sup> || E. \**n-garak-*: E. Dangla *ḡārk-o* | Migama, Bidiya *ḡarāk* ‘(crested) crane’ [*B. regulorum*].

Cush. E. \**ḡara(y)ḳ-*: HEC \**ḡaraḳen-t-*: Darasa *ḡaraḳen-sa* (pl.) | Sidamo *ḡaraḳess-a* (assim. of \**-n-*) ‘raven’ [*Corvus corax*] || Yaaku *ḡaryaḳ-a* ‘crow’ [*Corvus sp.*].

(?) Omot. S.: Ari *ḡuraak* (met.?) ‘crow’<sup>97</sup> [*Corvus sp.*].

► AADB 2633; SED II N<sup>o</sup> 91.

8.8. \**k<sup>w</sup>ar-* and \**k<sup>w</sup>arkay-* ‘k. of bird’

Sem. \**k<sup>w</sup>arkay-* (redupl.): Akk. *kurkû* ‘goose’ [*Anser sp.*] || Aram.: Syr. *kurḳay-ā* ‘grus’ [*Grus sp.*], Jud. *kurḳay-ā*, Mand. *kurki-a* ‘crane’ [*Grus sp.*] || Arab. *kurkiyy-* ‘grue’ [*Grus sp.*] || Gz. *k<sup>w</sup>arāki*, *korki* ‘crane’ [*Grus sp.*]<sup>98</sup>

(?) Egyp. (Pyr.) *tḥ* (if < \**kVr-*) ‘junger Vogel (Schwalben, Geier); junges Tier’.

Berb. \**-kVr-t*<sup>99</sup>: Ayr *a-krə-t* | E. Wlm. *a-kər-t* ‘nestling, young of a bird | Ahaggar *e-ker-t* ‘young of ostrich’.

Chad. \**k<sup>w</sup>ar-*, \**k<sup>w</sup>arkay* and \**k<sup>w</sup>ark<sup>w</sup>ir-* (redupl.): W. \**k<sup>w</sup>ar-*: Tangale *k<sup>w</sup>ar-tɛ*, *k<sup>w</sup>atr-ɛ* (met.) ‘guinea-fowl’ [*Numida meleagris*] | Polchi *koṛ-ə* | Buli *kúr* | Tala *kwor*, *kor* ‘hen’ || C. \**kur-*, \**k<sup>w</sup>arkay-* and \**k<sup>w</sup>Vrk<sup>w</sup>Vr-* (redupl.): Mbara *kūrākáy* | Gude *kūrākú-tà* | Lame *kòrók-ú* ‘dove’ [*Streptopelia sp.*] | Mofu *kwerekwer-e* | Munjuk *kor-o* ‘duck’ [*Anas sp.*] | Sukur *ta-kur* ‘hen’ and ‘dove’

<sup>93</sup> Acc. to CED #300, not an Arabism as presumed in SED II No. 89, but a genuine Chadic word.

<sup>94</sup> Cf. also HEC: Burji *gurrub-a* ‘crow’ which may be a Dullay lw.: in Burji \**ɣ* > *h* is rather expected.

<sup>95</sup> Acc. to CED: 41, Sem. \**ɣ* (conveyed *ibid.* by \**ḡ*) corresponds to what Stolbova reconstructs as P. Chad. \**fi* (a symbol accepted in Chadic but not in AA studies to convey \**ɣ*) yielding \**g-* in W. and E. Chad. As for the velar in the final position, all Chadic forms except Hausa may reflect \**-k* corresponding to \**-ḳ* in the other AA branches; Hausa, however, points to \**-k*, not \**-ḳ*. One wonders if the Hausa term could have been borrowed from another W. Chad. language and not *vice versa*; another possible explanation is the loss of “emphatization” in the position of the fourth radical.

<sup>96</sup> Jimi *gaurak-a*, Ngizim *ḡəwāarak*, *vāarak* ‘crown bird’ [*B. regulorum*] are probably loans from Hausa.

<sup>97</sup> Reflexes of the extremely infrequent voiced uvular *ɣ* in Omotic have not been reliably established.

<sup>98</sup> Claimed by some authors to be a series of borrowings: Aram. < Akk. < Sum.; Gz. < Arab. Neither is convincing in view of the AA parallels.

<sup>99</sup> Cf. Ayr *ta-kārāw* ‘k. of bird’.

[*Streptopelia sp.*] || E. *\*(ʔa-)ku(r)kir-* (redupl.): Kera *akórkórò* | Migama *kókkór* ‘duck’ [*Anas sp.*], *kúkkirá* ‘cock’ | Dangla *kókíra* | Jegu *kókóré* | Mubi *kòkóréy* ‘hen’.

Cush. *\*kar-*: N. *\*ka(r)karr-*: Beja *kaakarr-et* ‘hen’ || S. *\*karaʔ-*: Qwadza *kalaʔ-eto* ‘stork’ [*Ciconia sp.*].

Omot. N. *\*kur-* and *kurki-t-* (redupl.) ‘partridge, quail’ [*Coturnix sp.*, *Perdix sp.*]: Gamo *kúr-aččo* | Dorze *kor-ačo* | Zergula *kerk-ečče* | Male *karek-ačo* || S. *\*kurki-t-*: Hamar *korikiš-a*.

▲ Cf. Sum. *kurgi* [goose], wr. *kur-gimūšen*, *kurku<sub>2</sub>*, etc. (PSD), *kur-gimūšen* ‘crane’ (EDS #1478).

► AADB 377; cf. SED II No. 117; HSED 1505; Stolbova 2011 #277.

### 8.9. *\*k<sup>w</sup>ay-* and *\*k<sup>w</sup>ak<sup>w</sup>ay-* ‘k. of crow or bird of prey’

Sem. (a) *\*(ʔa-)kaw-*: Akk. (SB; LL) *akkû* || Aram. Jud. *kaw-tā* ‘owl’ [*Strigiformes sp.*] || Tgr. *kuw-a* ‘raven’ [*C. corax*]; (b) *\*k<sup>w</sup>ak<sup>w</sup>ay-* (redupl.): Gz. *k<sup>w</sup>āk<sup>w</sup>-ā* | Tgr. *kākay* ‘corbeau à capuchon blanc’ [*Corvus albicollis*] | Tna. *k<sup>w</sup>ak<sup>w</sup>-ə* ‘crow, raven’ [*Corvus sp.*].

Chad. *\*kawiy-*: W.: Bolewa *kāk-i* (redupl.) ‘crow’ [*Corvus sp.*] || C. *\*kVwiy-*: Mandara *kuy-è* ‘kite’ | Malgwa *kúúy-e* ‘falcon’ [*Falco sp.*] | Zulgo *kéwíy-è* ‘eagle’ [*Aquila sp.*] | Zime *kìw* ‘falcon, kite, eagle’ || E.: Mokilko *kāw-á* ‘white dove’ [*Columba livia domestica*]<sup>100</sup>.

Cush. *\*k<sup>w</sup>aw<sup>w</sup>ay-* (redupl.): N. *\*kwikay-*: Beja *kwiikway*, *kiik(?)ay* ‘crow; raven’ [*Corvus sp.*]<sup>101</sup> || E.: SA *\*kakaw-*: Saho, Afar *kaakoo* ‘raven’ [*C. corax*].

Omot. N. *\*ʔan-ka(w)-* ‘vulture’ [*Gyps sp.*]: Basketo *ánká* | Malo *anjó*.

► AADB 2641. Cf. SED II No. 123, Blažek 2003: 262.

### 8.10. *\*k<sup>w</sup>aʔay-* and *\*k<sup>w</sup>ak<sup>w</sup>ay-* ‘k. of gallinaceous bird’<sup>102</sup>

Sem. *\*kakay-* (redupl.): Syr. *kākāy-ətā* (*kokantō?*) ‘gallina’.

Berb. *\*-ka(H)y-aw* ‘cock’: Ghat *i-kai* | Ahaggar *e-kahi*, pl. *i-keh-an* | Ayr, E. Wlm. *ta-kəyy-a* | Semlal *a-kiy-aw*, pl. *i-kiyaw-ən* ‘poussin’.

Chad. (a) *\*k<sup>w</sup>ay-*: W. *\*k<sup>w</sup>ay-*: Gwandara *kwee* | Sura *kwéé* | Angas *kì* ‘hen’ | Guruntum *k<sup>w</sup>āy-a* ‘cock’ || C. *\*kway-*: Buduma *kúy-ē* ‘partridge’ [*Perdix sp.*]; (b) *\*k<sup>w</sup>ayk<sup>w</sup>-* (redupl.): W. *\*k<sup>w</sup>ayk<sup>w</sup>-*: Dera *kíyók* ‘goose’ | Fyer *kúkw-è* ‘hen’ || C. *\*kwak-*: Gulfei *kwaku* ‘hen’.<sup>103</sup>

Cush. *\*kaway-* and *\*kawkaw-* (redupl.): N. *\*kaw-*: Beja *kaw* ‘Perdrix Erkelii’ [*Pternistis erckelii*] || C. *\*kaway-*: Bilin *koy-a* | Qemant *kawiy-ā* ‘Frankolinus Erkelii’ [*Pternistis erckelii*] || E. *\*kawkaw-*: Yaaku *kəkə* ‘fowl’.

Omot. *\*kaway-* and *\*kaykay-* (redupl.): N. *\*(ʔa-)kayw-*: Male *kai* ‘bird’ | Yamma *akō* ‘chicken’ | Mao *wake* (met.) ‘fowl, chick’ | Chara *aka* | Kafa *akkō* | Mocha *akke* || S. *\*kaykay-*: Dime *kéekí* ‘partridge, quail’ [*Coturnix sp.*, *Perdix sp.*].

► AADB 325. Cf. SED II No. 123; Blažek 2003: 261.

### 8.11. *\*k<sup>w</sup>a(?)r-* ‘crow, raven’

Sem.: Eth. *\*k<sup>w</sup>ar-* and *\*k<sup>w</sup>ak<sup>w</sup>ar-* (redupl.) ‘crow, raven’ [*Corvus sp.*]: Gz. *kāker* ‘crow’ | Amh. *k<sup>w</sup>əra* (*kura*) ‘crow, raven’ | Selti *ḳure* | Wolaita *ḳuri* | Gur. *\*k<sup>w</sup>əra* ‘crow’.

Berb. S. *\*-ḳru-t-*: Ayr, E. Wlm. *a-ḳru-t* ‘corbeau’.

Chad. *k<sup>w</sup>ar-* and *kukar-* (redupl.): C.: Podoko *kākár-a* | Cuvok *k<sup>w</sup>ūr-əw* ‘crow’ || E.: E. Dangla *kḍr-é* ‘esp. de corbeau noir’ | Bidiya *kùraag-à* (met.; -g <\*k by dissim.?) ‘black crow’ [*Corvus sp.*] (cf. also *\*k<sup>w</sup>Vr-* ‘hawk’, *\*k<sup>w</sup>Vr(V)r-* ‘owl’ (Stolbova 2011 #406, 406b).

<sup>100</sup> Cf. also Fali-Jilbu, Gulfei *yukey* ‘hen’.

<sup>101</sup> Cf. Beja *éke* ‘Geier’ [*Gyps sp.*], *eeki* ‘white vulture, hawk’ [*Gyps sp.*, *Accipiter sp.*].

<sup>102</sup> It is hard to determine whether the present root is homonymous to 8.9. *\*k<sup>w</sup>ay-* and *\*k<sup>w</sup>ak<sup>w</sup>ay-* ‘k. of crow or bird of prey’ or if both eventually originate from a common protoform.

<sup>103</sup> Cf. also Chad. *\*kway(kway)-*: W.: Tangale *kwíyū* ‘dove’, Seya *kwok-kwok* ‘weaver-bird’; also Kwang *kóyō* ‘bird’.

Cush. C. \**ḳur-* and \**ḳwaḳur-* (redupl.): Bilin *kūāqūr-ā* ‘der Rabe’<sup>104</sup> | Khamta *qur-ā-* | Qwara *qur-ā* | Qemant *ḥor-ay* ‘raven’ [*Corvus corax*] || Aungi *kur-a* ‘crow’ [*Corvus sp.*] || E.: LEC \**ḳurr-*: Oromo *ḳurr-uu* | Bayso *ḳur-a* | HEC \**ḳurʔ(-an)-*: Hadiya *ḳoraan-ta* | Alaba *ḳur-a* | Burji *ḳorʔan-čō*, etc. ‘crow’ [*Corvus sp.*].

Omot. \**ḳur-* and \**ḳuraak-* (redupl.): N.: Wolaita *ḳuur-uwa* | Gamu, Zaysse *ḳur-o* | Dače *ḳor-aa-še* | Bworo *a-ḳoḳo-a* | Kafa *ku-re-ččo* | Yemsa *kur-a* ‘crow’<sup>105</sup> || S.: Ari *ḳuraak* ‘crow’ [*Corvus sp.*].

► AADB 1668; SED II No. 134.

## 9. Reptiles (selection)

### 9.1. \**ʔačhayl-* ‘(big) reptile’

Sem. \**ʔaḥhaly-* (met.) ‘a mythical reptile, dragon’: Syr. *ʔātaly-ā* ‘draco; stella, quae solem tegens eclipsim efficit’ | Mand. *tali-a* ‘fictive dragon causing eclipse’ || Tgr. *ʔashal-āt* ‘dragon’ | Tna. *ʔasāl-āt, ʔashal-āt* ‘animale favoloso, di smisurata grandezza e della specie del coccodrilla’.

Berb. \**Haššayl-*: Ghat *ašil* | Ahaggar *āššel* ‘serpent’ | Ayr *aššel* | E. Wlm. *aššol* ‘coulevre, gros serpent’ | Mzab *t-iššel-t* ‘vipère’.

(?) Cush. S.: Dahalo *táʔala* (met.) ‘puff-adder’ [*Bitis arietans*].<sup>106</sup>

Omot. \**ʔ/haylaš-* (met.) ‘crocodile’ [*Crocodylus sp.*]: N. \**ʔaylaš-*: Wolane *aylaaš-uwa* | Zala *aylaš-uwa* | Dawro *allaš-o* || S. \**haylaš-*: Ari *hayleš-a*.

► AADB 2606. Cf. SED II No. 20.

### 9.2. \**bač-* ‘snake, reptile’

Sem. \**baṭ-am/n-*: Akk. *bašm-u* ‘a horned serpent’ [*Cerastes cerastes*]; the constellation Hydra || Ebla *ba-ša-nu-um* ‘k. of snake’ || Ugr. *bṭn, bṭn-t* ‘serpent, dragon’ || Arab. *baṭan-* ‘coulevre, serpent’, *buṭn-* ‘genre de reptile’.

(?) Berb. \**buč-*: Shenwa *buš* ‘ramper (serpent, escargots, insects)’.

Cush. E.: LEC \**(ʔa-)bVč-*: Somali *abes-o* ‘kind of serpent’ | Oromo *bof-a* ‘snake’.

Omot. N. \**biʔas-* (met. < \**ʔi-bas-?*): Dizi (Sheko) *bias-u* ‘crocodile’ [*Crocodylus sp.*].

► AADB 2625, SED II No. 63.

### 9.3. \**bat-* ‘k. of snake or worm’

Sem. \**batt-ir-* and \**buttay-m-*: Akk. (SB) *mu-battir-u* ‘ein Wurm (?)’ || Gur.: Muher *buttayam-at* | Eža Chaha *butyam-at* | Ennemor. *butayäm-ata* | Gyeto *butyäm-ata* ‘vipère’.

(?) Egyp. *btn.w*, snake determinative.

Chad. \**bat-ar-*: C. \**bat-ar-*: Gudur *batar-a* and *botor-o* ‘vipère cornue’ [*Cerastes cerastes*] | Gisiga *botor-o* ‘Art Brillenschlange’ [*Elapidae sp.*] | Daba *m-burut-u* (met.) ‘ver de Guinée, draconculose’ || E. \**bVt-Vr-*: W. Dangla *bòttòr* ‘petit vers parasite du mil (sur épis)’ | Mokilko *ḥétér-é* ‘Würmer (Eingeweide)’.

Cush. E.: LEC \**būt-* ‘puff-adder [*Bitis arietans*]: Oromo *būt-ī* | Dasenech *būt-i?* | Konso *pūt-ota* | Arbore *būt-é*.<sup>107</sup>

<sup>104</sup> *kūā* is translated as ‘schreien, gäkern’.

<sup>105</sup> Acc. to Bender’s phonological correspondences, *k-* in Kafa and Yemsa can continue AA \**ḳ-*.

<sup>106</sup> *t̥* in Dahalo is usually considered to continue only AA \**t* – just as *d̥* is considered to continue AA \**d*. There are a couple of cases, however, of Dahalo *d̥* likely corresponding to Sem. *d̥* < AA \**ʒ* (cf. fns. 19 and 109). Interestingly, G. Takács, the foremost authority on AA consonantal reflexes in S. Cush., leaves empty cells for Dahalo reflexes of AA \**č* and \**ʒ* in his table of correspondences between PAA and S. Cush. (Takács 2011: 116).

<sup>107</sup> HEC: Burji *būt-ē* ‘puff-adder [*Bitis arietans*] and Golango *pūt-e* ‘Speischlange (Naia nigricollis)’ [*Naja nigricollis*] are loanwords from Oromo.

Omot. S. \**bVt*:- Dime *bɛt-á* ‘lizard’.

► AADB 2626; cf. SED II No. 63.

9.4. \**dawi*?- ‘k. of snake or worm’

Sem.: \**dawd*- (redupl.): Arab *dūd*-, *duwwād*- ‘ver, vermine’.

Egyp. *ddy* ‘Schlangename’.

(?) Berb.: \**diwdiw* (redupl.): Izayan *a-diuidiu* ‘k. of worm’.<sup>108</sup>

Chad. \**daʔiw*- and \**dawday*- (redupl.): W. \**daʔVw*- and \**dawday*:- Hausa *ɗa* ‘snake (the lightish-colored snake called dako)’ (<\**daʔ*-), *daudai* ‘k. of snake’ | Angas *dū* ‘lizard’ | Bolewa *did* ‘snake’ || C. \**dVdVw/y*:- Mofu *dédew* | Mada *dèdyè-kw* ‘a very poisonous snake’.

(?) Cush. N. \**dawɬ*:- Beja *doʔ-o* ‘maggot, worm, caterpillar’.

Omot. \**dawi*?- ‘python’ [*Pythonidae* sp.]: N. \**dawi*?-: Malo *dawe* | Oyda *dawwe* | Basketo *dawi* | Chara *doʔ-a* | Ginirra (Bencho) *dyah* | Dizi (Sheko) *däw-á* || S. \**da(H)w*:- Dime *dááú*.

► AADB 253, 368.

9.5. \**har*- ‘crocodile’

Chad. \**har-um*- ‘crocodile’ [*Crocodylus* sp.]: W. \**haram*:- Diri *hur-in* | Sha *hàr-àm* ‘some water monster (crocodile, hippo)’ | Bokkos *haràm* | Daffo-Butura *hàrà̀m* | Kulere *hàrà̀h*, etc. || C. \**hurum*:- Gisiga *hurom* | Mbara *húrúm* | Masa *hūrùm-nā*, etc. || E. \*(H)*urm*:- Lele *ürm-ō* | Kabalai *ɔrrm-ə*.

Cush. E.: Dullay: Gollango *háar-o* ‘crocodile’ [*Crocodylus* sp.] || S. \**harar*- (redupl.) ‘mythological giant snake’: Iraqw *hárár-îo* | Gorowa *haraari-yoda*.<sup>109</sup>

► AADB 4072; CED #332 (comp. to Gollango and Ongota).<sup>110</sup>

9.6. \**has*- ‘k. of big snake or crocodile’

Sem. \**na-haş*: Ugr. *nħš* ‘serpent, snake’ || Hbr. *nāḥāš* ‘snake’ (in one context interpreted as *Crocodylus vulgaris* [*Crocodylus vulgaris*]).<sup>111</sup>

(?) Egyp. (OK) *msh* (<\**mV-has*- with met.?) ‘crocodile’ [C. *vulgaris*].

Chad. W. \**ni-(H)wis*:- Sayanchi *nyíç-i* ‘snake’ | Bokkos *nyes* | Daffo-Butura *nis* ‘python’ [*Pythonidae* sp.].

Cush. E.: LEC \**ya-has*- ‘crocodile’ [*Crocodylus* sp.]: Somali *yeḥaas* | Rendille *yaḥ(a)s-í* | Boni *ṣaháas* (ṣ- <\**y*?) | HEC \**hamas*- (<\**mV-has* with met.) ‘snake’: Sidamo *hamas-o* | Hadiya *hamas-iččo* | Burji *hamas-i*.

▲ The multiple prefixes and metatheses may be due to tabooing.

► AADB 279, 336. Cf. Stolbova 2005 #502 and SED II: p. 211.

9.7. \**kVr*- ‘k. of big snake or crocodile’

(?) Egyp. (Pyr.) *ḫkr.w* ‘kind of snake’ (a variant root?).

Chad. \**kVr*:-<sup>112</sup> W. \**kVr*:- Bolewa *kúre-di* ‘snake’ | Dera *kôr-ì* ‘black cobra’ [*Naja melanoleuca*] | Tangale *kuro-t* | Pero *kùre-t* | Guruntum *kàrà-udi* ‘viper’ | Kulere *kígyér* (redupl.) ‘snake’ ||

<sup>108</sup> Cf. also *ti-dda*, *ti-ddi-t* ‘sangsue’ strangely quoted in Naït-Zerrad 1999: 274 without naming the language.

<sup>109</sup> Acc. to Kiessling, Mous, 2003: 132, “probably of Datoga origin”, which is less likely in view of the Chad. cognates.

<sup>110</sup> Ongota *haar-o* ‘crocodile’ compared *ibid.*, if recorded correctly, does not fit here because of *h*-.

<sup>111</sup> Cf. Arab. *ḥanaš*- ‘reptile, snake’ with irregular -š instead of -s and Akk. *šahan*, a snake-god acc. to some authors.

<sup>112</sup> In all Chadic languages represented here, *k* may continue both AA \**k* and \**ḳ*.



C. \**ku/ir-*: Hwona *kir-anja* | Gude *kir-ma* | Laamang *kár-ámá* | Mandara *kírr-we* | Glavda *kír-əwà* | Musgu *kúr-um* ‘crocodile’ [*Crocodylus* sp.] | Banana *kur-iyà* ‘snake’.<sup>113</sup>

Omot. S. \**kar-* ‘python’ [*Pythonidae* sp.]: Ari *qari* | Hamar *qáári*.

► AADB 266.

9.8. \**sib-* ‘k. of snake or worm’

Sem. \**šib-* and \**šibšib-* (redupl.): Akk. (MB) *šibb-u* ‘a snake’ || Gur.: Gyeto *sīb-a* | Chaha, Eža *šib-a* | Ennemor *šib-a* ‘k. of worm’ || Harsusi *šebšeb-ēt* ‘small red worm, centipede’ | Mhr. *šabšib*, Jib. *šabšéb* ‘red water worm’ [*Glycera* sp.].

Chad. E.: Lele *súb-ó* ‘worm’.

Cush. \**sib-*: C. \**sib-ar-* (with fossilized *-r* suffix) ‘snake’: Bilin *sabar-ā* | Khamir *sibr-ā* || E.: LEC: \**si(m)b-*: Oromo *siib-a* ‘worm’ | Somali *sumbay* ‘tapeworm’ [*Eucestoda* sp.].

Omot. N. \**šibb-* ‘crocodile’ [*Crocodylus* sp.] (with a semantic shift): Koyra *šəpp-o* | Ganjule *šəb-o* | Zaisse *šəbo*, *šəpp-o* | Zergula *šəp-o*.

► AADB 256; cf. SED II No. 200, HSED № 2227.

9.9. \**sVf-* ‘k. of snake’

Sem. \**šVp-*: Hbr. *šəpīp-ōn* ‘horned viper’ [*Cerastes cerastes*] || Syr. *šappāp-ā* ‘serpens’ || Arab. *siff-*, *suff-* ‘serpent; espèce de serpent tacheté de blanc et de noir’ || Tgr. *saf* ‘millepede’ [*Diplopoda* sp.] | Tna. *wäsf-at* ‘intestinal worm, hook-worm’ [*Ancylostoma* sp.] (K 1748).

Egyp. (RT) *sfy* ‘Name einer Schlange’.

Berb.: Zenaga *te-sfuf-əh* ‘a snake’.

Chad. E.: Kwang *sép-í* ‘snake’.

Omot. N. \**šip-*: Kafa *šef-ittoo* ‘sorta di lucertola’ | Bworo *šippir-a*, *šipr-a*<sup>114</sup> ‘worm’.

► AADB 254; SED II No. 207.

## 10. Water reptiles and amphibia (selection)

10.1. \**k<sup>w</sup>aɬ-* and \**k<sup>w</sup>aɬk<sup>w</sup>aɬ-* (redupl.) ‘frog’

Sem. \**k<sup>w</sup>aɬk<sup>w</sup>aɬ-* (redupl.) ‘frog’: Eža Msq. *k<sup>w</sup>äč-ä* | Cha. Enm. Gyt. *k<sup>w</sup>änč-ä* | End. *kōnč-ä* (with a secondary *-n*) || Mhr. *kāk-āt* | Hbt. *kakʔó-ot* | E. Jib. *kḏákḏ-át*.<sup>115</sup>

Chad. \**kawkaw* (redupl.): W.: Ngizim *kóokó* ‘small frog’ || C.: Buduma *kōko* ‘frog’.

Cush. E. \**kawɬ-* ‘frog’: Dullay \**koɬ-*: Tsamay *mu-koɬ-te* || Yaaku *koḵ-ontε* (redupl.).

Omot. N. \**ʔuḵ-* (met.?) and \**k<sup>w</sup>aḵ-* (redupl.) ‘frog’: Male *ʔuḵ-ane* | Mao (Hozo) *kwaḵ-a* | (Sezo) *kwaḵ-i*.

► AADB 3963, SED II No. 128.

10.2. \**čamb-* ‘some water creature’

Sem. \**šabb-*: Hbr. *šāb* ‘thorn-tailed lizard’ [*Uromastyx* sp.] || Syr. *šabb-ā* ‘garum, lacerta caudiverbera’ [*Lacerta caudiverbera*] <sup>116</sup> || Arab. *šabb-* ‘lézard, part. lézard d’Afrique’ <sup>117</sup> || Mhr. *žabb-it* | Hrs. *žebb-ēt* | Jib. *žəb* (pl. *žēt*) ‘(female) monitor lizard’ [*Varanus* sp.].

Berb. S. \**zVbb-* ‘k. of small insect living in water’: Ayr, E. Wlm. *ta-zəbb-e*.

Chad. W. \**čumb-*: Hausa *čumb-ē* ‘frog’.

<sup>113</sup> Cf. also Munjuk *kiyir* ‘ground worm’ and E.: Sokoro *kōre* ‘guinea worm’.

<sup>114</sup> With fossilized suffix *-r*.

<sup>115</sup> Cf. Jib. *šakšákət* ‘frog’, with metathesis.

<sup>116</sup> Aram. Jud. *šabbā*, *šābā* ‘thorn-tailed lizard’ [*Uromastyx* sp.], likely < Hbr. since \**š* is expected to yield *š*.

<sup>117</sup> Cf. also Arab. *ḏiʔb-* ‘animal living in water’.

Cush. S. \**ĉamb-Vʕ*:- Iraqw *ĉamb-ebé* ‘small insect drifting on top of water, tadpole’ | Alagwa *ĉemb-eʕu* | Burunge *ĉamb-eʕu* ‘frog’.

Omot. \**ĉV(m)b*:- Gimirra (Bencho) *ĉobm* ‘adder’ [*Vipera berus*]<sup>118</sup> | Yemsa *ĉop-a, šöp-ā* ‘fish’ | Kafa *ṭibb-ō* ‘sorta di verme che vive nell’acqua’.

► AADB 334; cf. SED II No. 221; Blažek 2008 #173; HSED #581; Dolgopolskiy 2008 #470.

### 10.3. \**dindal*- ‘small waer creature’

Sem. \**da(n)dil*:- Akk. (SB) *dālil*- ‘a small animal, probably a frog’ || Hbr. pB *nādāl* (met.) ‘polype, centipede’ || Aram.: Jud. *naddal*- id. (met.) | Syr. *dandāl-ā* ‘scolopendra vel millepeda’ || Jib. *dolol-ét* ‘kind of slow-moving snake’.<sup>119</sup>

Chad. E. \**didaln*- (met.): Dangla *didaln-yà* ‘limace’.

Cush. E.: HEC \**dindal*:- Darasa *daddal-ʕe* | Kambatta *diddil-ičču* | Sidamo *dandall-e* ‘lizard’.

► AADB 3916. Cf. SED II No. 68.

### 10.4. \**kʷar*- ‘frog’

Sem. \**kʷa/ir*- and \**kʷa(r)kʷir*- (redupl.) ‘frog’ (with different affixes): Syr. *yaḵrūr*- and *ʕaḵrūk*- | Jud. *ʕaḵrok-tā* || Arab. *ḵirr*-, *ḵurr*-, *ḵarr-at*- || Gz. *ḵāḵer*, *ḵʷarnanaʕāt* | Tgr. *ḵorəʕ*, *ʕanḵorəʕ* | Tna. *ḵʷərʕo*, *ənḵoroʕ*<sup>120</sup> | Amh. *ḵarnan-ot* | Har. *anḵurārah-ti*.

Egyp. (NE) *ḵrr* | Copt. \**ḵarār-aw* (redupl.) ‘frog’ (Vycichl 1983: 86-7).

Berb. \**ḵarw* and \**ḵarḵu/ir* (redupl.) ‘toad, frog’: Rif *a-ḵarḵur* | Iznassen *ḵarḵriw* ‘toad’ | Senhaja *a-ḵarḵur* | Ahaggar *a-ḵeru*, pl. *i-ḵer-ân* ‘frog’.

Chad. \**kir-an*- and \**ka(r)kir-an* ‘frog’: W.: Ngizim *kàrìnàkáu* || C. \**kir-in*- ‘frog’: Gava *kirè* | Mofu *kakəraŋ* (redupl.) | Daba *kìrrin* || E. \**kVrin*:- Birgit *kirén-ì* | Sokoro *kórin-ē*.

Omot. N. \**ʔu-ḵar*:- Zaysse *ʔooḵkar-o* | Zergula *ʔooḵēr-u* ‘frog’.

► SED II No. 137; AADB 270.

## 11. Fish

### 11.1. \**ʔirw*- ‘fish; spawn’

Sem. \**ʔirw*:- Akk. (SB) *e/urû-t-u* ‘(fish) spawn’.

Egyp. (Med.) *ʔr.t* ‘part of fish’ (if < \**ʔir-t*).

Chad. \**ʔirw*- ‘fish’: W. \**ʔuri*:- Hausa *ùùrii* || C. \**wuraH*- (met.?) ‘k. of fish’: Zime *wùràhù?* ‘Polypterus bichir’ [*Polypterus bichir*] || E. \**ʔir*:- Ndam *er-e* ‘fish’.

Omot. \**ʔur-t*- ‘fish’: N.: Sheko *or-us* || S.: Dime *or-co*.

► AADB 2027.

### 11.2. \**kal*- and \**kalul*- (redupl.) ‘k. of large fish’

Sem. \**kalul*:- (redupl.) Akk. (SB) *kulil*-, *kulull*- ‘a fabulous creature, part man and part fish’<sup>121</sup> || Mhr. *kell* | Jib. *kāl* ‘whale’.

Chad. \**kVl*:- W. \**kul-m*:- Hausa *kulm-a/e/i* ‘the name of a large fish’ || C. \**kalik*- (redupl.): Bura *kalik-o*, *kilakil-a* ‘a k. of fish’.

Cush. E. \**kallu-m*- ‘fish’: SA \**kullu-m*:- Afar *kullu(u)m* | LEC \**kallu-m*:- Somali *kallúun*, pl. *kalluum-o* | Bayso *kunnum-i* (assim.)

► AADB 2674; cf. Dolgopolskiy 2008 #1030 (apud Thomas).

<sup>118</sup> Cf. also *ḵonbet* ‘snake’.

<sup>119</sup> Cf. Arab. *duldul*- ‘hérisson’ and Chad. E.: Tumah *dāl* ‘rat géant’, hardly related.

<sup>120</sup> -ʕ in Tgr. and Tna. is non-etymological.

<sup>121</sup> Hardly a Sumerian loanword in view of the Sem. and AA cognates.

11.3. \**ka(w)ar-* ‘fish (generic?)’

Sem. \**ka(w)ar-* ‘k. of fish’:<sup>122</sup> Aram.<sup>123</sup>: Off. *kwr* ‘fish’ | Bab. *kəwār-* ‘fish’ | Jud. *ka(w)ar-ā* ‘fish in the cauf; fish in general’ | Mand. *kauar-a* ‘fish’ || Soq. *kúwerhor* ‘espèce de poisson’.<sup>124</sup>

Chad. \**kirw-* and \**kukirw-* (redupl.): W. \*(*ku*)*kirw-*: Hausa *kukkurw-ai* ‘name of fish’ | Bolewa *kerw-o* | Ngamo *kèrw-ò*, *kerh-o* || C. \**kikur-*: Mbara *kikur-a* ‘kind of fish (Tetraodon fahaka)’ [*Tetraodon lineatus*].

Cush. E.: LEC \**ka(w)ar-* ‘fish’: Dasenech *kaara* | Dullay \**ka(w)ar(-at)-* ‘fish’: Gawwata *haar-e* | Gollango *haar-e* | Harso, Dobase *haar-icce* | Tsamai *haar-ite*, etc.

Omot. S. \**kār-* ‘fish’: Hamer, Ongota *kāra* ‘fish’.

▲ Cf. Nub (Kenuzi, Dongola) *kāre* ‘fish’.

► AADB 372; cf. SED II No. 125; cf. Stolbova 2011 #282.

11.4. \**kar-tum-* ‘k. of fish’

Sem. \**kar-mut* (met.): Arab. *ḡarmūt-*, *ḡurmūt-*, pl. *ḡarāmīṭ* ‘*Silurus anguillaris*, k. of eel’ (Wehr 1020) [*Clarias anguillaris*] || MSA \**ṡarnīḡ* (met. <\**kar-nīṡ*): Mhr. *ṡarnīḡ* ‘name of a fish’ | Soq. *ṡárnīḡ*, *ṡáníḡ* | Boṡhari *ṡōnēḡ* ‘a kind of large black *ṡanṡad* (King macherel)’ [*Scomberomorus cavalla*] (Johnstone 1987).<sup>125</sup>

Cush. E. \**kur-ṡum-* ‘fish’: LEC \**kur-ṡum-*: Oromo *ḡurṡumm-i* | Dirayta *ḡurtum-et* | HEC \**ḡur-tum-*:<sup>126</sup> Sidamo, Darasa *ḡultiṡm-i* | Kambatta *ḡurtum-* | Alaba *ḡurṡum-et* | Hadiya, Burji *ḡurṡum-e*.

▲ Apparently a unique case of a compound word consisting of two root morphemes which are separately preserved only in NAA: \**kar-* ‘fish’<sup>127</sup> and \**tum-* ‘fish’<sup>128</sup> (*t* > *ṡ* in the compound terms influenced by *ḡ*).

► AADB 1054.

11.5. \**maw(-ay)-* ‘k. of fish’

(?) Egyp. *mr.t* (if <\**mVl-*) ‘ein Tier (zwischen Fischen genannt)’.

Chad. W. \**maw(-ay)-*: Galambu *māl-á* ‘eel’ [*Anguilla sp.*] | S.: Bauchi \**m<sup>w</sup>al(l)-aṡ-* (met.) || C. \**maw(-ay)-*: Mofu *mālāl-áy* (redupl.) ‘catfish’ [*Siluriformes sp.*] | Kotoko. *mul-nī* ‘poisson: *Gymnarchus niloticus*’ [*Gymnarchus niloticus*] | Musgu-Puss *malaw* (met.) ‘poisson: *Porcus Bayad Bayad*’ [*Bagrus bajad*] || E. \**maw(-)* (met.): Somrai *mawàl-à* ‘carpe’ [*Cyprinus sp.*].

Cush. E.: LEC \**maw(-ay)* ‘fish’: Somali *mālày* | Jiddu *mallay* | Boni *mālālái* (redupl.) | Baiso *mole* | Mossiya *mōle* | Bussa *mole* | HEC \**maw(-iy)-*: Sidamo *mawoliyā*.<sup>129</sup>

Omot. \**maw(-ay)-* ‘fish’: N.: Wolaita *moliy-a* | Dawro *moly-a* | Male *moll-o* | Dorze, Zergulla *mole* | Chara *mul-ā* | Bworo *mole* || S.: Galila *mol-ta* | Ari *mol-a*.

▲ Too much similarity between Chadic, Cushitic and Omotic forms raises suspicion of interborrowing which is, however, is hard to prove.

► AADB 3668; EDE III, 397.

<sup>122</sup> Cf. Gz. *kawwara* ‘to set traps, catch fish’.

<sup>123</sup> Hardly from Akk. (OB) *kamār-* ‘k. of a fish’ as claimed by some authors.

<sup>124</sup> Cf. also *kér* ‘kind of shark’

<sup>125</sup> Though a series of metathetic words, their quadriconsonantal roots leave little doubt about their links with the Cush. terms, whether related or borrowed.

<sup>126</sup> Borrowing from Omoto is not to be ruled out.

<sup>127</sup> Preserved, e.g., in Hausa *ḡarāyā* ‘fish’.

<sup>128</sup> NAA \**tVm-* ‘k. of fish’: Egyp. (Med.) *tm.t* ‘k. of fish’; Chad. C. \**tum-*: Munjuk *tum* ‘pêche’, *tumi* ‘pêcher’, Musgu *tum* ‘Fisch’ (rel. to Chad. \**tVm-* ‘to fish, fishing’ CED #743). Cf. also Sem. Eth. \**taman-*: Geez *taman*, Tgr., Amh. *tāmān* ‘snake, dragon’.

<sup>129</sup> Acc. to EDE 2008, < Omotic.

## Typological considerations

Extensive discussion addressing the issue of the Proto-Afrasian homeland in a chronological context can be found in our previous paper (Militarev, Nikolaev 2020). In this part, we only need to make an additional point in light of further evidence presented by semantic reconstruction in the zoonymic field.

The difficulty of reconstructing zoonyms for wild animals at the Afrasian level is hardly unique from a typological point of view. Thus, in Proto-Indo-European, the chronological depth of which is approximately half the depth of Proto-Afrasian, precise semantics can be reconstructed for only a small subset of the terms for wild animals (*\*bhebhru-* ‘beaver’, *\*lūnks-* ‘lynx’, *\*włk(w)o-* ‘wolf’, *\*h<sub>2</sub>lōp-* ‘fox’, etc.); interpretation of the rest remains difficult. Several examples will suffice:<sup>130</sup>

(1) PIE *\*h<sub>2</sub>eit-* (~*\*h<sub>3</sub>-*) ‘k. of lesser bovid’: Old Indian *éta-* m., *étā* f. ‘a k. of deer or antelope’ | Baltic: Latvian *àita*, dial. *àite* ‘sheep (generic)’ (WP I: 3);

(2) PIE *\*kerw-* ‘k. of larger bovid’: Slavic *\*kŏrovā* ‘cow’ and *\*korvō* ‘ox’ (Old Polish *karw* ‘old ox’) | Baltic *\*kaŕwiā* f. : Lithuanian *kárvė* ‘cow’ and *\*kurwa-* m. ‘bull’: Old Prussian *curwis*, acc. *kurwan* ‘bull’ | Germanic *\*xiru-ta-z*, *-ēn*, *\*xiru-tu-z* ‘deer’ [*Cervus elaphus*] | Latin *cervus* m. ‘deer’ [*Cervus elaphus*]; *cerva* f. ‘hind’ | Celtic: Welsh *carw*, Cornish *carow*, Breton *karo* ‘deer’ [*Cervus elaphus*] | Albanian *ka*, pl. *qe* ‘ox’. ♦ The root is sometimes considered to be derived from *\*k’erh<sub>2</sub>(w)-* ‘horn’, but the Balto-Slavic non-*satem* forms contradict this etymology. (WP I: 403 f.);

(3) PIE *\*dām-*, *\*dam-* ‘k. of lesser bovid’: Greek (suffixal) *δαμᾶλης* m. ‘young steer’, *δαμᾶλη* f. ‘young cow, heifer’, *δάμᾶλις* ‘young cow, heifer’, dial. ‘pig’, *δάμᾶλος* (Hdn.) ‘calf?’ | Germanic *\*tam-Vr(-Vl)-* (suffixal): Low Austrian German *zamer*, *zamerl* ‘junger Ochs’ | Latin *damma* or *dāma* ‘a general name for beasts of the deer kind: fallow-deer [*Dama dama*], chamois [*Rupicapra rupicapra*]; buck, doe, antelope’ | Celtic *\*dam-*: Cornish *da* ‘fallow-deer [*Dama dama*]’, Old Irish *dam* ‘ox’, *dam allaid* ‘deer’ [*Cervus elaphus*]; suffixal: Welsh *dafad*, Old Cornish *dauat*, Breton *dañvat* ‘sheep’. ♦ The root is sometimes considered to be derived from *\*demh<sub>3</sub>-* ‘to tame’, but its nominalization and “animal” semantics can be projected back to Proto-Indo-European in any case (WP I: 72 f.);

(4) PIE *\*sing’h-* ‘k. of large feline’: Proto-Tokharian *\*šēnś-äke* (suffixal) ‘lion’ [*Panthera leo*]: Tokh. A *śisäk*, B *šecake* | Old Indian *siṃhā-* m. ‘Id.’ | Armenian *inš* ‘leopard’ [*P. pardus*]. (WP II: 508);

(5) PIE *\*włp-* ‘k. of wild canine or feline’: Iranian (suffixal): Middle Persian *gurp-ak*, Persian *gurb-a* ‘(domestic) cat’ | Baltic (suffixal): Lithuanian *vilp-iš-ī-s* m. ‘wild cat [*Felis silvestris*]’ | Germanic *\*wúlfa-z* ‘wolf’, *\*wulbió* ‘she-wolf’ [*Canis lupus*] (a contamination with *\*włk(w)o-* ‘wolf’) | Latin *volpēs*, *vulpēs* ‘fox’ [*Vulpes vulpes*] (and *lupus* ‘wolf’, a contamination with *\*włk(w)o-* ‘wolf’ and *\*h<sub>2</sub>lōp-* ‘fox?’). (WP I: 317 f.);

(6) PIE *\*bhēl-*, *\*bhel-* ‘k. of small predator or rodent’: Slavic *\*bělv*, *\*běla* ‘squirrel’ [*Sciurus vulgaris*] (East Slavic only; contaminated with *\*bělv* ‘white’) | Latin *fēlēs* f. ‘wild cat’ [*Felis silvestris*], also ‘marten, ferret, polecat’ | Celtic (suffixal) *\*bel-ego-*: Welsh *bele* ‘marten’ [*Martes martes*] | Germanic (suffixal) *\*bel-ik-*: Old High German *bilih* m., *bilihmūs* ‘dormouse [*Glis glis*]; shrew [*Sorex sp.*]’, Middle High German *bilch* f., German *Bilch* m. ‘dormouse’ [*Glis glis*] (WP II: 177).

Examples like these, which can be easily multiplied, show that the Proto-Afrasian semantic picture as presented in the evidence assembled in this paper, with its mix of (sometimes) precise and (more often) vague, genus- or family- rather than species-based, reconstructions, should not be considered as being out of line with the commonly accepted standards of se-

<sup>130</sup> All the Indo-European examples are quoted according to Sergei Nikolaev’s online database for Indo-European etymology, hosted by the «Tower of Babel» project (<https://starlingdb.org>).

mantic reconstruction for acknowledged families — even though, of course, this does not imply that researchers in the future should not strive for more semantic precision through more careful analysis of whatever lexicographic or philological evidence is available to them.

### Appendix: Alphabetic index of reconstructed roots

- \**ʔačhayl-* ‘(big) reptile’ 9.1.  
 \**ʔandaw-* ‘mouse’ 7.2.  
 \**ʔar-* ‘hare’ 7.3.  
 \**ʔa/usk-ay-* ‘k. of canine or hyena’ 5.3.  
 \**ʔaway-* or \**yawaʔ-* (met.) ‘k. of canine’ 5.1.  
 \**ʔaw/ys-* ‘k. of canine’ 5.2.  
 \**ʔaʒur-* ‘k. of feline, hyena or viverra’ 6.1.  
 \**ʔirw-* ‘fish (generic?); spawn’ 11.1.  
 \**ʔV(n)çaw-* ‘k. of smaller carnivoran or rodent’ 7.1.  
 \**(ʕa-)gwir-* ‘k. of bird’ 8.6.  
 \**ʕariw-* ~ \**ʕiraw-* ‘k. of feline’ 6.2.  
 \**baʔy-* ‘leopard or lion’ 6.3.  
 \**baʔis-* ‘k. of feline (*Felis silvestris*)’ 6.4.  
 \**baʔVš-* ‘k. of canine or hyena’ 5.4.  
 \**bač-* ‘snake, reptile’ 9.2.  
 \**bar-* ‘k. of bird (pigeon, partridge)’ 8.1.  
 \**bat-* ‘k. of snake or worm’ 9.3.  
 \**bawih-* ‘k. of canine’ 5.5.  
 \**bawr-* ‘k. of large carnivorous animal’ 7.4.  
 \**çiraʕ-* ‘k. of bird’ 8.2.  
 \**çap-* ‘k. of smaller carnivoran or rodent’ 7.5.  
 \**çamb-* ‘some water creature’ 10.2.  
 \**dab-* ‘k. of large animal’ 7.6.  
 \**da(ʔ/y)m-* ‘k. of feline’ 6.5.  
 \**dang<sup>(w)</sup>Vl/r-* ‘monkey; dwarf’ 4.1.  
 \**dVr-* ‘k. of bird’ 8.3.  
 \**dawiʔ-* ‘k. of snake or worm’ 9.4.  
 \**gVd-* ‘k. of canine or hyena’ 5.6.  
 \**guray-* ‘k. of monkey’ 4.2.  
 \**giwar-* ‘k. of feline’ 6.6.  
 \**g<sup>w</sup>am-* ‘k. of bird’ 8.5.  
 \**g<sup>w</sup>ay-* and \**g<sup>w</sup>ag<sup>w</sup>ay-* (redupl.) ‘k. of bird’ 8.4.  
 \**ɣuraʔ-* ‘crow’ 8.7a.  
 \**ɣurayb-* ‘crow, raven’ 8.7b.  
 \**ɣurayk-an-* ‘k. of bird (crane, raven)’ 8.7c.  
 \**har-* ‘crocodile’ 9.5.  
 \**has-* ‘k. of big snake or crocodile’ 9.6.  
 \**kal-* and \**kalul-* (redupl.) ‘k. of large fish’ 11.2.  
 \**kur(-ay-)* ‘k. of canine’ 5.7.  
 \**kawit-* ‘dog’ 5.8.  
 \**kawar-* ‘fish (generic?)’ 11.3.  
 \**k<sup>w</sup>ihān-* ‘dog’ 5.9.  
 \**k<sup>w</sup>ay-* and \**k<sup>w</sup>ak<sup>w</sup>ay-* ‘k. of crow or bird of prey’ 8.9.  
 \**k<sup>w</sup>aʔay-* and \**k<sup>w</sup>ak<sup>w</sup>ay-* ‘k. of gallinaceous bird’ 8.10.  
 \**KVʒim-* or \**KVʒim-* ‘dog’ 5.10.  
 \**ka(ʔa)yr-* ‘monkey’ 4.3.  
 \**k<sup>w</sup>Vr-* ‘k. of big snake or crocodile’ 9.7.  
 \**k<sup>w</sup>ar-tum-* ‘k. of fish’ 11.4.  
 \**k<sup>w</sup>aʕ-* and \**k<sup>w</sup>aʕk<sup>w</sup>aʕ-* (redupl.) ‘frog’ 10.1.  
 \**k<sup>w</sup>ar-* ‘frog’ 10.4.  
 \**k<sup>w</sup>a(ʔ)r-* ‘crow, raven’ 8.11.  
 \**labiʔ-* ‘lion’ 6.7.  
 \**layč-* ‘k. of large feline’ 6.8.  
 \**mawl-ay-* ‘k. of fish’ 11.5.  
 \**mary-* or \**mayr-* ‘k. of feline’ 6.9.  
 \**sib-* ‘k. of snake or worm’ 9.8.  
 \**sVf-* ‘k. of snake’ 9.9.  
 \**sa/ik<sup>w</sup>-* ‘k. of lesser animal’ 7.7.  
 \**sawr-* and \**sarw-* ‘k. of feline’ 6.10.  
 \**wahr-* ‘k. of canine; hyena’ 5.11.  
 \**wanš-* ‘k. of canine’ 5.12.  
 \**ʒaʔib-* ‘k. of scavenger’ 7.8.  
 \**ʒag<sup>(w)</sup>-il/r-* ‘k. of monkey’ 4.4.

### Abbreviations

#### Names of languages

AA – Afrasian = Afro-Asiatic; Akk. – Akkadian; Amh. – Amharic; Arab. – Arabic; Aram. – Aramaic; Berb. – Berber; Bibl. – Biblical Aramaic; Cha. – Chaha; Chad. – Chadic; Copt. – Coptic; Cush. – Cushitic; Dem. – Demotic; E. Wlm – East Tawlllemmet; Ebl. – Ebla; Egyp. – Egyptian; End. – Endegen; Enm. – Ennemor; ESA – Epigraphic South Arabian; ESud – East Sudanic; Eth. – Ethiopic; Ez. – Eza; Gog. – Gogot; Gyt. – Gyeto; Gur. – Gurage; Gz. – Ge’ez; Har. – Harari; Hbr. – Hebrew; Hbt – Hobyot; HEC – Highland East Cushitic; Hrs. – Harsusi; Jib. – Jibbali; Jud. – Judaic Aramaic; LEC – Lowland East Cushitic; Mand. – Mandaic; MB – Middle Babylonian; Mhr. – Mehri; Min. – Minaean; MK – Middle Kingdom; Msq. – Masqan; MSA – Modern South Arabian; Muh. – Muher; NAA North Afrasian; NE – New Kingdom; NS – Nilo-Saharan; Oakk – Old Akkadian; OB – Old Babylonian; Off. – Official

Aramaic; OK – Old Kingdom; PAA – Proto-Afrasian; PAlt – Proto-Altaic; PCO – Proto-Cushitic-OmotiC; Phoen. – Phoenician; PIE – Proto-Indo-European; PNorthCauc. – Proto-North-Caucasian; PSEBC – Proto-Semitic-Egyptian-Berber-Chadic; Qat. – Qatabanian; S. Cush. – Southern Cushitic; SA – Saho-Afar; Sab. – Sabaic; Sam. – Samaritan; Sel. – Selti; Sem. – Semitic; SNil – South Nilotic; Soq. – Soqotri; Sum. – Sumerian; Syr. – Syriac; Tgr. – Tigre; Tna. – Tigrinya (Tigray); Ugr. – Ugaritic; Wol. – Wolaita; YB – Young Babylonian.

#### Terminology

coll. – collective; dial. – dialect; dissim. – dissimilation; f. – feminine; gen. – generic (term); id. – idem; k. – kind; lw. – loanword; m. – masculine; met. – metathesis; pers. – personal; pl. – plural; pref. – prefix; redupl. – reduplication; sg. – singular; sp. – species; subst. – substantive; suff. – suffix.

#### Other notational symbols

|| separates branches within a language family

| separates subbranches or groups

- denotes morphemic boundaries<sup>131</sup>

#### Abbreviations of sources

AADB = Militarev, Stolbova 2020

AHw = von Soden 1965–1981

BK = Biberstein-Kazimirski 1860

CAD = Brinkman et al. 1956–2010

CED = Stolbova 2016

DRS = Cohen et al. 1970–1993

EDE = Takács 1999, 2001, 2008

EDS = Parpola 2015

HSED = Orel, Stolbova 1995

LSF = Attinger 2019

PSD = Anon. 2019 (The Pennsylvania Sumerian Dictionary)

SED II = Militarev, Kogan 2005

WP = Walde, Pokorny 1926–1930

## References

- Anon. 2019. *The Pennsylvania Sumerian Dictionary, ePSD2 Public Beta 5 (built 2019-05-24)*. Online resource at: <http://psd.museum.upenn.edu/>.
- Attinger, Pascal. 2019. *Lexique sumérien-français (textes traduits dans Attinger février 2019)*. Ms., online at: <http://www.iaw.unibe.ch/atinger>.
- Bargery, George P. 1934. *A Hausa-English Dictionary and English-Hausa Vocabulary*. London: Oxford University Press.
- Bassano, Francesco da. 1918. *Vocabolario tigray-italiano e repertorio italiano-tigray*. Roma: Casa editrice italiana & C. de Luigi.
- Biberstein-Kazimirski, Albert de. 1860. *Dictionnaire arabe-français. Vols. 1–2*. Paris: Maisonneuve et Co.
- Blažek, Václav. 2003. Fauna in Beja Lexicon. In: Ilya Smirnov (ed.). *Studia Semitica: Fs. for Alexander Militarev*: 230–294. Moscow: Russian State University for the Humanities.
- Blažek, Václav. 2007. Beja historical phonology: Consonantism. In: Azeb Amha, Maarten Mous, Graziano Savà (eds.). *OmotiC and Cushitic Language Studies. Papers from the Fourth Cushitic OmotiC Conference (Leiden, April 2003)*: 125–156. Köln: Rüdiger Köppe Verlag.

---

<sup>131</sup> In some cases, especially for Chadic lexemes, boundaries drawn between the stem and affixes, or between several affixes, may not be exact, since perfect morphemic analysis requires more detailed knowledge of each individual language's morphology and morphophonemics than possessed by the authors.

- Blažek, Václav. 2008. Lexicostatistical comparison of Omotic languages. In: John D. Bengtson (ed.). *Hot Pursuit of Language in Prehistory*: 57–148. Amsterdam: John Benjamins.
- Blažek, Václav. 2013. Indo-European zoonyms in Afroasiatic perspective. *Journal of Language Relationship* 9: 37–54.
- Blažek, Václav. n.d. *Simiae*. Unpublished Ms.
- Brinkman, John A., Miguel Civil, Ignace J. Gelb, A. Leo Oppenheim, Eruca Reiner (eds.). 1956–2010. *The Assyrian Dictionary of the Oriental Institute, the University of Chicago*. Chicago, Illinois, U.S.A.: Oriental Institute; Glückstadt, Germany: J.J. Augustin erlagsbuchhandlung.
- Caprile, Jean Pierre. 1975. *Lexique Tumat-Français*. MSA 5. Berlin: Reimer.
- Cerulli, Enrico. 1951. *Studi etiopici. IV. La lingua caffina*. Roma: Istituto per l'Oriente.
- Cohen, David, Jean Cantineau, François Bron, Antoine Lonnet. 1970–2010. *Dictionnaire des racines sémitiques ou attestées dans les langues sémitiques*. Paris: La Haye / Mouton: Leuven.
- Cosper, Ronald. 1994. *South Bauchi Lexicon. A Wordlist of Nine South Bauchi (Chadic) Languages and Dialects*. Halifax.
- Dozy, Reinhart. 1927. *Supplément au dictionnaires arabes. T. I-II*. Leiden / Paris: E. J. Brill / Maisonneuve.
- Dolgopolskiy, Aron B. 1973. *Sravnitel'no-istoričeskaja fonetika kušitskix jazykov* [Comparative-Historical Phonetics of the Cushitic Languages]. Moskva: Nauka.
- Dolgopolskiy, Aharon B. 2008. *Nostratic Dictionary*. Cambridge, UK: McDonald Institute for Archaeological Research University of Cambridge.
- Erman, Adolf, Hermann Grapow. 1957–1971. *Wörterbuch der ägyptischen Sprache*. I–V. Berlin: Akademie-Verlag.
- Foucauld, père Charles de. 1951–1952. *Dictionnaire touareg-français*. Paris: Imprimerie Nationale de France.
- Freytag, Georg Wilhelm. 1833. *Lexicon arabico-latinum. T. I-IV*. Halle: Halis Saxonum.
- Johnstone, Thomas M. 1987. *Mehri Lexicon and English-Mehri Word List*. London / New York: Routledge.
- Jungraithmayr, Herrmann, Dymitr Ibrizimow. 1994. *Chadic Lexical Roots. Vols. I-II*. Berlin: Dietrich Reimer Verlag.
- Kiessling, Roland, Maarten Mous. 2003. *The Lexical Reconstruction of West-Rift Southern Cushitic*. Köln: Rüdiger Köppe Verlag.
- Kraft, Charles H. 1981. *Chadic Wordlists I-III*. Berlin: D. Reimer.
- Lane, Edward W. 1863–1893. *An Arabic-English Lexicon. Vols. I-VIII*. London / Edinburgh: Williams and Norgate.
- Leslau, Wolf. 1979. *Etymological Dictionary of Gurage (Ethiopic). Vol. III: Etymological section*. Wiesbaden: Otto Harrassowitz.
- Leslau, Wolf. 1987. *Comparative Dictionary of Ge'ez (Classical Ethiopic)*. Wiesbaden: Otto Harrassowitz.
- Lukas, Johannes. 1953. *Die Sprache der Tubu in der Zentralen Sahara*. Berlin: Akademie-Verlag.
- Militarev, Alexander, Leonid Kogan. 2005. *Semitic Etymological Dictionary. Vol. II: Animal Names*. Münster: Ugarit-Verlag.
- Militarev, Alexander, Olga Stolbova. 2021. *Afroasiatic Etymology Database*. Available online at: <http://starlingdb.org>.
- Militarev, Alexander, Sergei Nikolaev. 2020. Proto-Afrasian names of ungulates in light of the Proto-Afrasian homeland issue. *Journal of Language Relationship* 18/3-4: 199–226.
- Nait-Zerrad, Kamal. 1998–2002. *Dictionnaire des racines berbères*. Paris-Louvain: Peeters.
- Orel, Vladimir, Olga Stolbova. 1995. *Hamito-Semitic Etymological Dictionary*. Leiden / New York / Köln: Brill.
- Parpola, Simo. 2015. *Etymological Dictionary of the Sumerian Language. Part 1. Lexical Evidence*. Winona Lake, Indiana: Eisenbrauns.
- Prasse, Karl-Gustav. 1969. *A propos de l'origine de H touareg (tahaggart)*. *Historisk-filosofiske Meddelelser* 43/3. Copenhagen: Det Kongelige Danske Videnskaberne Selskab.
- Sasse, Hans-Jürgen. 1976. *The Consonant Phonemes of Proto-East-Cushitic (PEC): A First Approximation*. Malibu: Undena Publications.
- Sasse, Hans-Jürgen 1982. *An Etymological Dictionary of Burji*. Hamburg: Helmut Buske Verlag.
- Soden, Wolfram von. 1965–1981. *Akkadisches Handwörterbuch*. Wiesbaden: Otto Harrassowitz.
- Stolbova, Olga V. 2005. *Chadic lexical database. Issue I*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Stolbova, Olga V. 2009. *Chadic Lexical Database, Issue III*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Stolbova, Olga V. 2011. *Chadic Lexical Database, Issue IV*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Stolbova, Olga V. 2016. *Chadic Etymological Dictionary*. Moscow: Institute of Oriental Studies of the Russian Academy of Sciences.

- Stolbova, Olga V. 2019. *Chadic Lexical Database, Issue V*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Stolbova, Olga V. 2021. *Chadic Lexical Database, Issue VI*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Takács, Gábor. 2001. *Etymological Dictionary of Egyptian. Vol. 2: B-, P-, F-*. Leiden / Boston / Köln: Brill.
- Takács, Gábor. 2008. *Etymological Dictionary of Egyptian. Vol. 3: M*. Leiden / Boston: Brill.
- Takács, Gábor. 2011. *Studies in Afro-Asiatic Comparative Phonology: Consonantism*. Berlin: Dietrich Reimer Verlag.
- Vycichl, Werner. 1983. *Dictionnaire étymologique de la langue copte*. Leuven: Peeters.
- Walde, Alois, Julius Pokorny. 1926–1930. *Vergleichendes Wörterbuch der indogermanischen Sprachen*. Bd. I–II. Berlin / Leipzig: W. de Gruyter.
- Wehr, Hans. 1958. *Arabisches Wörterbuch für die Schriftsprache der Gegenwart*. Wiesbaden: Otto Harrassowitz.
- Wölfel, Josef Dominik. 1965. *Monumenta linguae Canariae. Die kanarischen Sprachdenkmäler. Eine Studie zur Vor- und Frühgeschichte Weissafrikas*. Graz: Akademische Druck- und Verlagsanstalt.

А. Ю. Милитарев, С. Л. Николаев. Праафразийские названия некопытных животных в свете проблемы афразийской прародины

В статье представлена вторая порция относительно полного тезауруса праафразийских зоонимов, подготовленного авторами — более шестидесяти реконструированных названий обезьян, собачьих, гиен, кошачьих, грызунов и других млекопитающих, птиц, пресмыкающихся, земноводных и рыб, отражающих максимально полную (на современном уровне наших знаний) картину этих видов фауны в период накануне распада праафразийского языка (11 тыс. до н.э. согласно глоттохронологическим подсчетам) и в зоне обитания праафразийцев, две различные точки зрения на локализацию которой (Левант и Восточная Африка) также обсуждаются авторами.

*Ключевые слова:* афразийская прародина; афразийские (афро-азиатские) языки; этимология зоонимов; реконструкция культурной лексики.



## Once again on the history and validity of the Sino-Tibetan bifurcate model

Sino-Tibetan is the hypothesis that postulates a bifurcate genetic relationship between Sinitic and Tibeto-Burman. The history of the subgrouping of its components as well as its overall membership are briefly though exhaustively summarized. Special attention is given to the methodological shortcomings of a recent Bayesian approach that further professes to substantiate this bifurcate Sino-Tibetan model. Future directions and broader affinities of Trans-Himalayan are also discussed.

*Keywords:* Trans-Himalayan family; Sino-Tibetan family; Bayesian inference; language subgrouping.

### 1. Introduction

The Trans-Himalayan family (also referred to as Sino-Tibetan or Tibeto-Burman) is a very large linguistic family, comprising around 600 languages (Turin 2007, Owen-Smith and Hill 2014) spoken across a wide geographic area. It is the second populous linguistic family of the world, with nearly 1.4 billion speakers (second only to Indo-European), and one of the most diversified families, encompassing languages with very different morphological systems, from isolating languages such as Chinese and Burmese, to polysynthetic languages such as rGyalrongic or Kiranti. ‘Sino-Tibetan’ is the hypothetical linguistic family postulating a bifurcate genetic relationship between Sinitic and all other Trans-Himalayan languages, which are thus grouped under a “Tibeto-Burman” rubric.

Today the existence of a Sino-Tibetan/Trans-Himalayan family, whatever be the subgrouping of its components, is widely accepted among linguists (Greenberg 1996: 134; LaPolla 2001: 225). Since the beginning of the 21st century, a number of publications, including scientific articles, books and monographs have been dedicated to the subject, e.g. Ting and Sūn (2000, 2001), van Driem (2001), Beckwith (2002),<sup>1</sup> Gong (2002), Matisoff (2003), Thurgood and LaPolla (2003), Saxena (2004), Sagart *et al.* (2005), Owen-Smith and Hill (2014), Hill (2019), and so on, and the field is becoming more well-covered. Nevertheless, disagreement still persists on some crucial aspects of Sino-Tibetan/Trans-Himalayan, especially for what regards its overall membership, as well as the subgrouping of its components. While some scholars accept a bifurcate Sino-Tibetan family (Bradley 1997; LaPolla 2002; Thurgood and LaPolla 2003; Matisoff 2003), others contend that no serious linguistic evidence which may warrant the separation of Sinitic from the rest of the family has been shown (van Driem 2001, 2005, 2007, 2011; DeLancey 2014). Thus, they have suggested renaming the family with the more agnostic term ‘Trans-Himalayan’ (van Driem 2014; Owen-Smith and Hill 2014). The present survey article provides an overview of both historical and recent works in Sino-Tibetan/Trans-Himalayan linguistics, with an eye toward a few issues which may be of particular interest to both specialists and non-specialists of this branch of linguistics. But before discussing in more detail these

---

<sup>1</sup> This author, however, contends that Sinitic and Tibeto-Burman are not related. For ST discussions, see in particular Bradley (2002: 73–112) and Beckwith (2002: 113–157).

aspects, a remark seems to be necessary. The existence of Sino-Tibetan/Trans-Himalayan has not yet been proven (Handel 2008: 423), because it cannot, and, by the same token, it cannot be disproved. Just like any other early linguistic relationship, including those which are universally or almost-universally accepted such as Indo-European, Uralic and Semitic, Sino-Tibetan/Trans-Himalayan, too, necessarily remains confined into the realm of non-demonstrable hypotheses. Some of these hypothetical, now-lost *Ursprachen* are, indeed, well-documented, whereas others are, instead, less-thoroughly explored. The question, nonetheless, is not whether they may or may not be definitely proven (cf. Handel 2008: 423)—because they cannot—but instead how convincing is the evidence that may be brought in support of these specific hypotheses. Although this statement may sound trivial, as it should be well-known to any serious worker in the field of historical linguistics, there are specialists who, occasionally, seem to be unaware of this fact.

## 2. The historical classification of Sino-Tibetan

The history of the Sino-Tibetan hypothesis has been illustrated in a number of publications by George van Driem (1997 et seq.), which are perhaps the most informative sources on the argument. This section, however, does not constitute a recapitulation of van Driem's detailed history of the Western linguistic development of the hypothesis of a Sino-Tibetan language family. While a section of this length is necessarily incomplete, the present paragraph demonstrates that the history of Sino-Tibetan is more controversial, debated and complex than usually assumed. According to van Driem, workers in Sino-Tibetan divide in two types: those who work in the received tradition of John Leyden's Indo-Chinese paradigm, which he believes to be what most specialists now refer to as Sino-Tibetan, and those who operate within the framework of Julius Klaproth's Tibeto-Burman. Van Driem reckons Sinitic as one of the many branches of Tibeto-Burman and therefore he still refers to the family as a whole as Tibeto-Burman (or Trans-Himalayan) instead of Sino-Tibetan. Van Driem's arguments in favour of a Tibeto-Burman/Trans-Himalayan family chiefly rely on three points: (i) Chinese and Limbu, a Kiranti language spoken in Nepal, exhibit certain vestiges of early morphological connections (van Driem 1997: 463–471). He also adduced lexical evidence which further purports to prove the existence of this early, now-lost “Sino-Bodic” connection (van Driem 1997: 471–484); (ii) from a historical point of view, the alleged bifurcation of Sino-Tibetan was nothing else than a mere scholastic artefact. Chinese appeared to be sensibly different from other Tibeto-Burman languages only because early reconstructions of Chinese, especially those championed by the great Swedish Sinologist Bernhard Karlgren (1889–1978) were methodologically flawed and thus unable to reveal many of its morphophonemic characteristics (see below for a more detailed discussion); (iii) extra-linguistic evidence, such as archaeology and the history of migrations, seems to support his subgrouping (van Driem 2001, 2007; cf. Handel 2008: 429). Other scholars have dismissed van Driem's subgrouping of Sino-Tibetan as an example of ‘neosubgroupitis’ (cf. Matisoff 2000); still others have rejected the arguments in favour of both Sino-Bodic and Trans-Himalayan because they find them “less than convincing due to problems of fact and argumentation” (LaPolla 2016). Although certain problems are recognized,<sup>2</sup> it is not necessarily mistaken to agree with van Driem on two fundamental

<sup>2</sup> The present writer chiefly disagrees on the following points: (i) no matter what is our opinion of Heinrich Julius von Klaproth (1783–1835), Klaproth's work is merely of literary interest, and does not indicate an extraordinary foresight. Klaproth worked mainly in the received strand of nineteenth century linguistics, with its particular

points: (i) the bifurcate Sino-Tibetan model was never anything more than a mere impressionistic label; (ii) had more refined reconstructions of the sound system of Old Chinese been available to Benedict and other pioneers, they probably would not have outlined a bifurcate Sino-Tibetan phylum.

We start from the second point. For more than forty years, Karlgren's reconstructions have been almost the sole reference available in Europe, as well as in East Asia and the United States. However, Karlgren's works were more historical than comparative. In fact, like any other before (and perhaps even after) him, Karlgren made use of the "comparative method," a method which he openly distrusted (Karlgren 1915–1926: 45), only when the identification of the phonetic value of a certain sound class was not otherwise knowable. Furthermore, being a severe opponent of phonemics, he apparently lost himself into the triviality of reconstructing as many phonetic contrasts as the number of sound classes of medieval Chinese rime tables required. This led him to reconstruct a system which was studded with contrasts that were hardly describable with the acoustic-articulatory attributes known from other living languages, including other Sino-Tibetan/Trans-Himalayan tongues. But, more relevant to our case is the fact that Karlgren (1931) embarked in an open polemic with Walter Simon (1893–1981), when he tried to compare Chinese with Written Tibetan in the attempt at finding the solution to certain problems pertaining the reconstruction of final consonants in Old Chinese. Karlgren did not deny that Chinese was a branch of Sino-Tibetan (which he called Indo-Chinese). Rather, he thought that the radical phono-semantic changes that had occurred in Chinese lexicon, as well as the huge time-depth that existed between Sinitic and its sister languages, rendered infeasible one-to-one comparisons between Chinese and Tibetan in Indo-European fashion, let alone the identification of regular sound laws. Hence, Karlgren urged to build a stronger knowledge for Chinese lexicon. This was the idea which lied behind Karlgren's emphasis on word families (*cízú* 詞族). Unfortunately, the complexity of Karlgren's approach to word families was such that later scholars, such as Robert Shafer (1967: 9), completely misunderstood Karlgren's real intentions, as it is clearly illustrated by Shafer's critique to Karlgren's alleged advocacy for a "relaxation" of *Lautgesetze*.

When Paul Benedict moved to Berkeley in 1938 to join Kroeber's Sino-Tibetan Philology project, Karlgren (1915–1926) was the only complete system of reconstructions available to him, and his reconstructed Old Chinese was visibly different from Tibeto-Burman. In fact, only several years after Karlgren's death, the structure of Old Chinese morphemes and roots began to be better understood, especially after Sagart (1993: 238, 255–257) criticized the absence of a framework to handle affixation processes in Baxter's system (1992). We cannot know whether Benedict and other pioneers would have accepted a bifurcate Sino-Tibetan family, had more refined reconstructions of Old Chinese been available to them, but there are good reasons to believe that they probably would not have outlined such a bifurcate scheme, if Karlgren's system, as many other systems reconstructed after him, had contained an \*s- causative and denominative prefix (Mei 1985: 334–343, 1989; Baxter and Sagart 1998: 53; Schuessler 2007: 52–54) which could also form iteratives and intensives (Schuessler 2007: 53), an \*m- prefix for the direction/causativity dichotomy (Takashima 1996; J. Sun 1993), a \*-t suffix for transitivization (Benedict 1972: 98–102; Michailovsky 1985; van Driem 1988), and a nominalizing \*-n suffix (LaPolla 1994; Jin 1998; Sagart 1999 [but abandoned in later publications]; Schuessler 2007).

---

emphasis for the typological classification of languages. Although Klapproth's work may be regarded as better than the work of other scholars of his time, his classifications too were based on criteria that can no longer satisfy contemporary scientific requirements; (ii) the "fallen leaves" model of the family appears to imply the existence of a *Stammbaum* of which, however, we have no trace (cf. LaPolla 2016: 291).

The other point discussed by van Driem is that the establishment of a bifurcate Sino-Tibetan phylum has never been verified in accordance with serious linguistic methods, using accepted protocols of evaluation of results. Although we may agree on the fact that Sino-Tibetan linguistics is relatively young (cf. Matisoff 1991: 469), and that only Benedict and few other scholars may be credited for pioneering Sino-Tibetan linguistics, we must remark, however, that certain embryonic forms of Sino-Tibetan or Trans-Himalayan subgrouping can be traced to earlier times. As argued in this section, van Driem claims that workers in Sino-Tibetan linguistics divide in two groups: those who followed the Indo-Chinese tradition started by John Leyden (1775–1811), and those who instead still adhere to Julius von Klaproth's Tibeto-Burman model, which was allegedly “accepted by British scholars such as Forbes, Houghton and Cust” (van Driem 2007: 215). However, this dichotomy is not necessarily correct. Surely, the name ‘Tibeto-Burman’ was soon popularized by British scholars, such as Capt. Charles James Forbes (?–1879) and Robert Needham Cust (1821–1909), but Chinese was never explicitly included within this phylum. For instance, Robert Cust (1881), in an article presented at the Fourth International Congress of Orientalists in Florence (1878), grouped together some non-Aryan languages of ‘East India.’ He distinguished six language families: Dravidian, Kolarian, Tibeto-Burman, Khasian (now also part of Austroasiatic), Tai and Mon-Anam (both being Austroasiatic). He never mentioned Chinese, therefore we do not know whether in his opinion Chinese was part of that Tibeto-Burman family, or whether it was reckoned as a linguistic isolate. For what regards Forbes's ‘Tibeto-Burman,’ as explained by he himself, the term was nothing else but the convenient designation of a very wide family of languages which corresponded more or less to Max Müller's Gangetic and Lohitic ‘classes’ (Forbes, 1878:210). He mainly revised the previous Tibeto-Burman scheme outlined by the American missionary Francis Mason (1799–1874), who included, in this family, the dialects spoken by eleven different Burmese tribes, such as “Burmese, Arracanese, Mugs, Kanyans, Toungooers, Tavoyers, Yos, Yebains, Pyus, Kados and Danus” (Forbes 1878: 211). Forbes left only six of Mason's original eleventh tribes under his Burman sub-phylum, to which he added also Naga, Singpho and Karen. As we can see, Julius von Klaproth could not exercise a decisive influence over these two British scholars. Rather, it seems that Max Müller (1823–1900) and his Turanian family exerted a much wider influence.<sup>3</sup> Müller is credited with being the first to have outlined the existence of a Tibeto-Burman phylum (Grierson and Konow 1909: 13; Campbell and Poser 2008: 113). In fact, Müller's *Bothiya* class of languages includes many Trans-Himalayan idioms. Most of these languages had been extensively studied by Brian Houghton Hodgson (1801–1894), who also championed the Turanian hypothesis, and had pointed out similarities in the use of personal pronouns between languages of the Circassian area and rGyalrongic. Müller divided *Bothiya* into two groups, Lohitic and Sub-Himalayan (or Gangetic). They included languages such as Burmese, Tibetan, Garo, Naga, Bodo, Kachin, Kuki-Chin, Singpho, Mising and many other dialects of North Assam (Müller 1854: 97–126).

What most scholars have overlooked, however, is the fact that a Trans-Himalayan phylum had been described long ago by a member of the Indian Civil Service, Joseph Edward Lyall Brandreth (1823–1907). Writing in the *Journal of Royal Asiatic Society* (1878), Brandreth outlined the richest and most exhaustive classification of Trans-Himalayan languages. Brandreth consulted the many communications that Robert Cust had made to the Philological Society, as well as other works and reports written by Sir George Campbell (1824–1892) and Sir William Wilson Hunter (1840–1900). Brandreth's classification of Trans-Himalayan languages was also grounded on Müller's Turanian, and his very aim was to correct some deficiencies

<sup>3</sup> For more information about ‘Turanian’, see Orlandi (2020).

that the Turanian class contained (Brandreth 1878: 2). Brandreth divided the non-Indo-European (non-Aryan in his words) languages of India into six groups, *viz.* Dravidian, Kolarian, Tibeto-Burman, Khasi, Tai, and Mon-Anam. For what regards Brandreth's classification of Tibeto-Burman languages, he further divided this huge family into nineteen different classes on the basis of mutual intelligibility and other grammatical features. That this was a Trans-Himalayan classification, rather than a Tibeto-Burman one, is highlighted by the fact that although the family was referred to as 'Tibeto-Burman,' Brandreth collocated Tibetan, together with Sherpa and other dialects spoken across Sikkim and Nepal, under the second class, whereas Burmese stood as an isolate in his sixteenth class. Although Brandreth recognized certain phonetic parallelisms between the two languages, he kept them separated because "Burmese has not the determinative syllables or letters of the Tibetan, for which the tones are the substitutes," and also because "[t]he Burmese verb has no person endings, but it has a plural suffix *kra*" (Brandreth 1878: 22). Verbal aspects, plus other features related to word order were reckoned by Brandreth as diagnostic for the sub-classification of Tibeto-Burman (i.e. Trans-Himalayan) languages.

As we can see, there was no pure dichotomy between Indo-Chinese and Tibeto-Burman (perhaps the only dichotomy was between monophyletic vs polyphyletic origin of Asian languages). These terminologies were often imprecise and covered different types of linguistic families. Furthermore, claiming that a dichotomy was likewise existing between followers of Leyden's Indo-Chinese and Klaproth's Tibeto-Burman would require us to ignore both Brandreth's Trans-Himalayan (clearly a different type of hypothesis) and Müller's Bothiya. It is also unlikely that the racist views of Indo-Chinese workers were mostly concerned with typological classification and had little to do with genetic subgrouping, as claimed by van Driem (2004, 2007, 2014). It is true that, e.g., monosyllabism was often associated with backwardness, yet many influential scholars such as Jean-Pierre Abel-Rémusat (1788–1832) and Charles-Joseph de Harlez de Deulin (1832–1899) were unconvinced of this erroneous causality.<sup>4</sup> German scholars had rejected the monosyllabic view at least since Theophilus (Gottlieb) Siegfried Bayer (1703: 106), and, most importantly, Georg von der Gabelentz (1840–1893), in his article presented at the Fourth International Congress of Orientalists (1878), ventured into a *genealogical* classification of Indo-Chinese languages, i.e. not based on typological criteria. Gabelentz started his investigation from morphology. After having identified a resemblance in the morphology of Indo-Chinese languages, he sought to establish the *Lautverschiebungsgesetze* (phonetic displacement) between these tongues. In the attempt to obtain as reliable data on lesser known Indo-Chinese languages as possible, Gabelentz mainly worked on second-hand materials, especially on those studies which had been published some years earlier in the *Journal of the Indian Archipelago*. An intellectual step forward is remarked by the fact that he did not compare lexical items in their modern forms. For instance, when he compared the word for 'ear' in Chinese, Karen and Tibetan, he did not take the character 耳 in its Mandarin clothing but in its medieval Chinese reading \*ni (Gabelentz 1881: 289, 292). Moreover, although Gabelentz observed a phonetic correspondence between Thai liquid consonant *l* and Cantonese *h* (Mandarin *f*), he also cautioned about some alleged cases of cognancy between Chinese, Thai and Hmong. Gabelentz also remarked that monosyllabism proved nothing, and was not a diagnostic feature for typological, let alone genealogical classification. To sum up, Gabelentz's *Indochinesisch* encompassed Chinese, Siamese (Thai), Shan, Kiranti, Lao, Miao (Hmong), Tibetan and Burmese. Some years later, Conrady (1896, 1915), like Trombetti (1905) and other proponents of the Indo-Chinese theory, grouped Tibeto-Burman from one side

<sup>4</sup> See, respectively, Abel-Rémusat (1825: 47, 60) and Harlez (1878: 67–88)

(whose subgroup also included a *nepalische und assamesische Gruppe*) and Sino-Siamese from the other. He studied extensively the grammar of these languages, especially of Chinese, and also recognized some important connections between certain types of tones and voiced initials.

However, van Driem (2007: 221–226, 2011) is probably correct when he concludes that the Indo-Chinese legacy still defines current Sino-Tibetan phylogeny, not because Sino-Tibetanists still accept the typological classification and/or the racial prejudices which were the run of the mill until the first half of nineteenth century, but because the Sino-Tibetan hypothesis still defines a Tibeto-Burman taxon which encompasses all non-Sinitic languages after the removal of Daic (wrongly reckoned as being a branch of Sino-Tibetan by Whitney (1867: 336), Kuhn (1889), Trombetti (1905), Finck (1909), and others). Van Driem's other critics argue that those who subscribe to the standard bifurcate Sino-Tibetan phylogeny certainly do not do so because of influence from nineteenth century typologists, but mainly because of Li Fang-kuei's analysis of the languages of China and also later comparative works by Benedict (1972, 1976) and Matisoff (1973, 1991, 2003) (LaPolla 2016: 288). However, Li Fang-kuei's Sino-Tibetan (1937) is a "macro Sino-Tibetan" taxon, encompassing also Kra-Dai (*ZhuàngDòngyǔ* 壯侗語) and Hmong-Mien (*MiáoYáoyǔ* 苗瑤語), in Gabelentz and Conrady fashion.<sup>5</sup> Lamenting that Western scholars had scarce familiarity with 'tones,' which he regarded as important as vowels and consonants, Li sought to explain the striking similarities between certain Kra-Dai, Hmong-Mien and Sinitic languages, by assuming that they were inherited from a common ancestor, and not by assuming that they reflect areal linguistics or other types of linguistic contacts.<sup>6</sup>

Furthermore, Robert Shafer (1955: 98–99, 1965: 445–446), whom Benedict (1975) saw as the first Sino-Tibetanist, also described the Sino-Tibetan family as encompassing Daic, albeit he rejected Konow's opinion (1909: 1) according to which Sino-Tibetan was divided into an eastern "Chinese-Siamese" sub-family and a western "Tibeto-Burman" one (Shafer 1955: 94).<sup>7</sup>

Another early work on Sino-Tibetan was written between 1964 and 1965 by Charles Frederick Voegelin (1906–1986) and Florence Marie Voegelin (1927–1989). This lengthy work, which was divided into five fascicles published in *Anthropological Linguistics*, also presents a Sino-Tibetan phylum which encompasses Sinitic, Hmong-Mien, Kra-Dai (called Kam-Tai and including a south-western group [Tai-Shan languages], a central group [Nung-Tho languages], and a northern group [Kam-Sui languages]), Naga-Kuki-Chin, Karen, Lolo-Burmese, Tibetan and 'Gyarung-Mishmi.' The Sino-Tibetan family, as presented by Voegelin and Voegelin (1964–1965), is visibly very similar to the *Indochinesisch* family described by Gabelentz and Conrady.

Benedict (1972) is of particular interest. Although he proposed a bifurcate Sino-Tibetan model, he was inclined to question the adaptability of traditional tree models for Tibeto-Burman. In fact, he regarded Kachin (i.e. Jingpho) as being central to the family, both geographically and typologically (Handel 2008: 427). Hence, the sub-grouping of Tibeto-Burman languages is not represented in a family tree, but radiates out from Kachin. To put it in Matisoff's words (1978: 2), it represents "an interlocking network of fuzzy-edged clots of languages, emit-

<sup>5</sup> This is the view which most Chinese scholars (e.g. Xíng Gōngwǎn 1991, Sūn Hóngkǎi 1991, Pān Wùyún 2003, Zhèngzhāng Shàngfāng 2003, Ting and Sūn 2006) still hold today. The approach toward the issue occasionally remains politically motivated, and national and political preferences also figure, though not consistently. On the other side, scholars from Peking University implicitly contend that Chinese and Tibeto-Burman are not related (e.g. Guō Xīliáng 2002, 2003), mainly because of the time gap which occurs between the writing systems of the two main branches of Sino-Tibetan (Chinese and Tibeto-Burman). It is not very clear, however, whether these scholars do not differentiate between language and writing (also a possibility), or whether they claim that the time gap in the emergence of written documentation renders infeasible the correct application of the comparative method.

<sup>6</sup> This claim has been comprehensively disproved. See DeLancey (1990).

<sup>7</sup> Daic was removed from Sino-Tibetan by Benedict. See Benedict (1942).

ting waves of mutual influence from their various nuclear ganglia.” Matisoff (1978) also agreed on the fact that evidence from Tibeto-Burman languages does not support the integrity of a *Stammbaum*.<sup>8</sup> What is most striking, however, is the fact that at the end of his laborious scientific work on Sino-Tibetan (1972), Benedict seemed to favour the idea that the connections between Sinitic and Tibeto-Karen might be of a substratum and superstratum kind (Künstler 2019: 19):

It might be argued that the ST elements constitute only a superstratum in Chinese, and that the substratum is of distinct origin. In historical terms, the Chou people might be regarded as bearers of a ST language, which became fused with, or perhaps immersed in, a non-ST language spoken by the Shang people.<sup>9</sup>

Although at a first glance one may get the impression that Benedict supported the integrity of a bifurcate Sino-Tibetan family, as the title of his book suggests, at the end of his epoch-making study, Benedict (1972: 195–196) exposes four major arguments against the alleged kinship between Sinitic and Tibeto-Karen: (i) lack, from the Sinitic side, of the fairly elaborate Tibeto-Karen morphological system; (ii) presence of a restricted number of common roots; (iii) differences in the phonological system of the two branches which hardly become reconcilable at some points; (iv) non-correlation between the tonal systems of the two stocks.

When Benedict worked on his project in 1942, Karlgren’s system of Old Chinese reconstructions was the only system available to him. Although Karlgren opposed to the idea that Sinitic lacked morphology, he tried to show that proto-Chinese possessed a perspicuous system of inflection in Indo-European fashion. This type of flectional morphology, of course, was very different from the derivational morphology exhibited by many Trans-Himalayan dialects, and Benedict could not relate it to any language of his Tibeto-Karen stock. Because those languages grouped together under the Sino-Tibetan phylum lacked exactly those morphological anomalies which facilitated the studies of the pioneer Indo-Europeanists,<sup>10</sup> workers in Sino-Tibetan had to plunge into comparative philology, trying to identify cognate roots mainly on the basis of overall phonetic similarities plus semantic congruity. This was exactly what Karlgren sought to avoid when he proposed his approach toward word-families, and it is probably this fact that led many linguists (perhaps including Benedict) to think that the genetic relationship between Sinitic and Tibeto-Burman relied only on a number of common lexical roots, from which only tentative phonological generalizations could be deduced (Künstler, 2019: 18–25).

Benedict (1976), nonetheless, devoted another influential work on Sino-Tibetan after his *Conspectus* was edited and published in 1972 by Matisoff. The main findings of this *aggiornamento* are that Sino-Tibetan is a well-established family, that Daic and Hmong-Mien should remain confined outside the Sino-Tibetan family, and that lexical analysis supported the taxonomic arrangement of the *Conspectus*, which setted Sinitic apart from Tibeto-Burman, albeit the position of Karen remained indeterminate.<sup>11</sup> It is important to remark, however, that

---

<sup>8</sup> Various Sinologists such as Edwin Pulleyblank (1991: 442) have also argued that the *Stammbaum* model is inappropriate for Sinitic languages. Many dialectologists, especially from China (e.g. Zhang et al. 2018), agree on this point. Nevertheless, others have pointed out that many arguments which seek to reject the traditional tree model are grounded on a series of misconceptions towards the nature and the scopes of the *Stammbaum* model (Orlandi 2019a: 132–134). For what regards how the tree model should be utilized in historical linguistics, see Jacques and List (2019). See also Kalyan and François (2019) for a competing hypothesis.

<sup>9</sup> Benedict (1972: 197).

<sup>10</sup> Indo-European shows a wide range of “aberrant features,” such as thematic and athematic conjugations not reducible to the form of active plus affixes. Sino-Tibetan, for typological causes, cannot show this wide range of morphological markers into a string, nor can show a lucid system of apophony (so called *Ablaut*).

<sup>11</sup> However, few scholars have exposed or at least recognized the limitations of the Tibeto-Burman comparanda contained in Benedict’s *Conspectus*. In fact, Benedict’s reconstruction of Tibeto-Burman roots resembles

Benedict's view on the classification of Sino-Tibetan language was haphazard and at times self-contradictory. As he favoured the idea that proto-Sino-Tibetan was an SOV language, he was tempted to consider the SVO word-order in Sinitic and Karen a sufficient reason for treating Sino-Karen as the outgroup of the family.

In the meanwhile, contributions to the field of historical Chinese phonology had been done, most notably by Li Fang-kuei (1971) and Nicholas Bodman (1973). Benedict (1976: 172) was struck by "the numerous special lexical links between Chinese and Tibetan" which were cited by Bodman (1973), who suggested that Chinese stands in a special relationship to Tibetan. Notwithstanding Bodman's efforts to clarify the position of Sinitic within the Sino-Tibetan family, Benedict's previous Sino-Tibetan scheme practically remained unaltered. However, we must remember that the morphology of Old Chinese was still poorly explored, let alone understood, and that Benedict's argument for a bifurcation of the family was based on a tally of core vocabulary cognates (Swadesh 100-word list), which is not the most recommended practice in historical linguistics. Later discoveries of Old Chinese morphological mechanisms, in fact, have dramatically sharpened the burden between Sinitic and Tibeto-Burman languages.<sup>12</sup>

To conclude this lengthy paragraph, (i) although the alleged intellectual dichotomy between Leyden's Indo-Chinese and Klaproth's Tibeto-Burman is problematic and does not faithfully portrays the vibrant academic endeavours of nineteenth century linguistics, it is probably accurate to assert that the Sino-Tibetan phylum as understood by Chinese scholars (including Li Fang-kuei, *pace* LaPolla) and pioneers such as Shafer and Voegelin and Voegelin is a repackaging of the late Indo-Chinese orthodoxy; (ii) the various taxonomic proposals presented in this paragraph, to which one could also include the two challenging models of Starostin (Sino-Kiranti) in 1994 and Bodman (Sino-Himalayan) in 1980 which sought to defy the traditional Sino-Tibetan scheme, are a clear indicator of the debated status of the *Stammbaum* of the family; (iii) the bifurcation of Sinitic vs non-Sinitic varieties of Trans-Himalayan was probably due to the inadequateness of reconstructions of Old Chinese; (iv) bifurcate Sino-Tibetan is just one among many other models that have been proposed from nineteenth century onward.

### 3. Common Innovations of non-Sinitic Languages

While it is not possible to assess all the attempts to justify the bifurcated Sino-Tibetan model, some of the proposed common innovations of the non-Sinitic languages may equally deserve a brief discussion.

Schuessler (2007: 30–31) already observed that some Chinese words seem to correspond to "Tibeto-Burman" final \*-ʔ, whereas other words seem to correspond to TB final \*-k. Baxter

---

more a teleological "reconstructive exercises," where proto-forms are often postulated through an imaginative process of syllable canon, than a reconstruction arrived by means of the orthodox comparative method. Beckwith (2002: xiii–xiv) is one of the few scholars to have exposed this problem: "[T]he only premodern Tibeto-Burman languages cited (unsystematically) in the book are what Benedict calls 'Written Tibetan' and 'Written Burmese'. These terms are in fact equatable with whatever is contained in, respectively, Jäschke's Tibetan-English dictionary and Judson's two Burmese dictionaries, both of which, though excellent, contain many modernisms alongside numerous archaisms."

<sup>12</sup> Regarding Benedict and Matisoff's reconstructions, there has been a heated debate in *Cahier de Linguistique Asie Orientale* on word-families and reconstructions between Fellner and Hill (2019ab) and Handel (2019), Thurgood (2019) and Schuessler (2019). While the present writer does not agree with every objection raised by Fellner and Hill (2019), it is felt that their advocacy for a stricter adherence to the traditional comparative method is essentially correct.



(1992: 182) speculated about the existence of a \*-kʔ final cluster. Baxter and Sagart (2014) adduced good evidence in favour of the existence of a contrast between \*-k and \*-ʔ codas in Old Chinese. Sagart (2017) suggested that for the correspondence between Old Chinese \*-k and Tibeto-Burman (i.e., non-Sinitic) \*-k it would be natural to postulate a proto-Sino-Tibetan coda \*-k, whereas for the correspondence between Old Chinese \*-ʔ and Tibeto-Burman \*-k, a compelling hypothesis would be to reconstruct proto-Sino-Tibetan \*-q. Under this scenario, which Sagart referred to as scenario A, the merger of proto-Sino-Tibetan \*-k and \*-q codas would constitute a shared innovation uniquely observed among non-Sinitic languages. The other scenario (scenario B), however, would imply to treat the final glottal stop as a secondary development, perhaps through dialect stratification. This explanation would definitely rule out the alleged merger, provided that some doublets with an alternation between \*-k and \*-ʔ are attested in Old Chinese. Evidence for the existence of this type of doublets was adduced by Baxter (1992: 205), as Sagart himself (2017: 106) acknowledged in his study.<sup>13</sup> Nevertheless, Sagart writes that

Explanation B is simpler and we could prefer it on that ground. However evidence external to Sino-Tibetan argues that explanation A is correct. An argument that the ST and Austronesian language families are genetically related was made in Sagart (2005), supported by sound correspondences, morphological parallels and arguments from cultural history, in particular relating to agriculture.

Unfortunately, Sino-Tibeto-Austronesian (STAN) remains unproven (cf. Blust 1995; Li 1995), and I have already shown in my rebuttal of STAN (2018: 6–7) that Old Chinese \*-ʔ actually corresponds to multiple proto-Austronesian (PAN) final consonants without explained conditions, and in some cases it is even unmatched. Furthermore, it is true that certain affixes may fossilize in one language while remaining productive in another, however, under normal circumstances, external evidence should not take the place of internal evidence, all the more since Austronesian and Trans-Himalayan are not demonstrably related. In other words, I would rather be inclined to agree with Sagart that scenario B is simpler (Occam's razor) and, as such, preferable to scenario A. Note also that it has been demonstrated that uvulars are cross-linguistically disfavoured in a phonemic system without glottals also co-occurring (Sylak-Glassman 2014: 39).

For what regards other non-Sinitic innovations, it is often mentioned the merger of \*a and \*ə, proposed by Handel (2008) and discussed by Hill (2019: 29–31), in languages other than Chinese. It is true that the merger likely occurred in both Tibetan and Burmese, but it still needs to be demonstrated that it also occurred in all the non-Sinitic tongues of the Trans-Himalayan family. Also, sound changes involving vowels are quite frequent, and without further corroborative evidence from other domains (e.g., morphology, syntax) it would be premature to assume a bifurcation between Sinitic and non-Sinitic languages. Also, it would be helpful to explain why phonemic evidence should be more important than other types of evidence. For instance, DeLancey (2010, 2015, 2017) has written some articles on the study of verb agreement and verbal person marking, which is a near-neglected system in Trans-Himalayan. He spoke convincingly in favour of the existence of a central branch within Trans-Himalayan on the basis of morphological considerations. He showed that there is strong morphological

<sup>13</sup> Old Chinese reconstructions are in Baxter's system (1992): 採 \*sri(k)ʔ > 'gather' 耨 \*srjik 'reap'; 采 \*sri(k)ʔ 'colourful' 色 \*srjik 'colour'. However, an anonymous reviewer has rightly pointed out to me that, in this specific case, Baxter (1992) postulated an odd reconstruction which violates its own reconstruction system, since OC \*sr- does not yield MC \*s-, as it normally should. The semantic correlation between such pairs does not appear to be sufficiently strong.

evidence connecting Jinghpaw with both Northern Naga and Kuki-Chin (as in the classification of Voegelin and Voegelin 1965). Furthermore, according to DeLancey (2017), verbal person marking systems are lacking in Tibetic, Tani, Lolo-Burmese (and also Sinitic), but are widely diffused among other clades of the Trans-Himalayan family, most conspicuously in rGyalrongic, Qiang, Naga, Kuki-Chin, Kiranti and West Himalayan.

In a more recent study, Jacques (2019) has proposed that, contrary to common assumptions, the labial causative prefixes found in various Trans-Himalayan languages of North-eastern India are not innovations, but related to labial causative prefixes found in rGyalrongic and, possibly, also in other branches of Trans-Himalayan, including perhaps Sinitic. If this is correct, it may have further implications for what concerns the subgrouping of Trans-Himalayan languages.

Morphology plays a central role in historical linguistics, especially when attention is on homologous and not analogous features.<sup>14</sup> However, it is also true that structural features may be less informative than other types of evidence, such as lexicon, for reconstructing linguistic phylogenies, all the more since a huge typological gap exists between morphologically rich rGyalrongic and Kiranti languages on the one hand, and Lolo-Burmese and Chinese on the other hand (Arcodia and Basciano 2020), as in the case of Trans-Himalayan. Since morphology can also be dramatically lost, some authors feel that comparing morphology would be especially tricky, as it will only be informative for the morphologically rich clades of the family. Indeed, the affixes which are not attested in Chinese may either never have existed, or have been lost without leaving traces (cf. DeLancey 2014; Gong 2017; Zhang, Lai and Jacques 2019).

From the phonological side, potential cases of common innovations between Lolo-Burmese and rGyalrongic (and even Chinese) have been adduced in recent publications, such as Jacques and Michaud (2011).<sup>15</sup> In their work on proto-Naish (Na, Naxi and Laze), Jacques and Michaud (2011: 491) have demonstrated that a merger between \*ŋ- and \*ɲ- (> \*ɲ-) is likely to have occurred in that proto-language. Something similar was described in a well-known source such as Baxter (1992: 194) is thought to have happened in Old Chinese as well, as clearly pointed out also by the authors of that study (2011: 492): e.g., Old Chinese \*hn- and \*hl- > medieval Chinese \*t<sup>h</sup>- or \*ɣ-, depending on the type of syllable.

Even if lexicon is given a special status, the subgrouping of Trans-Himalayan is not likely to bifurcate between Sinitic and non-Sinitic. Zhang, Jacques and Lai (2019) have arranged a substantial list of lexical cognates between Old Chinese and rGyalrongic, including some body parts. While it is not thought that everything is correct (e.g., rime \*-eŋ, p. 84, should be dropped), some of the lexical comparanda adduced in this study, indeed, are fairly robust. Other promising cases of lexical innovations which point toward a closer relationship between Burmic and rGyalrongic are, instead, listed in Sagart et al. (2019b). Although not all the cases adduced by Sagart et al. (2019b: 26) have convinced the present writer, we may list at least four of the most convincing ones, viz. Japhug *tu-rts<sup>h</sup>ʁz* ‘lung’ and Old Burmese *ʔachut* ‘lung’; Japhug *u-ɛʁri* ‘the front side’ and Old Burmese *rhe?* ‘the front side’; Japhug *zdum* ‘cloud’ and Old Burmese *tim* ‘cloud’; Japhug *nuuqambuumbjom*, Wobzi Khroskyabs *jmbjêṃ* ‘to fly’ and Old Burmese *pjam* ‘to fly’. While the Japhug form is much longer than the Burmese one, the au-

<sup>14</sup> See, for instance, the role that morphology plays in the validation of Kiranti (Jacques 2017, but Gerber and Grollman (2018) for a contrasting view) or in the definition of rGyalrongic (J. Sun 2000).

<sup>15</sup> The following examples, however, do not necessarily reflect the opinion of the present writer. For I do not think that the pieces of evidence mentioned below are robust enough to imply a closer relationship between the languages in question, and, in any case, the authors of such proposals never claimed so. Nevertheless, these examples may indicate where further work might prove rewarding.

thors have argued that it is due to the presence of a reduplication as well as of additional prefixes. Another good piece of evidence (which, incidentally, contradicts also a cognate found in the STEDT) in favour of a Tibeto-Derung connection, adduced in the supplementary material in Sagart (2019b), is the parallel between Derung *tu<sup>31</sup> wan<sup>53</sup>* and Japhug *txjpa*. In particular, they argue that the comparison with Tibetan, Chepang and Kaman adduced in the STEDT (#471) is probably mistaken, because the Tibetan forms *wal* (Amdo) and *kha<sup>55</sup> wa<sup>53</sup>* (Derge) included do not have final \*-l (Sagart et al. 2019b: 25).

Also, although phylogenetic studies do not support the integrity of Sino-Bodic, potential cases of morphological and lexical connections between Chinese and Kiranti languages such as Limbu and Dumi were presented by van Driem (1997). More specifically, van Driem (1997: 466) has shown that the morphological alternations in the stem finals of the type \*/-k ~ -ŋ/, \*/-t ~ -n/ and \*/-p ~ -m/ exhibited by medieval Chinese verbs, preserved in the form of polyphonic readings of Chinese characters, “are precisely the type of regular morphophonological alternations manifested by the various classes of verb stem in Kiranti languages such as Limbu and Dumi”. Lexical isoglosses are also included in van Driem’s detailed study on Sino-Bodic, albeit they are less palatable for a number of motives.<sup>16</sup> However, two interesting parallels (cf. van Driem 1997), namely Old Chinese \*hju? 手 ‘hand’ vs Limbu *huk* ‘arm, hand’, and Chinese \*wan 王 ‘king’ vs Limbu *haŋ* ‘king’, may also be discussed here. The Old Chinese initial of the former is uncertain (Schuessler, 2009:178). Some authors such as Baxter (1992, 1995) and Schuessler (2009) reconstruct an initial \*h- phoneme, whereas Unger (1995: 133–135) reconstructed an Old Chinese nasal initial \*n- mainly on the basis of textual glosses and graphic alternations. Unger (1995) points out that the *Shì míng* 釋名 (200 CE), a gloss dictionary of Hàn dynasty (202 BC–220 AD), glosses this word with the character *xū* 須, which possibly had a cluster initial composed of a sibilant prefix plus a nasal phoneme, e.g. \*sn-.<sup>17</sup> In addition, the *Guǎngyùn* 廣韻 uses the graph *chǒu* 杼 for *Qièyùn* 切韻 *niǔ* 杻 (also written as 狃), as indicated by Wáng Lì (1982: 231). Unger, thus, suggested that the word for ‘hand’ may be related to that for ‘finger’, viz. *niǔ* 狃 (or 丑). Alternatively, if the Old Chinese initial really was \*n-, Schuessler (2007: 469) proposes that the word *shǒu* 手 ‘hand’ perhaps may be derived from *shōu* 收 \*nhju ‘to gather up, to collect, harvest.’ In this case the rising tone would be the endoactive morpheme: ‘hand’ < lit. ‘the thing which is doing the act of collecting.’ If this is not correct, and the original initial consonant was \*h-, then Limbu *huk* may eventually be a working hypothesis. Note, however, that in Limbu *h-* has several origins (Jacques 2017: 193–194), and that Kiranti languages have very different words for ‘arm, hand’. For what regards the word for ‘king,’ the etymology is uncertain and a number of explanations have been proposed. Difficulties in the phonology of the sister languages (e.g. Written Burmese *dbaŋ* ‘power’ vs Old Chinese \*wan 王) seem to exclude the Trans-Himalayan origin of this word. Alternatively, Schuessler (2007: 507–508) has indicated that it may be associated with an Austroasiatic homophone: Old Khmer *van* ~ *vān(n)* ‘royal palace’ (also Tai *wan* ‘palace’), which in turn is cognate to *luəŋ* ‘king’ (Tai *luan* ‘royal’). Note that Shaughnessy (1991: 197) has pointed out an example in bronze inscriptions where \*wan 王 refers to a royal place and not to the king. Other

<sup>16</sup> While certain cognate sets present regular correspondences with proto-Kiranti, other cognate sets agree more closely with the initial of the modern Limbu form (van Driem 1997: 480). To give but one example, in van Driem (1997), Old Chinese \*prōŋ 邦 ‘country’ is compared with Limbu *pa:ŋphe*: ‘village.’ The final syllable *-phe*: is unmatched. In addition, the word for ‘country’ is possibly the same etymon as \*poŋ 封 (Wáng Lì 1982: 388). Schuessler (2009: 169) has also shown that this lexeme may form an area etymon together with Lushai *puŋH* and Written Burmese *phuŋ-po*, further complicating van Driem’s purported cognancy.

<sup>17</sup> The reader should be alerted, however, on the fact that sources based on late-Han era phonology cannot be taken at face value. Indeed, it is well-known that major changes had already occurred, especially in Type B syllables.

less parsimonious theories have indicated a poorly plausible connections between \*wan and \*gwan 狂, because of certain theories which link archaic Chinese kingship to shamanism (Keightley et al. 1995: 132). In this case, the Sino-Bodic comparison may enter this controversy as a valid alternative to all other explanations. Corroborative evidence in favour of Sino-Bodic has been adduced in Hill (2011: 715). In the explicit attempt to reconcile Matisoff and Gong's presentations of the origins of Written Tibetan *o*, he pointed out that both Sinitic and Tibetan merged proto-Trans-Himalayan \*o and \*ow as *o* (although he concludes that Written Tibetan has three additional origins for *o*, viz. \*wa, \*wə, \*aw).<sup>18</sup>

To sum up, the subgrouping of Trans-Himalayan languages still remains unclear and debated. However, with the fine tuning provided by recent scholarship, it is becoming more evident that the bifurcation between Sinitic and non-Sinitic languages is scarcely supported by linguistic facts.

#### 4. Recent Bayesian Models in Trans-Himalayan Linguistics and the Validity of Bifurcate Sino-Tibetan

As argued in sections 1 and 2, although the kinship of Trans-Himalayan languages is not disputed,<sup>19</sup> the subgrouping of its components still remains debated. Recently, two important studies that employ computational approaches have been dedicated to the 'origin and spread' of Sino-Tibetan, one by Zhang et al. (2019) in *Nature*, and one by Sagart et al. (2019) in *PNAS*. In the present section special attention is given mostly (but not exclusively) to the former. In fact, although Sagart et al. (2019) call their tree Sino-Tibetan (perhaps as a face-saving practice), their study apparently comes up with a Trans-Himalayan tree, where the first split is between Sino-Sal (or Sino-Brahmaputran) and the rest of the family.

Zhang et al. (2019) applied "a particular model of probability testing known as Bayesian phylogenetic modelling" (LaPolla 2019: 1). While Bayesian modelling, a character-based method, also frequently makes use of the Swadesh 100-word list and partly resembles distant-based methods in that they are a measure of lexical agreement, this approach is different from lexicostatistics, because it does not claim that lexicon is replaced at a constant rate. Yet, it is necessary to distinguish between tools and methods. Bayesian phylogenetic models are a mathematical tool, and Bayes' rule is a "vote counting" procedure (Jones and Love 2011; Anderson 2011: 189, see below for more details). Now, there are two procedures which one may follow in order to evaluate a proposed method: one is testing its practicability (perhaps re-reproducing the experiment); the other is assessing the validity of the basic assumptions which the method requires. The second procedure is followed here.

The first objection is addressed to Zhang et al.'s (2019: 112) basic assumption, according to which, as a consequence of the fact that "language carries cultural information," "the evolution of language provides insight into prehistoric human *culture* (emphasis added)." It may be true that in a homogeneous cultural area, especially if it is characterized by the lack of historical discontinuity (is this the case?), the relationship between cultural and linguistic similarities may, as Sapir (1921: 234) said, "move along parallel lines." However, this writer is afraid that the authors of this study were driven rightly into the *cul-de-sac* which Sapir (1921: 234) warned about, in his famous though misconceived passage:

<sup>18</sup> But see Matisoff (2015) for a rebuttal.

<sup>19</sup> Among the methodologies that have been employed to "prove" the existence of Sino-Tibetan/Trans-Himalayan, we may list core vocabulary cognate counts (Benedict 1976), lexical analysis (Baxter 1995), morpho-syntactic relations (van Driem 1997), and shared innovations (Burling 1983; Thurgood 2003).

It goes without saying that the mere content of language is intimately related to culture. A society that has no knowledge of theosophy need have no name for it; aborigines that had never seen or heard of a horse were compelled to invent or borrow a word for the animal when they made his acquaintance. In the sense that the vocabulary of a language more or less faithfully reflects the culture whose purposes it serves it is perfectly true that the history of language and the history of culture move along parallel lines. But this superficial and extraneous kind of parallelism is of no real interest to the linguist except in so far as the growth or borrowing of new words incidentally throws light on the formal trends of the language. The linguistic student should never make the mistake of identifying a language with its dictionary.

Second, the entire enterprise is flawed by the *contradictio in adjecto* inherent in their approach. As Zhang et al. (2019: 115) themselves explain, in order to understand the prehistory of human populations as well as their *culture*, they collated 949 lexical root-meanings across 109 Sino-Tibetan languages from the STEDT database, “[a]ccording to the items in the Swadesh 100-word list,” which, unfortunately, seeks to develop a ‘culturally independent’ test list of universally used meanings.<sup>20</sup>

Third, it seems that Zhang et al.’s (2019: 114) understanding of some historical linguistic concepts such as ‘comparative method,’ ‘glottochronology’ and ‘language classification’ is imprecise:

In historical linguistics, the comparative method—using abundant contemporary materials and historical documents—is an approach that has widely been used to establish language relationships. Glottochronology, which is an extension of the comparative method, uses lexical data to estimate absolute divergence times. However, glottochronology has considerable limitations (such as assuming a constant rate of language evolution) and does not account for different evolutionary rates of languages owing to contact, environmental change or varied rates of substitution among different categories of words.

First, the comparative method is *not* used to establish linguistic kinship (Hoenigswald 1960: 119; Newman 1970: 39; Thomason and Kaufman 1988: 201–202). The comparative method is a multi-task enterprise. On the one hand, it aims at producing as output a phonemic inventory, which is supposed to be a more or less close approximation of the phonological system of a language *X*, by taking as input phonemically transcribed lexical items from related languages (Hoenigswald 1960: 64); on the other, it is the method by which the hypothesis of an

---

<sup>20</sup> To put in the words of Bergsland and Vogt (1962: 151, also quoted in Teeter 1963: 645), “It is difficult to see ... how the various social, cultural, and linguistically patterned factors of lexical change could be determined from short word lists designed ... to eliminate those very factors.” Another internal contradiction can also be observed in the lexicostatistical approach developed by Morris Swadesh. For instance, given that lexicostatistics per se relies on a list of culturally independent lexical items, it is of crucial importance that this list be universally valid. Hence, practitioners and supporters of lexicostatistics have been very active in revising and gradually shortening this list, at the expense of respect for the principles of statistical sampling, which by definition would imply that more trust should be placed in *longer* lists of lexical items, not shorter ones. The reality is that these aims are opposed in practice. As Teeter (1963: 642) has rightly pointed out, “the closer we get to universal validity, the fewer the items we have.” Consider, for example, the Yakhontov list, which represents a mere 35-word subset of the Swadesh list. Although I do not contest the choice of specific lexical items, it is undeniable that this list may not consist of enough items to handle statistically. It is not a case that, as Holman et al. (2008) have shown, in identifying the kinship between the various Sinitic languages, the Yakhontov-35 list produces less accurate results than the original Swadesh-100 list. As an anonymous reviewer has pointed out to the present writer, however, it should be clarified that the various lexicostatistical tests of varying length may be more or less useful depending on the chronological distance that is measured by the tests. The fact that the Yakhontov-35 list is less useful for measuring dialectal differentiation within the framework of Sinitic is due to the fact that the so-called “stable items” items on them would show few—if any—variation, in the same way longer lists would produce scarce results for measuring the distance between the most distant branches of Trans-Himalayan, since the less-stable items would likely end up being replaced.

*already presumed* genetic relationship is demonstrated. While its ancillary procedure (i.e. ‘internal reconstruction’) in the broader sense might refer to any reconstructive technique that is not comparative,<sup>21</sup> the comparative method is the “procedure whereby morphs of two or more sister (i.e. genetically related) languages are matched in order to reconstruct the ancestor language” (Hoenigswald 1960: 119). In their supplementary material, they claim that “[t]here are challenges for the application of the historical linguistic comparative methods to systematically investigating the ST languages,” because “the time depth of the ST language is far beyond all the historical accounts and linguistic documentation.” This, however, holds true for many other linguistic families, some of which do not even possess a written documentation. This fact, however, is not known to have impeded the application of the comparative method (Campbell 1994, Orlandi 2019b).

Second, why glottochronology is supposed to be an extension of the comparative method we do not learn. Glottochronology, a method which Broca developed out of Dumont d’Urville’s “lexicostatistics,” is a method for estimating the time length since two or more languages diverged from a common ancestor tongue. The employment of the comparative method, on the other hand, involves (i) the establishment of systematic phonological correspondences in words of same or related meaning; (ii) the reconstruction of phonemic systems; (iii) the establishment of grammatical (especially morphological) correspondences; (iv) the reconstruction of grammatical systems (Thomason and Kaufman 1988: 202). Since glottochronology provides no way of going beyond lexical analysis, it is hard to believe that it is an extension of the comparative method.

Furthermore, one may also point out certain grave mistakes, such as putting the location of Tibetan in Lhasa rather than in the Yarlung/Chongye valley (Hill 2019: 3; Takeuchi 2012: 4; Stein 1972). However, what is most striking is that they did not publish their underlying data, but just relied on one scholar’s cognate judgments who strongly believed the theory Zhang et al. (2019) claim to confirm. All we are presented with is a table with ST reconstructions (mostly PTB) marked for presence/absence of reflexes in daughter languages in their supplementary materials. The list does not include a number of well-known languages such as Meitei, Newar, or Lepcha. Furthermore, it would be useful to point out that not including Sinitic languages besides modern Mandarin and so called “Old Chinese” (not properly a language itself, but a set of hypothetical and reconstructed sound classes) could potentially represent a methodological shortcoming. For instance, Sagart et al. (2019a) also include other Sinitic languages, such as Guǎngzhōuhuà, Cháozhōuhuà, Xìngníng huà, Lónggǎnghuà, etc. Sinitic is a much more diversified sub-family than they assume, but just how short is the list of languages included within their Sinitic group is particularly stunning.

In addition, because not every reader may have familiarity with Bayesian phylogenetic modelling, it is necessary to briefly introduce this method. Bayesian inference is a method of statistical inference where Bayes’ theorem [ $P(H|E) = P(H) \times P(E|H)/P(E)$ ] is used to update the probability for a hypothesis as more evidence becomes available.<sup>22</sup> Bayes’ theorem describes

<sup>21</sup> In a stricter sense, by ‘internal reconstruction’ is intended the comparison of allomorphs in order to arrive at a reconstructed *invariant* form of a morpheme. The reconstruction entails sound change and may also say something about allophonic relationships, though not necessarily.

<sup>22</sup>  $H$  is the hypothesis under investigation and  $E$  is the new evidence. The symbol “|” means “given”, so that  $P(H|E)$  represent the probability that the hypothesis is true given the new evidence, i.e. ‘posterior probability’, whereas  $P(H)$  is the prior probability (i.e. the probability that the hypothesis is true based on prior evidence).  $P(E|H)$  represents the so called ‘likelihood function’.  $P(E)$  is the dominator and represents the probability of the new evidence.

the probability of an event based on prior knowledge of related conditions, and in so doing it seeks to specify the optimal way to combine old evidence with new information. Bayesian phylogeny in linguistics is a character-based method which typically involves three procedures: encoding, representation and interpretation (Heggarty 2006: 184; Campbell 2013: 474). With the exception of few applications (e.g. Eska and Ringe 2004), in most Bayesian phylogenetic approaches, lexical items are the only characters coded.

More specifically, Bayesian phylogenetic applications involve the following procedures: (i) compiling a list of basic vocabulary items (normally the Swadesh 100-word list); (ii) finding their equivalents in the languages investigated; (iii) searching for or relying on cognates already identified by linguists; (iv) convert them into binary or multi-state representations; (v) using a computer software to construct a phylogenetic tree (Campbell 2013: 476–477). Of course the software produces a myriad of trees, which need to be further assessed. Bayesian methods have experienced a tremendous progress in recent years, following the fine-tuning of estimation and probability theories (Chater et al. 2006), which prompted many scholars to adopt Bayesian approaches in several areas of science. Because Bayesian inference is thought to be able to explain all the data without invoking arbitrary explanations, and to allow uncertainties to be estimated in a natural way, Bayesian approaches are not only at the forefront of phylogenetic development, but are also enjoying a position of prominence in historical linguistics, with authors who have gone as far as claiming that they constitute the future of historical linguistics (cf. Verkerk 2017).<sup>23</sup>

Proponents of Bayesian phylogenies have been active in reassuring that these quantitative approaches should not be regarded as substitute for linguistic innovations, and that statistical models will not replace the role of historical linguists. Many historical linguists (cf. Holm 2007, Campbell 2013: 482–490) argue that, although Bayesian models are certainly more sophisticated than distance-based models such as glottochronology and lexicostatistics, they are equally subject to the limitations resulting from their near-exclusive reliance on lexical items, which nonetheless are widely regarded as being incapable of revealing a sufficient range of linguistic change.<sup>24</sup> Greenhill et al. (2017) have demonstrated that, however, structural features are generally less informative and more prone to change than basic lexicon for reconstructing linguistic phylogenies.

As argued above, Bayesian phylogenetic modelling, like any regression model, attempts to infer a tree-generating model that is consistent with the priors in the given data. This does not imply that Bayesian inference can validate a given linguistic genealogy, nor does it mean that it can measure the probability that the studied hypothesis is true. While it is not denied that they might be useful to historical linguistics, these methods, more often than not, can at best yield results that are very close to those hypothesized by historical linguists. In fact, it is not very clear how they might advance our knowledge of the linguistic families they are applied to. For successful attempts in utilizing Bayesian models simply led to a recapitulation of existing theories, whereas other less successful attempts (e.g., Foster and Toth 2003; Serva and Petroni 2008) led to results that are too absurd to be taken seriously (such as, e.g., placing Al-

<sup>23</sup> Other specialists, instead, disagree. See Eska and Ringe (2004), Donohue, Denham and Oppenheimer (2012), Pereltsvaig and Lewis (2015).

<sup>24</sup> Lexicon had a central role in the first applications of “lexicostatistics” by Dumont d’Urville (1790–1842) not because lexicon was reckoned as the most stable part of a language, but because vocabulary was all the evidence available to him. Also, the alleged scarce “borrowability” among the items of the famous Swadesh 100-word list has been criticized (Haspelmath and Tadmor 2009). However, supporters of Bayesian inference (e.g., Greenhill and Gray 2012; Scarborough 2019) in linguistics claim that ‘borrowings’ are not a major issue in Bayesian approaches.

banian together with Romance languages as one branch of Indo-European, cf. Serva and Petroni 2008). For this reason, many historical linguists are left with the impression that these Bayesian models are “reinventing the wheel” (Campbell 2013: 485). For instance, Gray et al. (2000, 2009) argue that their classification of Austronesian languages matches the comparative method well (but curiously enough, one of the most eminent experts of Austronesian linguistics, Robert Blust, who incidentally has supplied the cognancy judgments which their Bayesian experiment relied upon, has not abandoned the conventional scheme of Austronesian languages in favour of the “new” Bayesian phylogeny), but it offers few, if any, original insights which we did not already know. Even within the framework of Sino-Tibetan, the classification of languages whose placement within the family is controversial or not sufficiently understood is not clarified.<sup>25</sup> The phylogenetic study by Zhang et al. (2019) would have been much more useful, had they clarified, e.g., the subgrouping of languages such as Lepcha and Newar that are difficult to classify. As such, despite its prestige and growing diffusion, this Bayesian study has yet to deliver on its promise to advance the empirical study of ST/Trans-Himalayan historical linguistics.

With Zhang et al.’s (2019) understanding of the ‘comparative method’ as a tool which merely explores the lexicon of two or more languages, and their identification of language with vocabulary, it is not surprising that they chose to put emphasis only on lexical data. However, it is well known since the times of Lorenzo Hervás y Panduro (1735–1809) and James Burnett, Lord Monboddo (1714–1799) that genetic relationship must be proven by grammatical evidence (as Meillet and Sapir believed), especially paradigmatic morphology (i.e. the combination of morphological markers into a string). To put in Nichols’s (1996: 64) words:

The diagnostic evidence is grammatical, and it combines structural paradigmaticity (usually multiple paradigmaticity) and syntagmaticity with concrete morphological forms. The Indo-Europeanists’ intuitive feel for what was diagnostic evidence of relatedness corresponds to a computable threshold of probability of occurrence, and the main purpose of this chapter has been to give a simple rule of thumb for judging evidence of relatedness. A grouping can be regarded as established by the comparative method if and only if it rests on individual-identifying evidence.

Nevertheless, certain aspects outlined in Zhang et al. (2019) may, indeed, deserve more attention. For instance, the close relationship outlined in Zhang et al. (2019) between Tani and Digarish or Naga and Kuki Chin (but already acknowledged by DeLancey 2012 et seq.) is probably correct. On the other hand, Sagart et al. (2019a) placed Tibetan together with Derung, and Sinitic with Sal.<sup>26</sup> Potential cases of common lexical innovations are offered in the supplementary material (Sagart et al. 2019b: 24–25), and this might be a good way to encourage further works to follow this direction.

---

<sup>25</sup> Actually it was not even discussed.

<sup>26</sup> DeLancey (2015) reckons the Sal languages as being a sub-branch of the wider Central Tibeto-Burman group. The Sal sub-family is generally represented with three branches, consisting of Bodo-Garo, Konyak and Kachinic or Jingpho-Luish languages (Burling 2003: 175; Thurgood 2003: 11). Divergent opinions are also encountered. Shafer originally put Konyak and Bodo-Garo under his ‘Baric’ rubric. Bradley (1997: 20) also combines them into a single sub-branch. Matisoff (2013: 20, 2012) divides Sal languages into two main branches, Bodo-Garo and Jingpho-Konyak. This division was foreshadowed by Benedict (1972:7), who noted that Bodo-Garo, Konyak and Kachin languages shared, together with the extinct Chairel language (a Luish language), distinctive roots for ‘sun’ and ‘fire.’ Van Driem (2001: 397–398, 403) originally included Bodo-Garo, Konyak, Dhimalish and Kachin-Luic in his Brahmaputran branch. However, in his most recent publications, he presents this family as encompassing only Bodo-Garo and Konyak. See also Coupe (2012).



Nevertheless, it should also be remembered that the consensus trees provided in both studies are not “just so” stories, but only partial representations of the Trans-Himalayan *Stammbaum*. For instance, Zhang et al. (2019) obtained a posterior probability of 0.6 for the non-Sinitic group, which means that this group is only “weakly supported”.<sup>27</sup> In Sagart et al.’s (2019) consensus tree, the Sino-Sal branch has a low posterior probability (0.33, cf. Sagart et al. 2019b), albeit it remains the highest candidate. This means that either Sinitic is the outgroup or that it sides with Sal, though they are not necessarily the outgroup.

#### 4. Future directions: the reconstruction of the parent language and the broader affinities of Trans-Himalayan

As argued in sections 1 and 3, the *Stammbaum* of Trans-Himalayan is still debated. Regarding language subgrouping, the identification of shared innovations is the most recommended practice of historical linguistics. However, as LaPolla (2001: 245) rightly pointed out, very few of the proposed subgrouping schemes for ST give reasons for their outlines. A number of recent publications such as Honda (2014), Hyslop (2014) and Čašule (2014), all in the volume edited by Thomas Owen-Smith and Nathan W. Hill, nonetheless tried to sharpen the demarcation line between innovations, inheritances and borrowings. What emerges from these new endeavours, however, is that our knowledge of Trans-Himalayan historical phonology is still limited (Owen-Smith and Hill 2014: 6). In order to distinguish innovations from borrowings, the reconstruction of proto-Trans-Himalayan (or of its branches) seems to be indispensable (cf. Hyslop 2014). Previous attempts at reconstructing the ancestor language of Trans-Himalayan (though called Sino-Tibetan) include Coblin (1986), Gong (1995) and Peiros and Starostin (1996). Reconstructions of proto-Tibeto-Burman are found in Benedict (1972) and Matisoff (2003). Coblin’s work (1986), one of the first attempts at incorporating more refined OC reconstructions, unfortunately suffers from some shortcomings. First, his handlist shows a “non-historical” treatment of linguistic material (Künstler 2019: 63). In fact, we find the proto-words for ‘iron’ or ‘metal’ (Coblin 1986: 98–99), although we have no evidence that the material cultures which are often associated to Tibeto-Burman populations effectively possessed iron and/or metal in general. If they had no knowledge of iron, it is hardly credible that they had a word for it. Second, this putative handlist of Sino-Tibetan words presents methodological peculiarities. For far too many of his reconstructed linguistic units, typically consisting of long strings of reconstructed phonemes, are tolerated to merge with zero in several attested Sino-Tibetan languages without explained conditions.

Gong (1995) has worked extensively on the most well-documented languages of the family, *viz.* Written Tibetan, Written Burmese, and Old Chinese (mainly as reconstructed by Li Fang-kuei, albeit with some minor revisions). The present article shall not treat in detail Gong’s reconstructions, in accordance with its aim of discussing principle and not practice. It suffices to point out that, perhaps, reconstructing a wide family, comprising around 600 languages, just from the surface of three of them will produce, in all probability, a proto-language plagued by the Berkson’s fallacy, no matter how rigorous is the reconstruction attempted. In fact, if we reconstruct the ancestor language of a very wide and still insufficiently explored family merely on the basis of Written Tibetan, Written Burmese and Old Chinese, we inevitably introduce a bias towards what is regular and frequent in the morphemic system of these three languages, which would further tempt us to exclude all deviant features a priori. Future

<sup>27</sup> I wish to thank an anonymous reviewer for his/her thoughtful comment on this aspect.

reconstructions of the proto-language will have to take into account also other languages in addition to Tibetic, Burmic and Sinitic.

Uncertainties and difficulties with the reconstruction of proto-Trans-Himalayan are testified by a significant divergence in the typology of the proto-languages. While some scholars, most notably LaPolla (1989, 1992, 1994), have argued that verbal agreement systems are independent innovations in various “Tibeto-Burman” branches of the family, others such as van Driem (1993), DeLancey (1989, 2010) and Jacques (2012) hold the view that the proto-language was a highly synthetic tongue whose subgroups later experienced varying degrees of phonological erosion (cf. Owen-Smith and Hill 2014). While none of these typological proposals has weathered the years free of difficulties, this writer’s predilection is for the reconstruction of a synthetic proto-language. DeLancey (2014) argues that a process of creolization, following their emergence as *linguae francae*, was the cause of the simpler typology of the analytic languages of Trans-Himalayan. This makes perfectly sense also in view of the process of urbanization and state centralization which is frequently associated with languages such as Chinese. In addition, McWhorter (1998: 798, 2001) lists three typological features which distinguish creoles from other languages, *viz.* (i) little or no inflectional affixation; (ii) little or no use of tone to lexically contrast monosyllables or encode syntax; (iii) semantically regular derivational affixation. However, others (cf. Ansaldo and Matthews 2001: 315–317) have rightly pointed out that this is of little help in East and South Asian context. Other features of creole languages have been listed by Aronoff et al. (2005: 302–312, 326). Some of these include (i) little inflection, or very little derivation, largely consisting of affixation with no simultaneous morphology; (ii) absence of polymorphemic classifier constructions. Most importantly, they have pointed out that the sociolinguistic conditions of usage in a given community may vary, and these variances may give rise to a difference in the grammatical properties of the creole (Aronoff et al. 2005: 304). For instance, certain sociolinguistic factors may favour a creole continuum with mesolectal varieties that are grammatically more similar to the dominant, lexifier language (Holm 1988). Although these ideas are far from being presentable at the moment, this is something future works will inevitably have to take into account.

We can now turn to the broader affiliation of Trans-Himalayan. Since a discussion about all these proposed distant classifications of Sino-Tibetan goes far beyond the scope of the present article, only some lesser-known attempts at linking Trans-Himalayan with other families are surveyed here. These proposals are not discussed in the specific. Overall, Sino-Tibetan has been linked with Na-Dene (Sapir 1920; Bengtson 1994) and Athabaskan (Shafer 1952, 1957, 1969; Swadesh 1952, 1965), Yeniseian and Caucasian (Starostin 1982, 1984, 1989, 1991; Starostin and Ruhlen 1994; Nikolaev 1991), Austroasiatic and Hmong-Mien (Starosta 2005), Indo-European (Chang 1988; Pulleyblank 1983, 1995, 1996), Austronesian (Sagart 1994, 2005, 2011, 2016, 2017), Hmong-Mien, Kra-dai, Austroasiatic and Austronesian (Pān 2003; Zhèngzhāng 2003). This last one, known as Sino-Austric (*Huá’Ào* 華澳) or Yangtzean, is a megalomaniac and very speculative distant relationship. Although it has been only poorly and vaguely described, one hesitates to cite specific examples from this putative though unclear list of genetic factors which are supposed to purport the integrity of this megalophylum only because one is not willing to give further currency to these pseudo-linguistic speculations. However, one comes up with the impression that mere geographical proximity, together with certain political adumbrations, are at the basis of this megalophylum. Sino-Caucasian has been surveyed elsewhere (van Driem, 2005b), therefore is not discussed here. Sagart’s Sino-Tibeto-Austronesian (STAN) is perhaps the most promising among these proposals of distant genetic relationship. It is grounded on both linguistic and archaeological (especially for what regards the domestication of two types of cereals) arguments. Linguistically, it was rejected by Blust

(1995, 2014) with two devastating critiques. More recently, the theory has been re-examined by Orlandi (2018), who appears to be more open-minded to it, although he also concludes that the evidence in favour of STAN is still insufficient for proving this distant relationship, given the non-obvious relationship of the languages involved. Sino-Dene is much less popular, and has not gained much support besides that of its original proponents. Although the theory goes back to Sapir, connections between Trans-Himalayan and the languages of the Americans were already noted by the English engineer and philologist Henry Harcourt Hyde Clarke (1815–1895).<sup>28</sup> He also argued that Lepcha, a Trans-Himalayan language spoken by Lepcha people in Sikkim and parts of West Bengal, Nepal and Bhutan, was genetically related to Chadic (Houssa in Clarke’s terminology). Of course, the history of Sino-Dene is unrelated to these absurd speculations. In brief, in 1920, Sapir noted that Na-Dene was fundamentally different from the other languages spoken in North America. He wrote a series of letters to Alfred Kroeber where he enthusiastically spoke of a connection between Na-Dene and “Indo-Chinese.” In 1925, a supporting article summarizing his thoughts, albeit not written by him (cf. Bengtson 1994: 208), entitled “The Similarities of Chinese and Indian Languages,” was published in *Science Supplements*. Sapir mainly compared Na-Dene classifiers with a similar verb prefix *s-* found in Tibetan. The Sino-Dene hypothesis never gained foothold in the United States, besides Sapir’s circle, though it was later revitalized by Shafer (1952, 1957, 1969) and Swadesh (1952, 1965). Nevertheless, their Sino-Tibetan-Athabaskan connections were assailed by other scholars (e.g. Matisoff, note 13 of Benedict’s *Conspectus* 1972: 3; Campbell 1988: 593). Today very few scholars have expressed their predilection for Sino-Dene. A notable supporter is Bengtson, who also holds the view that Sino-Dene and STAN could be reconciled at sufficient time depths (pc, 2018).

### 5. Remarks on the Trans-Himalayan *Urheimat*

The present article intentionally overlooked the ongoing debate on the purported Trans-Himalayan *Urheimat*. Some remarks, however, are also recommendable. Some scholars favour autochthony and argue for a Trans-Himalayan/Sino-Tibetan *Urheimat* in present-day northern China; others favour language diversification and argue for an *Urheimat* in northern India and Nepal.<sup>29</sup> Trans-Himalayan, however, is not the sole linguistic family where we do have two conflicting theories about the placing of a homeland. Even the hypothetical homeland of a well-explored family such as Indo-European is still controversial. One theory purports that

<sup>28</sup> See Clarke (1881: 118). However, Clarke misinterpreted the kinship of his Kolarian family, which was supposed to include Sino-Tibetan languages (Kuki-Chin-Naga, Bodo), Dravidic (Konda) and Austroasiatic (Sora) languages. Clarke believed that other Himalayish languages, generally put under the Indo-Chinese phylum, were instead related to Ugric tongues. In Clarke’s view, Ugric languages encompassed Samoyedic, Hungarian (Magyar), Khanty (Ostyak), Mordvinic, Mari (Cheremis), Udmurt (Votyak) and Lapp. Clarke saw lexical evidence for a genealogical relationship between Ugric and the Himalayan languages of East Nepal, such as Bantawa (Rungchengbung), Bahing and other languages which today are classified as Kiranti. In addition, there were other languages which show a genetic relationship with Ugric, albeit they were reckoned as being more distantly related: Takpa (Tshangla) and the Qiangic language Manyak (Muya). Kiranti languages of west Nepal were also considered genetically related, though not as close as the Kiranti tongues of east Nepal. Sunwar, Gurung, Magar, Newar and Tamang (Moormi). Clarke further included Vayu, Abhor-Minyong (Abor group) and Orunodoi (Sivasagar, previously Sivasagar), but left out the so-called ‘prehistoric languages of north-east Nepal,’ such as Bodo, Kachari (both being Bodo-Garo languages) and Dhimal (Clarke 1878: 47).

<sup>29</sup> See, for instance, Blench and Post (2014).

Indo-European originated somewhere within the Steppes near the Caspian Sea around 5,000~7,000 years ago. While there is no direct archaeological evidence for their dispersal north to Scandinavia, south to the Mediterranean, west to Europe and east to India, many linguists accept this provisional *Urheimat* because, based on what is known about rates of language changes, Indo-European could not have existed prior to 7,000~8,000 years ago. Nevertheless, the British archaeologist and palaeolinguist Colin Renfrew (1990) has pointed out that evidence from climatic changes and technological developments would suggest that Indo-Europeans originated as early as 9,000 years ago in Anatolia (cf. Bouckaert et al. 2012).

The typological classification of languages is generally inaccurate and can tell us few about the genetic relationship of languages.<sup>30</sup> The identification of the *Urheimat*, nonetheless, is a typological-cultural enterprise, as we are supposed to identify the typology of the speakers of a reconstructed language with the typology of early cultures, as reconstructed by means of archaeological data (Kitson 1997: 184). In fact, the association of a hypothetical group of proto-speakers with their possible homeland is an intuitive assessment of the time covered by the internal changes reconstructible for the *Ursprache* and by those socio-archaeological cultures which they are associated with (Kitson 1997: 183–184). Many of the homeland theories related to Trans-Himalayan are strictly connected to the assumption according to which “language families arise through demographic processes driven by favourable changes in food procurement” (Sagart et al. 2019a: 10320; cf. Diamond and Bellwood 2003). This ‘language dispersal’ model is not new, as it has been emphasized by Renfrew (1973 et seq.) and Bellwood (1991 et seq.), and was pioneered a long time ago by von Heine-Geldern (1914). Renfrew associated the spread of Indo-European languages in Europe with the spread of agriculture in the Old Continent, which he interpreted by the ‘wave of advance’ model of Ammermann and Cavalli-Sforza (1973, 1979, 1984; cf. Campbell and Poser 2008: 337). A wealth of publications have demonstrated that language and agriculture may follow separate vectors, and at least for what regards many other linguistic families, the model is simply untenable (Kaufman 1976; Vansina 1995; Gardner 2000; cf. Campbell and Poser 2008: 344–346). It would be also a good practice to illustrate why the language/agriculture spread model is better than the generally accepted ‘centre of gravity’ model, based on the principle of minimum moves and maximum diversifi-

---

<sup>30</sup> Some authors, however, claim otherwise. See, for instance, Greenhill et al. (2010), Wichmann and Saunders (2007), Dediu (2011), Wichmann (2013). Some of these studies are indeed very insightful and informative about how to use typological features. However, they do not invalidate the statements that typological classification cannot tell us much about the genealogical classification of languages: Hopi is very different typologically from Aztec, Chamic from Rukai, and Manchu from any other Tungusic language. Within the framework of Trans-Himalayan, we have SVO languages such as Chinese and Kiranti, and SOV languages, highly synthetic languages such as rGyalrongic and isolating languages such as Burmese. Yet Chamic and Rukai are Austronesian, and Manchu is a Tungusic language. Japanese and Korean are typologically close to the point that, in most cases, one can literally translate morpheme by morpheme from one language to the other, and yet they are unrelated. It is clear that typology is not a reliable criterion for language classification, albeit one may effectively look at typology in a different way. However, even some of these typological studies had to conclude that the degree of confidence toward the acceptance of hypotheses concerning genetic relationships generated by algorithms on the basis of typological data is still unknown to us. Indeed, the present writer suspects that this task is at odds with the genealogical classification of languages. Postulating a genetic relationship is a very different type of ability from the fast processing toward a specific end that the new quantitative approaches seek to pursue. It is well-known that the genetic relationship among languages is heuristic in principle, and relies upon some key human abilities such as intuition and common sense. It is very difficult to devise an algorithm for these key abilities, since all that an algorithm can do is to behave in a logical way (which is quite different from ‘reasoning’). In fact, there is a neat difference between the type of fluid intelligence demanded by historical linguistics and the type of ‘crystallized’ intelligence that comes from prior learning and/or past experience which is proper to quantitative approaches, including Bayesian inference.

cation, which is widely used for assigning linguistic homelands to accepted linguistic families (Campbell and Poser 2008), or why it is preferred over other approaches which go beyond the traditional comparative method, such as the ‘punctuated equilibrium’ model (Dixon 2002), or the ‘spread zone’ approach (Nichols 1992, 1997).

Genetic studies, sometimes, seem to support these cultural and typological identifications. However, language diversity cannot be explained by genetic factors, albeit the relationship between the two is often stimulating. For example, it is well-known that, as a result of what evolutionary biologists call the ‘founder effect,’ genetic diversity is higher in Africa and diminishes with increasing distance from it. Curiously, languages in Africa exhibit the highest number of phonemes, including a peculiar type of plosive consonants produced with a velaric ingressive (velaric suction in the Ladefoged and Williamson feature system) airstream, and the farther away phonemic inventories are from Africa, the simpler they become, touching their lowest levels in the languages of the Pacific. Yet this may be only a coincidence, as several exceptions are found in a number of languages of the north-west coast of the Americas. While it is not denied the usefulness of linguistic palaeontology, a term coined by Adolphe Pictet (1859–1863) in a romantic attempt to reconstruct the world of the proto-Indo-Europeans, it should be remembered that the primary scope of language reconstruction is to test and validate an assumed genetic relationship between languages, and not to give indications about the culture or the homeland of their alleged speakers.

## 6. Conclusions

The detailed history of the Western linguistic development of the hypothesis of the Sino-Tibetan family presented in this article has demonstrated that the hypothesis of a Sino-Tibetan phylum bifurcated into Sinitic and non-Sinitic (i.e. Tibeto-Burman) has never been really substantiated, and as such many scholars may not necessarily be mistaken in claiming that it reflects an obsolete view (cf. McColl et al. 2018: 362). The more agnostic term Trans-Himalayan is the most neutral and accurate name for the family (Simon and Hill 2015: 381). Given that the recent Bayesian study by Sagart et al. (2019) has produced a Trans-Himalayan tree and that the one by Zhang et al. (2019), which purports to substantiate the bifurcate Sino-Tibetan model, contains many methodological peculiarities, it is concluded that, for the present at least, the bifurcate Sino-Tibetan model is weakened rather than enhanced by these statistical approaches. Hugely important is the fact that, even though one is persuaded of similar Bayesian approaches, the two most recent works of this kind have come up with two divergent scenarios, one of which produced a Trans-Himalayan tree.

## References

- Abel-Rémusat, Jean-Pierre. 1825. *Mélanges Asiatiques*. Paris: Librairie Orientale de Dondey-Dupré Pères et Fils.
- Ammerman, Albert J., Luigi Luca Cavalli-Sforza, Colin Renfrew. 1973. The explanation of culture change: models in prehistory. In: Colin Renfrew (ed.). *A population model for the diffusion of early farming in Europe; Models in Prehistory*: 343–357. Pittsburgh: University of Pittsburgh Press.
- Ammerman, Albert J., Luigi Luca Cavalli-Sforza. 1979. The wave of advance model for the spread of agriculture in Europe. In: Colin Renfrew, Kenneth L. Cooke (eds.). *Transformations, mathematical approaches to culture change*: 275–293. New York, NY: Academic Press.
- Ammerman, Albert J., Luigi Luca Cavalli-Sforza. 1984. *The Neolithic transition and the genetics of populations in Europe*. Princeton, NJ: Princeton University Press.

- Anonymous. 1925. The similarity of Chinese and Indian languages. *Sci. Suppl.* 62(1607): xii.
- Ansaldo, Umberto, Stephen J. Matthews. 2001. Typical creoles and simple languages: The case of Sinitic. *Linguistic typology* 5(2): 311–325.
- Arcodia, Giorgio Francesco, Bianca Basciano. 2020. Morphology in Sino-Tibetan languages. In: Aronoff, Mark (ed.). *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press. doi 10.1093/acrefore/9780199384655.013.530.
- Aronoff, Mark, Irit Meir, Wendy Sandler. 2005. The paradox of sign language morphology. *Language* 81(2): 301–344.
- Atkinson, Quentin. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097): 957–960.
- Atkinson, Quentin D., Russell D. Gray. 2005. Curious parallels and curious connections – phylogenetic thinking in biology and historical linguistics. *Systematic biology* 54(4): 513–526.
- Atkinson, Quentin D., Geoff Nicholls, David Welch, Russell D. Gray. 2005. From words to dates: water into wine, mathemagic or phylogenetic inference? *Transactions of the Philological Society* 103(2): 193–219.
- Baxter, William H. 1992. *A handbook of Old Chinese phonology*. Berlin: Walter de Gruyter.
- Baxter, William H. 1995. ‘A stronger affinity... than could have been produced by accident’: A probabilistic comparison of Old Chinese and Tibeto-Burman. *Journal of Chinese Linguistics, monograph series* 8: 1–39.
- Baxter, William H., Laurent Sagart. 1998. Word formation in old Chinese. In: Jerome L. Packard (ed.) *New approaches to Chinese word formation: morphology, phonology and the lexicon in modern and ancient Chinese*: 35–76. New York: Walter de Gruyter.
- Bayer, Theophilus Siegfried. 1730. *Museum Sinicum*. St. Petersburg.
- Beckwith, Christopher (ed.). 2002. *Medieval Tibeto-Burman languages*. Leiden, Netherlands: Brill.
- Bellwood, Peter. 1991. The Austronesian dispersal and the origin of languages. *Scientific American* 265(1): 88–93.
- Bellwood, Peter. 1994. An archaeologist’s view of language macrofamily relationships. *Oceanic linguistics* 13: 391–406.
- Bellwood, Peter. 2000. The time depth of major language families: An archaeologist’s perspective. In: Colin Renfrew, April McMahon, Robert Lawrence Trask (eds.) *Time depth in historical linguistics*. Cambridge, UK: McDonald Institute for Archaeological Research.
- Bellwood, Peter. 2001. Early agriculturalist population diasporas? Farming, languages, and genes. *Annual review of anthropology* 30(1): 181–207.
- Benedict, Paul K. 1942. Thai, Kadai, and Indonesian: a new alignment in Southeastern Asia. *American anthropologist* 44(4): 576–601.
- Benedict, Paul K. 1975. *Austro-Thai language and culture, with a glossary of roots*. New Have: Human Relations Area Files Press.
- Benedict, Paul K. 1976. Sino-Tibetan: another look. *Journal of the American Oriental Society* 96(2): 167–197.
- Benedict, Paul K., James A. Matisoff. 1972. *Sino-Tibetan: a conspectus*. Berkeley, CA: University of California Press.
- Bengtson, John D. 1994. Edward Sapir and the “Sino-Dene” hypothesis. *Anthropological Science* 102(3): 207–230.
- Bergsland, Knut, Hans Vogt. 1962. On the validity of glottochronology. *Current Anthropology* 3(2): 115–153.
- Blench, Robert, Mark Post. 2014. Rethinking Sino-Tibetan phylogeny from the perspective of Northeast Indian languages. In: Thomas Owen-Smith, Nathan W. Hill (eds.). *Trans-Himalayan linguistics: Historical and descriptive linguistics of the Himalayan area*: 71–104. Berlin: De Gruyter Mouton.
- Blust, Robert. 1995. An austronesianist look at Sino-Austronesian. *Journal of Chinese Linguistics, monograph series* 8: 283–298.
- Blust, Robert. 2014. Some recent proposals concerning the classification of the Austronesian languages. *Oceanic linguistics* 53(2): 300–391.
- Bodman, Nicholas C. 1973. Some Chinese Reflexes of Sino-Tibetan S-Clusters. *Journal of Chinese linguistics* 1(3): 383–396.
- Bodman, Nicholas C. 1980. Proto-Chinese and Sino-Tibetan: Data towards establishing the nature of the relationship. *Cornell linguistic contribution Ithaca* 3: 34–199.
- Bouckaert, Remco, Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, Quentin D. Atkinson. 2012. Mapping the origins and expansion of the Indo-European language family. *Science* 337(6097): 957–960.
- Bradley, David. 1997. Tibeto-Burman languages and classification. In: David Bradley (ed.). *Tibeto-Burman languages of the Himalayas (Papers in Southeast Asian Linguistics, 14)*: 1–72. Canberra, Australia: Department of Linguistics, The Australian National University.

- Brandreth, Edward Lyall. 1878. On the non-Aryan languages of India. *Journal of the Royal Asiatic Society* 10(1): 1–32.
- Burling, Robbins. 1983. The Sal languages. *Linguistics of the Tibeto-Burman area* 7(2): 1–32.
- Burling, Robbins. 2003. The Tibeto-Burman languages of Northeastern India. In: Graham Thurgood, Randy J. LaPolla (eds.). *The Sino-Tibetan languages*: 167–192. London / New York: Walter de Gruyter.
- Campbell, Lyle. 1988. Review article of Language in the Americas, by J. H. Greenberg. *Language* 64: 591–615.
- Campbell, Lyle. 1994. Linguistic reconstruction and unwritten languages. In: R. E. Asher, J. M. Y. Simpson (eds.). *Encyclopedia of language and linguistics*: 3475–3480. Ann Arbor: Pergamon Press.
- Campbell, Lyle, William J. Poser. 2008. *Language classification: History and method*. Cambridge University Press.
- Čašule, Ilija. 2014. Burushaski kinship terminology of Indo-European origin. In: Thomas Owen-Smith, Nathan W. Hill (eds.). *Trans-Himalayan linguistics*: 181–220. Berlin: Walter de Gruyter.
- Chater, Nick, Joshua B. Tenenbaum, Alan Yuille. 2006a. Probabilistic models of cognition: Conceptual foundations. *Trends in cognitive sciences* 10(7): 287–291.
- Chater, Nick, Christopher D. Manning. 2006b. Probabilistic models of language processing and acquisition. *Trends in cognitive sciences* 10(7): 335–344.
- Clarke, Hyde. 1878. Himalayan origin and connection of the Magyar and Ugrian. *The Journal of the Anthropological Institute of Great Britain and Ireland* 7: 44–65.
- Clarke, Hyde. 1881. Osservazioni sulla lingua onorifica ovvero cerimoniale della tribù lepcha o rông di Dorjeling, nella regione Himalaia. In: *Atti del IV Congresso Internazionale degli Orientalisti*: 117–122. Firenze, Italia: Coi Tipi dei Successori Le Monnier.
- Coblin, Weldon South. 1986. *A sinologist's handlist of Sino-Tibetan lexical comparisons*. Nettetal: Steyler Verlag.
- Conrady, August. 1896. *Eine indochinesische Causativ-Denominativ-Bildung und ihr Zusammenhang mit den Tonaccenten, ein Beitrag zur vergleichenden Grammatik der indochinesischen Sprachen, insonderheit des Tibetischen, Barmannischen, Siamesischen und Chinesischen, von Dr. August Conrady*. Wien: Otto Harrassowitz.
- Conrady, August. 1915. Der althinesische Fragesatz und der steigende Ton. *Mitteilungen des Seminars für orientalische Sprachen* 1(18): 261–297.
- Coupe, Alexander R. 2012. Overcounting numeral systems and their relevance to sub-grouping in the Tibeto-Burman languages of Nagaland. *Language and linguistics* 13(1): 193–220.
- Cust, Robert. 1881. Sulle lingue non ariane delle indie orientali. *Atti del IV Congresso Internazionale degli Orientalisti Tenutosi in Firenze 1878*: 123–128. Firenze, Italia: Coi Tipi dei Successori Le Monnier.
- Dài, Qìngxìà. 1997. Guānyú Hànzàngyǔ fēnlèi de sǐkǎo [Thoughts on subgrouping in Sino-Tibetan]. *Yǔyán jiàoxué yǔ yánjiū* 4: 4–10.
- Dài, Qìngxìà, Huáng Bùfán, Fù Àilán, Rénzēng Wàngmǔ, Liú Júhuáng (eds.). 1991. *Zàngmiǎnyǔ shíwǔ zhōng* [Fifteen Tibeto-Burman languages]. Beijing: Běijīng yànshān chūbǎn shè.
- Dediu, Dan. 2011. A Bayesian phylogenetic approach to estimating the stability of linguistic features and the genetic biasing of tone. *Proceedings of the Royal Society B: Biological Sciences* 278(1704): 474–479.
- DeLancey, Scott. 1987. The Sino-Tibetan languages. In: Bernard Comrie (ed.). *The World's major languages*: 799–810. New York: Oxford University Press.
- DeLancey, Scott. 1989. Verb agreement in Proto-Tibeto-Burman. *Bulletin of the School of Oriental and African Studies* 52(2): 315–333.
- DeLancey, Scott. 1991. Sino-Tibetan languages. In: William Bright (ed.). *The International Encyclopedia of Linguistics*: 445–449. New York: Oxford University Press.
- DeLancey, Scott. 2010. Towards a history of verb agreement in Tibeto-Burman. *Himalayan Linguistics* 9(1): 1–39.
- DeLancey, Scott. 2014. Creolization in the divergence of the Tibeto-Burman languages. In: Thomas Owen-Smith, Nathan Hill (eds.). *Trans-Himalayan linguistics*: 41–70. Berlin / Boston: Walter de Gruyter.
- DeLancey, Scott. 2015. Morphological evidence for a central branch of Trans-Himalayan (Sino-Tibetan). *Cahiers de linguistique Asie orientale* 44(2): 122–149.
- DeLancey, Scott. 2017. Hierarchical and accusative alignment of verbal person marking in Trans-Himalayan. *Journal of South Asian languages and linguistics* 4(1): 85–105.
- Diamond, Jared M., Peter Bellwood. 2003. Farmers and their languages: the first expansions. *Science* 300(5619): 597–603.
- Dixon, Robert M. W. 2002. *Australian languages: Their nature and development (Vol. 1)*. Cambridge University Press.
- Donohue, Mark, Tim Denham, Stephen Oppenheimer. 2012. New methodologies for historical linguistics?: Calibrating a lexicon-based methodology for diffusion vs. subgrouping. *Diachronica* 29: 505–522.

- van Driem, George. 1988. Reflexes of the Tibeto-Burman \*-t directive suffix in Dumi Rai. In: David Bradley, E. J. Henderson, Martine Mazaudon, Richard Keith Sprigg (eds.). *Prosodic Analysis and Asian Linguistics: To Honour R. K. Sprigg*: 157–167. Canberra: Pacific Linguistics.
- van Driem, George. 1993. The Proto-Tibeto-Burman verbal agreement system. *Bulletin of the School of Oriental and African Studies* 56(2): 292–334.
- van Driem, George. 1997. Sino-Bodic. *Bulletin of the School of Oriental and African Studies* 60(3): 455–488.
- van Driem, George. 2001. *Languages of the Himalayas: An ethnolinguistic handbook of the greater Himalayan region* (Vol. 2). Leiden: Brill.
- van Driem, George. 2003. Review of Thurgood and LaPolla 2003. *Bulletin of the School of Oriental and African Studies* 66(2): 282–284.
- van Driem, George. 2005a. Tibeto-Burman vs Indo-Chinese: implications for population geneticists, archaeologists and prehistorians. In: Roger Blench, Laurent Sagart, Alicia Sanchez-Mazas (eds.). *The Peopling of East Asia*: 109–134. London: Routledge.
- van Driem, George. 2005b. Sino-Austronesian vs. Sino-Caucasian, Sino-Bodic vs. Sino-Tibetan, and Tibeto-Burman as default theory. *Contemporary issues in Nepalese linguistics*: 285–338.
- van Driem, George. 2007. The diversity of the Tibeto-Burman language family and the linguistic ancestry of Chinese. *Bulletin of Chinese Linguistics* 1(2): 211–270.
- van Driem, George. 2011. The Trans-Himalayan phylum and its implications for population. *Communication on contemporary anthropology* 5(1): 135–142.
- van Driem, George. 2014. Trans-Himalayan. In: Thomas Owen-Smith, Nathan W. Hill (eds.). *Trans-Himalayan linguistics*: 11–40. Berlin / Boston: Walter de Gruyter.
- Eska, Joseph F., Don Ringe. 2004. Recent work in computational linguistic phylogeny. *Language* 80: 569–582.
- Fellner, Hannes A., Nathan W. Hill. 2019a. Word families, allofams, and the comparative method. *Cahiers de linguistique Asie orientale* 48(2): 91–124.
- Fellner, Hannes A., Nathan W. Hill. 2019b. The differing status of reconstruction in Trans-Himalayan and Indo-European. *Cahiers de linguistique Asie orientale* 48(2): 159–172.
- Finck, Franz Nikolaus. 1909. *Die Haupttypen des Sprachbaus*. London: Teubner.
- Forbes, Charles James F. S. 1878. On Tibeto-Burman languages. *Journal of the Royal Asiatic Society* 10(2): 210–227.
- Gabelentz, Georg von der. 1881. Sur la possibilité de prouver l'existence d'une affinité généalogique entre les langues dites Indochinoises. *Atti del IV Congresso Internazionale degli Orientalisti Tenutosi in Firenze 1878*: 283–293. Firenze, Italia: Coi Tipi dei Successori Le Monnier.
- Gardner, Peter M. 2000. *Bicultural versatility as a frontier adaptation among Paliyan foragers of South India* (Vol. 2). Lewiston, MN: Edwin Mellen Press.
- Gerber, Pascal, Selin Grollmann. 2018. What is Kiranti? A critical account. *Bulletin of Chinese Linguistics* 11(1/2): 99–152.
- Gong, Hwang-cherng. 1995. The system of finals in Proto-Sino-Tibetan. *Journal of Chinese Linguistics, monograph series* 8: 41–92.
- Gong, Xun. 2017. Jiāróng yǔzǔ yǔyán yǔ shàngǔ Hànyǔ de xíngtàixué [The rGyalrong languages and the morphology of Old Chinese]. In: Haeree Park, Shaoxuan Cheng (eds.). *Gǔwénzì yǔ hànyǔ lìshǐ yīnyùn xué* [Ancient scripts and historical Chinese phonology]: 134–156. Shanghai: Fùdàn zhōnghuá wénmíng yánjiū zhuānkān.
- Greenberg, Joseph H. 1996. In defense of Amerind. *International journal of American linguistics* 62(2): 131–164.
- Greenhill, Simon J., Quentin D. Atkinson, Andrew Meade, and Russell D. Gray 2010. The shape and tempo of language evolution. *Proceedings of the Royal Society B: Biological Sciences* 277(1693): 2443–2450.
- Greenhill, Simon J., Russell D Gray. 2012. Basic vocabulary and Bayesian phylolinguistics: Issues of understanding and representation. *Diachronica* 29(4): 523–537.
- Greenhill, Simon J., Chieh-Hsi Wu, Xia Hua, Michael Dunn, Stephen C. Levinson, Russell D. Gray. 2017. Evolutionary dynamics of language systems. *Proceedings of the National Academy of Sciences* 114(42): E8822–E8829.
- Grierson, George A. 1909. *Linguistic survey of India: Vol. III, Part I, Tibeto-Burman family: Tibetan dialects, the Himalayan dialects and the North Assam group*. Calcutta: Superintendent of Government Printing, India.
- Guō, Xiliáng. 2002. Lishǐ yīnyùn xué yánjiū zhōng de jǐ gè wèntí— —bó Méi Zǔlín zài Xiānggǎng yǔyán xuéhuì niánhuì shàng de jiǎnghuà [Some problems related to historical Chinese phonology: Refuting Mei Tsu-lin's arguments at the Hong Kong Language Society annual meeting]. *Gǔ Hànyǔ yánjiū* 3: 2–9.
- Guō, Xiliáng. 2003. Yīnyùn wèntí dá Méi Zǔlín [Problems in historical phonology, A response to Mei Tsu-lin]. *Gǔ Hànyǔ yánjiū* 3: 2–17.



- Handel, Zev. 2008. What is Sino-Tibetan? Snapshot of a field and a language family in flux. *Language and Linguistics Compass* 2(3): 422–441.
- Handel, Zev. 2019. A brief response to Fellner and Hill's "Word families, allofams, and the comparative method". *Cahiers de linguistique Asie orientale* 48(2): 125–141.
- Harlez, Charles. 1878. Les langues monosyllabiques. *Actes du dixième congrès international des orientalistes*: 67–88. Leiden: Brill.
- Heggarty, Paul. 2006. Interdisciplinary indiscipline? Can phylogenetic methods meaningfully be applied to language data – and to dating language. In: Peter Forster, Colin Renfrew (eds.). *Phylogenetic methods and the pre-history of languages*: 183–194. Cambridge: McDonald Institute for Archaeological Research.
- Heine-Geldern, Robert. 1914. *Die Bergstämme des nördlichen und nordöstlichen Birma*. Wien: Elisabeth Stiglmayr.
- Hill, Nathan W. 2011. Multiple origins of Tibetan o. *Language and linguistics* 3(12): 707–721.
- Hill, Nathan W. 2019. *The historical phonology of Tibetan, Burmese, and Chinese*. Cambridge University Press.
- Hoenigswald, Henry. 1960. *Language change and linguistic reconstruction*. Chicago University Press.
- Holm, John A. 1988. *Pidgins and creoles: theory and structure. Vol. 1*. Cambridge University Press.
- Holm, Hans J. 2007. The new arboretum of Indo-European "trees", Can new algorithms reveal the phylogeny and even prehistory of Indo-European? *Journal of quantitative linguistics* 14: 167–214.
- Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica* 42(3/4): 331–354.
- Honda, Ideo. 2014. Internal diversity in the Tamangic lexicon. In: Thomas Owen-Smith, Nathan W. Hill (eds.). *Trans-Himalayan linguistics*: 131–154. Berlin/Boston: Walter de Gruyter.
- Hyslop, Gwendolyn. 2014. A preliminary reconstruction of East Bodish. In: Thomas Owen-Smith, Nathan W. Hill (eds.) *Trans-Himalayan linguistics*: 155–180, Berlin / Boston: Walter de Gruyter.
- Jacques, Guillaume. 2012. Agreement morphology: the case of Rgyalrongic and Kiranti. *Language and Linguistics* 13: 83–116.
- Jacques, Guillaume. 2017. A reconstruction of Proto-Kiranti verb roots. *Folia Linguistica Historica* 38: 177–215.
- Jacques, Guillaume, Alexis Michaud. 2011. Approaching the historical phonology of three highly eroded Sino-Tibetan languages: Naxi, Na and Laze. *Diachronica* 28(4): 468–498.
- Jacques, Guillaume, Johann-Mattis List. 2019. Save the trees. Why we need tree models in linguistic reconstruction (and when we should apply them). *Journal of Historical Linguistics* 9(1): 128–167.
- Jīn Lǐxīn. 1998. Hànzàngyǔ de míngcí hòuzhuì -n [The Sino-Tibetan noun suffix -n]. *Mínzú Yǔwén* 1: 43–48.
- Karlgren, Bernhard. 1915–1926. *Études sur la phonologie Chinoise* [Studies on Chinese phonology]. Archives d'études orientales. Vols I-IV. Leiden / Stockholm: E. J. Brill.
- Karlgren, Bernhard. 1931. Tibetan and Chinese. *T'oung Pao* 28: 25–70.
- Karlgren, Bernhard. 1933. Word families in Chinese. *Bulletin of the Museum of Far Eastern Antiquities* 5: 5–120.
- Kaufman, Terrence. 1976. Archaeological and linguistic correlations in Mayaland and associated areas of Meso-America. *World archaeology* 8(1): 101–118.
- Kalyan, Siva, Alexandre François. 2019. When the waves meet the trees. *Journal of Historical Linguistics* 9: 167–176.
- Kitson, Peter R. 1997. Reconstruction, typology, and the "original homeland" of the Indo-Europeans. In: Jasek Fisiak (ed.). *Linguistic reconstruction and typology* 96: 183–239. Berlin: Walter de Gruyter.
- Klaproth, Heinrich Julius. 1823. *Asia polyglotta*. Berlin: A. Schubart.
- Kuhn, Ernst. 1889. *Beiträge zur Sprachenkunde Hinterindiens*. München: Verlag der K. Akademie. In Commission bei G. Franz.
- Kuhn, Thomas S. 1962. *The structure of scientific revolutions*. Chicago University Press.
- Künstler, Mieczysław Jerzy. 2019. *The Sinitic languages: A contribution to sinological linguistics*. London: Routledge.
- LaPolla, Randy J. 1989. Verb agreement, head-marking vs. dependent-marking, and the 'deconstruction' of Tibeto-Burman morpho-syntax. In: *Annual meeting of the Berkeley Linguistics Society, Vol. 15*: 356–366. Berkeley: University of California.
- LaPolla, Randy J. 1992. On the dating and nature of verb agreement in Tibeto-Burman. *Bulletin of the School of Oriental and African Studies* 55(2): 298–315.
- LaPolla, Randy J. 1994. Variable finals in Proto-Sino-Tibetan. *Bulletin of the Institute of History and Philology Academia Sinica* 65: 131–173.
- LaPolla, Randy J. 2001. The role of migration and language contact in the development of the Sino-Tibetan language family. In: Alexandra Y. Aikhenwald, Robert M. W. Dixon (eds.). *Areal diffusion and genetic inheritance: Case studies in language change*: 225–254. Oxford University Press.

- LaPolla, Randy. 2013. Subgrouping in Tibeto-Burman. In: Balthasar Bickel, Lenore Grenoble, David Peterson, Alan Timberlake (eds.). *Language typology and historical contingency*: 463–474. Amsterdam: John Benjamins.
- LaPolla, Randy J. 2016. Once again on methodology and argumentation in linguistics. *Linguistics of the Tibeto-Burman area* 39(2): 282–297.
- LaPolla, Randy J. 2019. The origin and spread of Sino-Tibetan languages. *Nature, News and Views* 1–2. doi.org/10.1038/d41586-019-01214-6.
- Leyden, John. 1808. *On the languages and literature of the Indo-Chinese nations*. London: T. Hubbard at the Hindoostanee Press.
- Li, Fang-Kuei. 1936–1937. Languages and dialects. *The Chinese year book* (1936): 121–128. Reprinted in the *Journal of Chinese Linguistics* 1(1): 1–13.
- Li, Fang-kuei. 1971. Studies on archaic Chinese phonology. *Tsinghua Journal of Chinese Studies* 9(1/2): 1–61.
- List, Johann-Mattis, Mary Walworth, Simon J. Greenhill, Tiago Tresoldi, Robert Forkel. 2018. Sequence Comparison in Computational Historical Linguistics. *Journal of Language Evolution* 12: 130–144.
- Matisoff, James A. 1973. *The grammar of Lahu*. Berkeley: University of California.
- Matisoff, James A. 1978. *Variational semantics in Tibeto-Burman: The “organic” approach to linguistic comparison*. Berkeley: University of California.
- Matisoff, James A. 1991. Sino-Tibetan linguistics: present state and future prospects. *Annual review of anthropology* 20(1): 469–504.
- Matisoff, James A. 2002. On ‘Sino-Bodic’ and other symptoms of neosubgroupitis. *Bulletin of the School of Oriental and African Studies* 63(3): 356–369.
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction*. Berkeley: University of California Press.
- Matisoff, James A. 2013. Re-examining the genetic position of Jingpho: Putting flesh on the bones of the Jingpho/Luish relationship. *Linguistics of the Tibeto-Burman area* 36(2): 15–95.
- Matisoff, James A. 2015. On the demise of the Proto-Tibeto-Burman mid vowels. *Bulletin of the National Museum of Ethnology* 39(3): 375–395.
- McColl, Hugh, Fernando Racimo, Lasse Vinner, Fabrice Demeter, Takashi Gakuhari, J. Víctor Moreno-Mayar, George Van Driem. 2018. The prehistoric peopling of Southeast Asia. *Science* 361(6397): 88–92.
- McWhorter, John H. 1998. Identifying the creole prototype: Vindicating a typological class. *Language* 74(4): 788–818.
- Mei, Tsu-lin. 1970. Tones and prosody in Middle Chinese and the origin of the rising tone. *Harvard Journal of Asiatic Studies* 30: 86–110.
- Mei, Tsu-lin. 1989. The causative and denominative functions of the \*s-prefix in Old Chinese. *Proceedings of the Second International Conference on Sinology*: 31–51. Taipei: Academia Sinica.
- Mei, Tsu-lin. 2012. The causative \*s- and nominalizing \*-s in Old Chinese and related matters in Proto-Sino-Tibetan. *Language and Linguistics* 13(1): 1–28.
- Michailovsky, Boyd. 1985. Tibeto-Burman dental suffixes: evidence from Limbu (Nepal). In: Graham Thurgood, James A. Matisoff, David Bradley (eds.) *Linguistics of the Sino-Tibetan area: the state of the art; papers presented to Paul K. Benedict for his 71st birthday*: 363–375. Canberra: Dept. of Linguistics, Research School of Pacific Studies, The Australian National University.
- Müller, Friedrich Max. 1854. *Letter to Chevalier Bunsen on the classification of the Turanian languages*. Oxford.
- Müller, Friedrich Max. 1866. *Lectures on the science of language: Delivered at the Royal Institution of Great Britain in April, May, and June 1861*. London: Longmans Green.
- Newman, Paul. 1970. Historical sound laws in Hausa and in Dera (Kanakuru). *Journal of West African languages* 1(1): 39–51.
- Nichols, Johanna. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Nichols, Johanna. 1996. The comparative method. In: Mark Durie, Malcolm Ross (eds.). *The comparative method reviewed: Regularity and irregularity in language change*: 39–71. Oxford University Press.
- Nichols, Johanna. 1997a. The epicenter of the Indo-European linguistic spread. In: Roger Blench, Roger & Matthew Spriggs (eds.). *Archaeology and language I. Theoretical and methodological orientations*: 122–148. London: Routledge.
- Nichols, Johanna. 1997b. Modeling ancient population structures and movement in linguistics. *Annual review of anthropology* 26(1): 359–384.

- Nikolaev, Sergei L. 1991. Sino-Caucasian languages in America. In: Vitaly Shevoroshkin (ed.). *Dene-Sino-Caucasian languages: Materials from the first international interdisciplinary symposium on language and prehistory*, Ann Arbor, 8.–12. November 1988: 67–129. Bochum: Brockmeyer.
- Orlandi, Georg. 2018. Evaluating the Sino-Tibeto-Austronesian hypothesis. *Journal of Language Relationship* 16(1): 1–18.
- Orlandi, Georg. 2019a. On the classification of the Ng Yap dialects: some thoughts on the subgrouping of Sinitic languages. *Journal of Language Relationship* 17(1): 128–152.
- Orlandi, Georg. 2019b. Yǔyán chóngjiàn bùzhǐ yīlài shūmiànyǔ (biǎo yīn wénzì) [Language reconstruction does not rely on written languages (phonographic scripts)]. *Yǔyán kēxué* 18(3): 225–235.
- Orlandi, Georg. 2020. The state of the art of the genetic relationship of Japonic: the Turanian and Altaic Hypotheses. *International Journal of Eurasian Linguistics* 2(1): 29–69.
- Owen-Smith, Thomas, Nathan W. Hill (eds.). 2013. *Trans-Himalayan linguistics: Historical and descriptive linguistics of the Himalayan area* (Vol. 266). Boston / Berlin: Walter de Gruyter.
- Pān, Wùyún. 2000. *Hànyǔ lìshǐ yīnyùnxué* [Chinese historical phonology]. Shanghai: Shànghǎi jiàoyù chūbǎnshè.
- Peiros, Iliá, Sergei A. Starostin. 1996. *A comparative vocabulary of five Sino-Tibetan languages*. 6 vols. Parkville: Univ. of Melbourne, Dept. of Linguistics and Applied Linguistics.
- Pictet, Adolphe. 1859–1863. *Les origines indo-européennes ou, Les Aryas primitifs: essai de paleontologie linguistique / par Adolphe Pictet*. Paris: Joel Cherbuliez.
- Pulleyblank, Edwin G. 1983. The Chinese and their neighbors in prehistoric and early historic times. In: David N. Keightley (ed.). *The origins of Chinese civilization*: 411–466. Berkeley: University of California Press.
- Pulleyblank, Edwin G. 1991. Chinese dialect studies. *Journal of Chinese Linguistics, monograph series* 3: 429–453.
- Pulleyblank, Edwin G. 1995. The historical and prehistorical relationships of Chinese. *Journal of Chinese Linguistics, monograph series* 8: 145–194.
- Pulleyblank, Edwin G. 1996. Early contacts between Indo-Europeans and Chinese. *International Review of Chinese Linguistics* 1(1): 1–25.
- Renfrew, Colin. 1973. Problems in the general correlation of archaeological and linguistic strata in prehistoric Greece: the model of autochthonous origin. In: R. A. Crossland, Ann Birchall (eds.). *Bronze Age migrations in the Aegean: archaeological and linguistic problems in Greek prehistory*: 263–276. London: Duckworth.
- Renfrew, Colin. 1990 [1987]. *Archaeology and language: the puzzle of Indo-European origins*. Cambridge University Press.
- Renfrew, Colin. 1991. Before Babel: speculations on the origins of linguistic diversity. *Cambridge Archaeological Journal* 1(1): 3–23.
- Renfrew, Colin. 1994. World linguistic diversity. *Scientific American* 270(1): 116–123.
- Renfrew, Colin. 2003. World linguistic diversity and farming dispersals. In: Roger Blench, Matthew Spriggs (eds.). *Archaeology and language I: theoretical and methodological orientations*: 104–112. London: Routledge.
- Renfrew, Colin. 2000. At the edge of knowability: towards a prehistory of languages. *Cambridge Archaeological Journal* 10(1): 7–34.
- Renfrew, Colin. 2002. The emerging synthesis: the archaeogenetics of farming/language dispersals and other spread zones. In: Peter Bellwood, Colin Renfrew (eds.). *Examining the farming/language dispersal hypothesis*: 3–16. Cambridge: McDonald Institute for Archaeological Research.
- Ryder, Robin. 2009. *Phylogenetic models of language diversification*. Ph.D. dissertation. Oxford University. Online at: <https://tel.archives-ouvertes.fr/tel-00661866>.
- Sagart, Laurent. 1993. New views on Old Chinese phonology. *Diachronica* 10(2): 237–260.
- Sagart, Laurent. 1994. Proto-Austronesian and Old Chinese evidence for Sino-Austronesian. *Oceanic linguistics* 33(2): 271–308.
- Sagart, Laurent. 1999. *The roots of old Chinese*. Amsterdam: John Benjamins.
- Sagart, Laurent. 2005. Sino-Tibetan-Austronesian: An updated and improved argument. In: Laurent Sagart, Roger Blench, Alicia Sanchez-Mazas (eds.). *The peopling of East Asia: Putting together archaeology, linguistics and genetics*: 161–176. London: Routledge.
- Sagart, Laurent. 2011. The homeland of Sino-Tibetan-Austronesian: where and when. *Communication on contemporary anthropology* 5: 143–147.
- Sagart, Laurent. 2016. The wider connections of Austronesian. *Diachronica* 33(2): 255–281.
- Sagart, Laurent. 2017. A candidate for a Tibeto-Burman innovation. *Cahiers de linguistique Asie orientale* 46(1): 101–119.
- Sagart, Laurent, Roger Blench, Alicia Sanchez-Mazas (eds.). 2005. *The peopling of East Asia: Putting together archaeology, linguistics and genetics*. London: Routledge.

- Sagart, Laurent, Tze-Fu Hsu, Yuan-Ching Tsai, Yue-Te C. Hsing. 2017. Austronesian and Chinese words for the millets. *Language dynamics and change* 7(2): 187–209.
- Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin J. Ryder, Valentin Thouzeau, Simon J. Greenhill, Johann-Mattis List. 2019a. Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences* 116(21): 10317–10322.
- Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin J. Ryder, Valentin Thouzeau, Simon J. Greenhill, Johann-Mattis List. 2019b. Supplementary information for the Paper ‘Dated language phylogenies shed light on the ancestry of Sino-Tibetan languages’. *Proceedings of the National Academy of Sciences* 116(21): 1–38.
- Sapir, Edward. 1920. *Comparative Sino-Tibetan and Na-Dene dictionary*. Ms. American Philosophical Society 20a 3.
- Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace.
- Saxena, Anju (ed.). 2004. *Himalayan languages: past and present*. Berlin / New York: Walter de Gruyter.
- Schuessler, Axel. 2007. *ABC etymological dictionary of Old Chinese*. Honolulu: University of Hawaii Press.
- Schuessler, Axel. 2019. Comments on Fellner and Hill: “Word families, allofams, and the comparative method”. *Cahiers de linguistique Asie orientale* 48(2): 142–153.
- Shafer, Robert. 1952. Athapaskan and Sino-Tibetan. *International journal of American linguistics* 18(1): 12–19.
- Shafer, Robert. 1955. Classification of the Sino-Tibetan languages. *Word* 11(1): 94–111.
- Shafer, Robert. 1957. Note on Athapaskan and Sino-Tibetan. *International Journal of American Linguistics* 23(2): 116–117.
- Shafer, Robert. 1965. The Eurasian linguistic superfamily. *Anthropos* 1(6): 445–468.
- Shafer, Robert. 1966–1967. *Introduction to Sino-Tibetan. Vol. 1*. Wiesbaden: Otto Harrassowitz.
- Shafer, Robert. 1969. A few more Athapaskan and Sino-Tibetan comparisons. *International Journal of American Linguistics* 35(1): 67.
- Simon, Camille, Nathan Wayne Hill. 2015. Tibetan. In: Nicola Grandi, Livia Körtvélyessy (eds.). *Edinburgh handbook of evaluative morphology*: 381–388. Edinburgh University Press.
- Simon, Walter. 1929. Tibetisch-chinesische Wortgleichungen, ein Versuch. *Mitteilungen des Seminars für Orientalische Sprachen an der Friedrich-Wilhelms-Universität zu Berlin* XXXII 1: 157–228.
- Starosta, Stanley. 2005. Proto-East Asian and the origin and dispersal of the languages of East and Southeast Asia and the Pacific. In: Laurent Sagart, Roger Blench, Alicia Sanchez-Mazas (eds.). *The peopling of East Asia*: 210–226. London: Routledge.
- Starostin, Sergei A. 1982. Praenisejskaja rekonstrukcija i vneshnije sv’azi enisejskix jazykov [Proto-Yeniseian reconstruction and the external relations of the Yenisei languages]. *Ketskij sbornik*: 144–237. Leningrad: Nauka.
- Starostin, Sergej A. 1984. Gipoteza o geneticheskix sv’az’ax sinotibetskix jazykov s enisejskimi i severnokavkazskimi jazykami [Hypothesis on the genetic relationship of Sino-Tibetan languages with the Yeniseian and North Caucasian languages]. In: *Lingvisticheskaja rekonstrukcija i drevnejshaja istorija Vostoka* [Linguistic reconstruction and the ancient history of the East]: 19–38. Moskva: Nauka.
- Starostin, Sergei A. 1989. Nostratic and Sino-Caucasian. In: Vitaly Shevoroshkin (ed.). *Explorations in language macrofamilies*: 42–66. Bochum: Brockmeyer.
- Starostin, Sergei A., Merritt Ruhlen. 1994. Proto-Yeniseian Reconstructions, with Extra-Yeniseian Comparisons. In: Merritt Ruhlen (ed.). *On the origin of languages: Studies in linguistic taxonomy*: 70–92. Stanford, CA: Stanford University Press.
- Sūn Hóngkǎi. 2001. Shì lùn Zhōngguó jìngnèi Zàngmiǎn yǔ de pǔxì fēnlèi [A classification of Tibeto-Burman languages in China]. In: Tatsuo Nishida (ed.). *Languages and history in East Asia: A festschrift for Tatsuo Nishida on the occasion of his 60th Birthday, Vol. I*: 61–73. Kyoto: Shokado.
- Sūn Hóngkǎi, Chénglóng Huáng, Máocǎo Zhōu. 2001. *Róuruòyǔ yánjiū* [A study of Rouruo]. Beijing, China: Zhongyang Mínzú Dàxué Chubanshè.
- Sun, Jackson Tien-Shin. 1993. The linguistic position of Tani (Mirish) in Tibeto-Burman: A lexical assessment. *Linguistics of the Tibeto-Burman Area* 16(2): 143–188.
- Sun, Jackson Tien-Shin. 2000. Parallelisms in the verb morphology of Sidaba rGyalrong and Lavrung in rGyalrongic. *Language and linguistics* 1(1): 161–190.
- Swadesh, Morris. 1952. Review article of ‘Athapaskan and Sino-Tibetan’, by R. Shafer. *International journal of American linguistics* 18: 178–181.
- Swadesh, Morris. 1965. Lingvisticheskie sv’azi Ameriki i Evrazii [Linguistic ties of America and Eurasia]. *Etimologija* 1964 [Etymology 1964]: 271–322. Moskva: Nauka.

- Takashima, Kenichi. 1996. Language and palaeography. In: Gary F. Arbuckle (ed.). *Studies in early Chinese civilization: Religion, society, language and palaeography*. Vol. 1: Text. Vol. 2: Tables and notes. Kansai Gaidai University Publication.
- Takeuchi, Tsuguhito. 2012. Formation and transformation of Old Tibetan. In: Tsuguhito Takeuchi, Norihiko Hayaishi (eds.). *Historical development of the Tibetan languages. Proceedings of the workshop B of the 17th Himalayan languages symposium, Kobe, 6th–9th September*: 3–17. Kobe: Research Institute of Foreign Studies.
- Teeter, Karl V. 1963. Lexicostatistics and Genetic Relationship. *Language* 39(4): 638–648.
- Thomason, Sarah Grey, Terrence Kaufman. 1988. *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Thurgood, Graham. 2019. A review of Fellner and Hill’s “Word families, allofams, and the comparative method”. *Cahiers de linguistique Asie orientale* 48(2): 154–158.
- Thurgood, Graham, Randy J. LaPolla (eds.). 2003. *The Sino-Tibetan languages (Vol. 3)*. London: Routledge.
- Ting Pang-Hsin, Hóngkai Sun. 2000. *Hàn-Zàngyǔ yánjiū de lìshǐ huígù* [Retrospective history of Sino-Tibetan studies]. (*Hàn-Zàngyǔ tóngyǔáncí yánjiū* [Cognate words in Sino-Tibetan languages], 1). Nanning: Guǎngxī Mínzú Chubanshè.
- Ting Pang-Hsin, Hóngkai Sun. 2001. *Hàn-Zàng, Miáo-Yáo tóngyǔáncí zhuāntí yánjiū* [Special topics in Sino-Tibetan and Hmong-Mien comparison]. (*Hàn-Zàngyǔ tóngyǔáncí yánjiū* [Cognate words in Sino-Tibetan languages], 2). Nanning: Guǎngxī Mínzú Chubanshè.
- Trombetti, Alfredo. 1905. *L’unità d’origine del linguaggio*. Bologna: Libreria Treves di Luigi Beltrami.
- Turin, Mark. 2007. *Linguistic diversity and the preservation of endangered languages: A case study from Nepal*. Kathmandu: International Centre for Integrated Mountain Development (ICIMOD).
- Vansina, Jan. 1995. New linguistic evidence and ‘the Bantu expansion’. *The Journal of African history* 36(2): 173–195.
- Verkerk, Annemarie. 2017. Phylogenies. *Language dynamics and change* 7(1): 127–140.
- Voegelin, Charles Frederick, Florence Marie Voegelin. 1964–1965. *Languages of the World: Sino-Tibetan fascicle one-five*. Ms. Archives of languages of the world, Anthropology department, Indiana University.
- Wichmann, Søren, Arpiar Saunders. 2007. How to Use Typological Databases in Historical Linguistic Research. *Diachronica* 24(2): 373–404.
- Wichmann, Søren. 2013. Diachronic Stability and Typology. In: Claire Bower, Bethwyn Evans (eds.). *The Routledge Handbook of Historical Linguistics*: 212–224. London: Routledge.
- Xíng, Gōngwǎn. 1991a. Guānyú Hànyǔ Nándǎoyǔ de fāshēng xué guānxì wèntí — L· Shājiā’ěr “Hànyǔ Nándǎoyǔ tóngyuán lùn” shùpíng bǔzhèng [On the genetic relationship of Chinese and Austronesian — Review and corrections of L. Sagart’s ‘Sino-Austronesian theory’]. *Mínzú yǔwén* 3: 1–14.
- Xíng, Gōngwǎn. 1991b. Hànyǔ Nándǎoyǔ shēngmǔ de duìyìng — L· Shā jiā’ěr “hànyǔ nándǎo yǔ tóng yuán lùn” shùpíng bǔ zhèng [Sound correspondences in the initials between Chinese and Austronesian — Review and corrections of L. Sagart’s ‘Sino-Austronesian theory’]. *Mínzú yǔwén* 4: 23–35.
- Xíng, Gōngwǎn. 1991c. Hànyǔ nándǎo yǔ shēngmǔ jí yùnwěi fǔyīn de duìyìng — L· Shā jiā’ěr “hànyǔ nándǎo yǔ tóng yuán lùn” shùpíng bǔzhèng [Sound correspondences in the initials and final consonants between Chinese and Austronesian — Review and corrections of L. Sagart’s ‘Sino-Austronesian theory’]. *Mínzú yǔwén* 5: 13–25.
- Zhang, Menghan, Wu-Yun Pan, Shi Yan, Li Jin. 2018. Phonemic evidence reveals interwoven evolution of Chinese dialects. *arXiv preprint arXiv:1802.05820*.
- Zhang, Menghan, Shi Yan, Wuyun Pan, Li Jin. 2019. Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature* 569(7754): 112–115 doi.org/10.1038/s41586-019-1153-z.
- Zhang, Shuya, Guillaume Jacques, Yunfan Lai. 2019. A study of cognates between Gyalrong and Old Chinese. *Journal of Language Relationship* 17(1): 73–92.
- Zhèngzhāng Shàngfāng. 2003. *Shànggǔyīnxì* [Old Chinese Phonology]. Shanghai: Shànghǎi jiàoyù chūbǎnshè.

Георг Орланди. Еще раз к вопросу об истории и степени обоснованности бинарной модели классификации сино-тибетских языков

В рамках сино-тибетской гипотезы, как правило, постулируется бинарное разделение всех языков семьи на китайскую (синитическую) и тибето-бирманскую ветви. В данной статье кратко, но исчерпывающе излагается общая история гипотезы, за которой сле-

дует критическое обсуждение методологических недостатков недавно предложенной байесовской классификации, якобы подтверждающей достоверность такого бинарного разделения. В завершение статьи обсуждаются возможные пути дальнейшего развития сравнительного изучения трансгималайских языков.

*Ключевые слова:* трансгималайские языки; сино-тибетские языки; байесовские классификации; филогенетическая классификация языков.

## Устойчивость и частотность: есть ли корреляция?

Значение слова может быть названо базисным, если для носителей легче вспомнить уже существующее в языке слово с этим значением, чем заменять его заимствованием или производным. Устройство человеческого мозга таково, что проще вспоминается то, что встречается чаще (поскольку задействованные в этом нейронные связи усиливаются в процессе использования, согласно правилу Хебба). Эта гипотеза была проверена на 15 частотных словарях языков разных семей и ареалов. Двадцать одно сводешевское значение выражается словом из первой тысячи по частотности во всех рассмотренных языках; больше 60 слов почти в любом из рассмотренных языков входят в первые две тысячи, более 80 – в первые три.

Если слово оказывается высокочастотным в одном из языков, велика вероятность, что оно окажется высокочастотным и в другом. Исключения связаны с полисемией: многозначные слова частотнее однозначных. Поскольку полисемия может быть унаследована из языка-предка, представительный универсальный список значений, частотность которых не зависела бы от полисемии, составить невозможно. Частотность до некоторой степени зависит от образа жизни, тем не менее, слова из сводешевского списка сохраняют высокую частотность при любом жизненном укладе. При этом изменения частотности отдельных слов не связаны ни с общим происхождением, ни с ареальным влиянием.

Наблюдения над частотностью позволяют предложить следующие улучшения методики сбора стословников: для предельных глаголов выбирать форму перфектива, для непредельных – имперфектива, для глаголов движения – форму однократного направленного действия, для глаголов восприятия – форму ненамеренного восприятия. Для слова 'рука' лучше брать не анатомические, а функциональные контексты.

Степень устойчивости слова с определённым значением различается в разных языках, тем не менее, разные выборки базисной лексики демонстрируют примерно одинаковую связь с частотностью. Хотя прямой зависимости между частотностью лексики и её устойчивостью нет, всё же соотношение между базисностью лексики и её частотностью является неслучайным. Существует связь между частотностью (но не стабильностью) и ранним освоением слов. Вероятно, это следствие того, что частотность в большей степени связана с нынешним образом жизни, тогда как стабильность в большей степени отражает прошлые эпохи.

Частотность слов с одним и тем же значением – разная в разных языках и даже в одном языке в разное время. И это, видимо, универсальный закон: если бы было иначе, слова из списка не могли бы выпадать (потому что их бы часто употребляли, не забывали и не заменяли). Сам же список базисных значений – каков бы он ни был – имеет вероятностную природу: какие именно значения войдут в первые тысячи по частотности в данном конкретном языке в заданный момент времени – непредсказуемо, но в каждый момент в любом языке большая часть стословника в них входит.

*Ключевые слова:* базисная лексика; 100-словный список Сводеша; историческая устойчивость; частотность; усвоение языка.

Кажется логичным, что вхождение того или иного значения в список базисной лексики – т. е. лексики особо устойчивой, мало подверженной заменам и заимствованиям – основывается на том, что носителям языка оказывается проще вспомнить уже существующее слово, чем заменять его заимствованием или новообразованием. В силу устрой-

ства нашего мозга, то, что чаще встречается, вспоминается лучше, поскольку соответствующие нейронные контуры оказываются усилены (Hebb 1949).

Идея того, что более частотные слова должны лучше наследоваться, проводится в работе (Арапов, Херц 1974): М. В. Арапов и М. М. Херц предлагают не составлять списки базисной лексики, а использовать весь словарь языка, ранжировав слова на основе их частотности (Арапов, Херц 1974: 30–33). Список Сводеша авторы критикуют за то, что он, по их мнению, является не лингвистическим, а этнографическим (Арапов, Херц 1974: 26). Однако, как показал М. Н. Саенко (Саенко 2014: 50–53), слова сводешевского списка являются в большинстве своём достаточно частотными: в чешском языке (по словарю Jelínek et al. 1961) «64 слова вошли в первую тысячу, 78 в первые две тысячи самых частотных слов» (Саенко 2014: 51), в польском (по словарю Kurcz et al. 1990) «51 слово из стословника входит в первую тысячу, 76 в первые две тысячи наиболее частотных слов» (там же: 52), в русском (по словарю (Brown 2003), составленному на основе словаря (Частотный... 1977) объёмом 40 тысяч слов) «68 слов из стословника вошли в первую тысячу, 84 в первые две тысячи самых частотных слов русского языка» (Саенко 2014: 53).

Для проверки этой гипотезы за пределами славянской группы были проведены подсчёты по частотным словарям семи языков разных семей и ареалов: русского (Ляшевская, Шаров 2009), английского (Davies, Gardner 2010), турецкого (Aksan Y. et al. 2017), японского (Tono et al. 2013), арабского (Buckwalter, Parkinson 2011), грузинского (Шмальцель, Нозадзе 2012) и игбо (Anon. 2015–2017). Впоследствии в подсчёт были добавлены также данные частотных словарей китайского (Xiao et al. 2009), испанского (Davies, Hayward Dawies 2018), португальского (Davies, Preto-Bay 2008), французского (Lonsdale, Le Bras 2009), немецкого (Jones, Tschirner 2006), нидерландского (Tiberius, Schoonheim 2014), чешского (Čermák, Křen 2011) и русского (Sharoff et al. 2013) языков, вышедшие в той же серии и составленные по той же методике, что и словари английского, турецкого, японского и арабского языков, а также данные по польскому языку (Kurcz et al. 1990, цит. по: Саенко 2014: 51–52).

Словари английского, нидерландского, немецкого, французского, испанского, португальского, чешского, польского, турецкого, японского, китайского и арабского охватывают первые 5 000 наиболее частотных слов соответствующего языка, словарь грузинского – первые 3525 слов, словарь игбо – первую тысячу слов.

В нидерландском словаре Tiberius, Schoonheim 2014 частоты приведены отдельно по разным жанрам и группам жанров, общего списка, упорядоченного по частотности, нет, поэтому подсчёты по нему отличаются некоторой долей условности.

Сопоставление данных русского языка по словарям Ляшевская, Шаров 2009 и Sharoff et al. 2013 показывает, что, хотя корпуса и разные (в Ляшевская, Шаров 2009 использован Национальный корпус русского языка (НКРЯ), у Sharoff et al. 2013 – тексты с более, чем 75 тысяч веб-страниц, охватывающие 150 миллионов словоупотреблений), корреляция достаточно высока: почти все слова попадают в ту же полутысячу по частотности. Однако по сравнению со словарём Brown 2003 картина в Ляшевская, Шаров 2009 выглядит несколько более пессимистичной: в первую тысячу по частотности попали не 68, а лишь 63 слова из стословника, в первые две – не 84, а 81, что говорит о некоторой зависимости рангов от выборки.

Сходная ситуация с чешским языком: согласно словарю Jelínek, Večka, Těšitelová 1961, в первую тысячу попадают 64 слова, а согласно Čermák, Křen 2011 – лишь 57, но при этом в первые две тысячи в более позднем словаре вошли не 78, а 80 слов. Для большинства слов ранг различается не очень значительно. Самое большое расхождение – у слова *tamten* ‘тог’: в словаре Jelínek, Večka, Těšitelová 1961 он имеет ранг 8785 (крайне нетипич-



ный для слова с таким значением – в большинстве языков оно входит в первую тысячу), а в словаре Šermák, Křen 2011 – 1730.

В одной из работ С. А. Старостина показано, что частотность корневой морфемы в языке в некоторый момент времени не зависит от характера выбираемого текста (Старостин 1989: 18). Со словами это очевидным образом не так. Частотные словари основываются на корпусах письменных текстов, тогда как передача языка в череде поколений осуществляется прежде всего за счёт устной речи. Таким образом, наиболее информативен был бы частотный словарь, составленный на корпусе бытовых разговоров, но таких словарей в настоящее время не существует. Тем не менее, даже имеющиеся материалы позволяют, как кажется, сделать некоторые интересные наблюдения.

Выяснилось, что между рангами слов с одинаковым значением в разных языках обнаруживается достаточно высокая корреляция: если слово входит в первую тысячу наиболее частотных слов в одном языке, велика вероятность, что слово с этим значением войдёт в первую тысячу и в другом языке (см. Табл. 1). Если же слово в одном из языков является достаточно редким (не входит в первые три тысячи по частотности), то велика вероятность, что так же будет и в других языках. Из этого есть ряд объяснимых исключений: если слово совмещает в себе несколько значений, его частотность будет выше, чем у слова, имеющего лишь одно из соответствующих значений. Так, в чешском языке слово *těšíc* совмещает значения ‘луна’ и ‘(календарный) месяц’, в японском слово 羽 *hane* – ‘перо’ и ‘крыло’, в грузинском ზობო *p'iri* – это и ‘рот’, и ‘человек’ (ср. совр. рус. разг. *в одно было* ‘в одиночку’), в арабском صغير *saġīr* – это и ‘маленький’, и ‘молодой’, в русском слово *язык* – это и ‘tongue’, и ‘language’, а *рука* – и ‘hand’, и ‘arm’.

Из-за полисемии будет различаться частотность в разных семьях, унаследовавших от своих праязыков разные совмещения значений. Универсальный список – такой, чтобы в нём было представительное количество базисной лексики, но при этом ни в каком языке не было бы влияющей на частотность полисемии, составить, видимо, нельзя.

Интересно, что разные списки базисной лексики демонстрируют примерно одинаковую соотносённость с частотностью. В рассмотренных языках среди 38 слов, входящих в список «Лейпциг – Джакарта» (Haspelmath, Tadmor 2009: 67), но не входящих в список Сводеша, от 11 до 22 входят в первую тысячу по частотности, среди слов, входящих только в список Сводеша, но отсутствующих в списке «Лейпциг – Джакарта», таких от 19 до 29 (грузинский язык из этого подсчёта исключён, поскольку данный словарь составлен по выборке текстов, из которых большинство составляют статьи законов, и выборка тем самым оказывается смещённой; так что лишь 35 слов из стословника входят в грузинском частотном словаре в первую тысячу по частотности). При этом во всех без исключения языках (даже в грузинском!) число слов, входящих в первую тысячу по частотности, в списке «Лейпциг – Джакарта» несколько меньше, чем в списке Сводеша. Напротив, слов низкочастотных (т. е. таких, которые не менее, чем в двух из рассмотренных языков не входят в первые три тысячи) в списке «Лейпциг – Джакарта» несколько больше, чем в списке Сводеша.

Если рассматривать не отдельные слова, а списки в целом, можно отметить, что есть немало значений (21, что составляет чуть больше одной пятой списка), которые во всех рассмотренных языках выражаются словами, входящими в первую тысячу по частотности. В предварительном варианте подсчёта, проведённом на меньшем количестве языков (были задействованы только русский, английский, турецкий, японский, арабский, грузинский и игбо), таких слов было 22. Это слова: ‘go’, ‘good’, ‘head’, ‘I’, ‘know’, ‘man’, ‘many’, ‘name’, ‘new’, ‘person’, ‘road’, ‘say’, ‘small’, ‘this’, ‘thou’, ‘two’, ‘water’, ‘we’, ‘what’, ‘who’, ‘woman’. Возможно, в этот же список можно включить слово ‘not’: из рассмотрен-

не входят в первую тысячу слов среди входящих в первую тысячу в языке:	рус.	англ.	тур.	яп.	араб.	груз.	кит.	итбо
рус. (63 слова)	11 (все входят во вторую тысячу)	4 (все входят во вторую тысячу)	10 (1 из них не входит во вторую тысячу)	10 (3 из них не входят во вторую тысячу)	27 (12 из них не входят во вторую тысячу)	20 (9 из них не входят во вторую тысячу)	12	
англ. (57 слов)	5 (все входят во вторую тысячу)	1 (входит во вторую тысячу)	9 (1 из них не входит во вторую тысячу)	10 (3 из них не входят во вторую тысячу)	21 (6 из них не входят во вторую тысячу)	14 (5 из них не входят во вторую тысячу)	10	
тур. (71 слово)	12 (4 из них не входят во вторую тысячу)	13 (1 из них не входит во вторую тысячу)	17 (4 из них не входят во вторую тысячу)	22 (8 из них не входят во вторую тысячу)	36 (18 из них не входят во вторую тысячу)	23 (8 из них не входят во вторую тысячу)	15	
яп. (57 слов)	10 (2 из них не входят во вторую тысячу)	10 (1 из них не входит во вторую тысячу)	4 (1 из них не входит во вторую тысячу)	16 (5 из них не входят во вторую тысячу)	27 (13 из них не входят во вторую тысячу)	14 (6 из них не входят во вторую тысячу)	9	
араб. (52 слова)	2 (1 из них не входит во вторую тысячу)	5 (1 из них не входит во вторую тысячу)	3 (1 из них не входит во вторую тысячу)	10 (1 из них не входит во вторую тысячу)	21 (7 из них не входят во вторую тысячу)	13 (7 из них не входят во вторую тысячу)	11	
груз. (36 слов)	0	2 (1 из них не входит во вторую тысячу)	0	7 (2 из них не входят во вторую тысячу)	4 (2 из них не входят во вторую тысячу)	3 (2 из них не входят во вторую тысячу)	5	
кит. (47 слов)	5 (1 из них не входит во вторую тысячу)	5 (все входят во вторую тысячу)	0	4 (1 из них не входит во вторую тысячу)	9 (4 из них не входят во вторую тысячу)	13 (7 из них не входят во вторую тысячу)	5	
итбо (61 слово)	12 (3 из них не входят во вторую тысячу)	14 (3 из них не входят во вторую тысячу)	7 (3 из них не входят во вторую тысячу)	14 (3 из них не входят во вторую тысячу)	20 (6 из них не входят во вторую тысячу)	32 (15 из них не входят во вторую тысячу)	21 (9 из них не входят во вторую тысячу)	

Таблица 1. Сравнение частотностей слов с одинаковым значением в разных языках

ных языков оно отсутствует в турецком, поскольку в этом языке отрицание выражается аффиксом. В предварительном подсчёте было ещё слово 'one', которое оказалось за пределами первой тысячи в немецком языке (возможно, из-за того, что в составе числительных больше 20 оно имеет не такую форму, как в изолированном употреблении). Ещё для десяти слов ('all', 'come', 'eye', 'full', 'give', 'hand', 'heart', 'long', 'one', 'see') ранг больше тысячи показывает лишь один язык из рассмотренных.

В большинстве языков не менее 60 слов входит в первые две тысячи (среди рассмотренных языков отклоняются лишь нидерландский – 59 слов и грузинский – 56 слов), в первые три – не менее 70 слов (а чаще – более 80).

Для сопоставления частотности и устойчивости были взяты данные из работ, где списки базисной лексики ранжировались по степени стабильности: Старостин 2007 (с поправкой на то, что в настоящей работе не учитываются слова, добавленные С. Е. Яхонтовым; ни одно из них не входит по устойчивости в верхнюю полусотню), Holman et al. 2008, Поздняков 2014, Коровина 2019 для слов сводешевского списка, а также данные из работы Tadmor 2009: 67, подсчитанные для списка «Лейпциг – Джакарта», на 38 слов отличающегося от списка Сводеша.

Степень устойчивости слова в разных языках, как показывают подсчёты, проведённые на различных выборках и по различным методикам (см. Старостин 2007, Holman et al. 2008, Коровина 2019, Поздняков 2014), оказывается неодинаковой: так, слова 'cloud' или 'tail' оказываются чрезвычайно устойчивы в тюркских языках, но крайне неустойчивы в индоевропейских (Старостин 1989: 15), слово 'name' стабильно в языках Евразии, но не Африки (Поздняков 2014). Однако средняя частотность по списку является достаточно стабильной величиной: бóльшая половина списка обычно входит в первую тысячу по частотности, а за пределами третьей тысячи оказываются лишь 10–15 слов.

Есть слова, которые очень сохранны, но низкочастотны, например, 'louse' (ни в одном из рассмотренных языков слово с этим значением не входит в первые 5000 по частотности). Вероятно, это связано с изменением хозяйственного уклада: в современном мире педикулёз не особенно распространён, тогда как у охотников-собирателей дело обстояло иначе, некоторые народы вшей даже ели (Mowat 1958: 18).

Частотность может различаться в зависимости от ареала: так, в языках Севера слово 'seed' отсутствует или низкочастотно, в полинезийских языках отсутствует или низкочастотно слово 'horn'.

Таким образом, М. В. Арапов и М. М. Херц в некотором смысле правы, характеризуя список Сводеша как скорее этнографический, чем лингвистический. Однако чисто лингвистический список (в том смысле, который предлагается в работе Арапов, Херц 1974) составить нереально: верхние позиции по частотности занимают предлоги, союзы, частицы, артикли и вспомогательные глаголы, которые бывает довольно непросто перевести с языка на язык (Саенко 2014: 49), кроме того, то, что в каком-то одном языке выражается предлогом или т. п., в другом может выражаться аффиксом (например, в турецком частотном словаре отсутствует перевод слова «не», поскольку в турецком языке отрицание выражается внутри глагола). К тому же, предлоги, союзы, частицы, артикли и вспомогательные глаголы, во-первых, возникают в результате грамматикализации (часто с утратой части фонемного состава и морфемной структуры), а во-вторых, подвергаются дальнейшей грамматикализации (с дальнейшей редукцией). Поэтому подход, основанный на списках устойчивой базисной лексики, представляется более продуктивным.

Редкими, но устойчивыми являются все названия нечеловеческих частей тела ('feather', 'horn', 'tail'; 'wing') и насекомых ('louse'; 'ant'), а также 'fingernail', 'dry' и 'liver'. Частотны и устойчивы местоимения и первые числительные, а из знаменательных слов –

‘name’ и ‘water’. Частотны и неустойчивы слова, содержащие оценки: ‘good’, ‘small’ и ‘many’. Редки и неустойчивы слова ‘bark’, ‘big’ и ‘belly’.

При этом среди слов, имеющих высокую частотность (входящих в первую тысячу) во всех рассмотренных языках, 6 входят в первую десятку наиболее сохранных слов (по ранжированию С. А. Старостина), 12 – в первую треть, 3 – в последнюю десятку, 8 – в последнюю треть, а середины почти нет. Зато слова, которые входят в первые две тысячи, представляют собой среднюю часть списка. По подсчётам Е. В. Коровиной, в первую десятку по устойчивости входит слово ‘fingernail’, которое во всех рассмотренных языках не вошло даже в первые 2,5 тысячи, а слово ‘many’, завершающее список устойчивости, во всех рассмотренных языках входит в первые полтысячи наиболее частотных слов. К. И. Поздняков отмечает, что «наиболее универсальны начало и конец списка. Наиболее стабильные значения универсального списка являются наиболее стабильными в большинстве языковых семей. Наименее стабильные значения являются таковыми в большинстве семей. Средняя часть “универсального списка” наименее универсальна» (Поздняков 2014: 206).

При любой ранжировке стословника – и по Старостину, и по Вихману, и по Позднякову, и по Коровиной, и по рангам списка «Лейпциг – Джакарта» – слов верхней половины по устойчивости, которые имеют низкую частотность (более, чем в одном языке выходят за пределы первых трёх тысяч), несколько меньше, чем низкочастотных слов из нижней половины.

Безусловно, частотность по современным словарям показывает срез лишь одного из бесчисленных этапов эволюции. В современных языках высокие позиции в частотном списке занимают слова, связанные с современными реалиями – ‘государство’, ‘система’, ‘проблема’, ‘закон’, ‘производство’, ‘политический’ и т. п. В языках, на которых говорили люди другого социально-экономического уклада, выше могли быть слова типа ‘сеять’, ‘пасти’, ‘дойти’; когда люди перемещались в другие природные зоны, менее частотными становились слова типа ‘семя’ или ‘рог’; с победой над педикулёзом ушло в число редких слово ‘вошь’. Но интересно, что даже в такой ситуации значительная часть базисной лексики всё равно сохраняет высокие позиции. Таким образом, можно сделать вывод, что хотя прямой зависимости между частотностью лексики и её устойчивостью нет, всё же соотношение между базисностью лексики и её частотностью является неслучайным.

Наблюдения над частотностью разных лексем заставляют задуматься о том, что, возможно, стоило бы внести некоторые корректировки в методику отбора единиц сводешевского списка.

В соответствии с принципами, изложенными в работе Kassian et al. 2010, следует включать в список наиболее частотный и стилистический нейтральный вариант обозначения соответствующего понятия (Kassian et al. 2010: 48). В соответствии с этим для предельных глаголов, видимо, следует брать форму совершенного вида, а для непредельных глаголов – несовершенного вида: частотности соответствующих словарных единиц различаются очень сильно (см. Табл. 2 и Табл. 3).

Единственно исключение из этих закономерностей – глагол ‘bite’, у которого в русском языке частотности совершенного и несовершенного вида практически одинаковы: *кусать* – 8,5 ipm, *укусить* – 8,3 ipm. Возможной причиной этого является то, что характер соотношения между совершенным и несовершенным видом в данном случае – промежуточный между тем, что можно наблюдать в глаголах *есть/съесть* и *пить/выпить*, с одной стороны, и *жечь/сжечь* и *убивать/убить*, с другой.

Для глаголов перемещения наиболее частотной является форма однократного направленного перемещения (см. Табл. 4), для глаголов восприятия – форма ненамеренно-

го восприятия (см. Табл. 5 и Табл. 6). Интересно было бы сравнить частотность разных форм в других языках, имеющих систему аспектуальных противопоставлений.

Частично рекомендации такого типа отмечены в Kassian et al. 2010: так, для 'say' указано, что имеется в виду однократный речевой акт, для глаголов местоположения 'sit', 'stand' и 'lie', а также для глагола 'sleep' – что следует собирать формы, обозначающие именно состояние, а не переход в него. Для 'see' и 'hear' указано, что следует собирать именно формы ненамеренного восприятия – но лишь для 'see' сказано, что нужны формы именно несовершенного вида. Вероятно, сходную рекомендацию имеет смысл распространить и на глагол 'hear'.

Для глагола 'go' сказано, что нужна форма длительного действия (Kassian et al. 2010: 62), что не очень хорошо отличимо от 'ходить' (ср. *Он каждый день ходит на озеро*), для глагола 'swim' приводятся контексты как для 'плыть', так и для 'плавать', при глаголе 'fly' (Kassian et al. 2010: 60) отмечено, что если в языке различаются корни для 'лететь' и 'летать' (а также для 'идти' и 'ходить'), то допустимо включать в список оба корня на правах синонимов. Между тем, чем меньше в списке синонимов, тем точнее результат. Так что, возможно, стоит и на глаголы перемещения распространить принцип максимизации частотности и требовать для них форм направленного перемещения.

Предельные глаголы				
Единица стословника	Несов. вид	Частотность (ipm)	Сов. вид	Частотность (ipm)
burn (tr.)	жечь	14,1	сжечь	21,6
come	приходить	218,2	прийти	523,3
die	умирать	62,0	умереть	179,7
give	давать	370,7	дать	573,1
kill	убивать	49,8	убить	144,7
say	говорить	1755,0	сказать	2396,6
Непредельные глаголы				
Единица стословника	Несов. вид	Частотность (ipm)	Сов. вид	Частотность (ipm)
drink	пить	200,9	выпить	129,9
eat	есть	94,7	съесть	43,2
see	видеть	818,2	увидеть	452,4
hear	слышать	256,1	услышать	160,4
know	знать	1713,8	узнать	238,1
sleep	спать	221,9	заснуть	29,7
			уснуть	26,5

Таблица 2. Сопоставление частотностей совершенного и несовершенного вида (по Ляшевская, Шаров 2009).

Единица стословника	Состояние	Частотность (ipm)	Переход в состояние	Частотность (ipm)	Вторичный имперфектив	Частотность (ipm)
lie	лежать	318,1	лечь	63,5	ложиться	39,5
sit	сидеть	538,1	сесть	175,9	садиться	77,9
stand	стоять	419,3	встать	172,1	вставать	69,1

Таблица 3. Частотности глаголов местоположения (по Ляшевская, Шаров 2009).

Единица стословника	Направленное движение	Частотность (ipm)	Ненаправленное движение	Частотность (ipm)
fly (v.)	лететь	82,9	летать	46,6
go	идти	957,1	ходить	296,6
swim	плыть	42,5	плавать	33,1

Таблица 4. Сопоставление частотностей глаголов направленного и ненаправленного перемещения (по Ляшевская, Шаров 2009).

Единица стословника	Невольное восприятие	Частотность (ipm)	Целенаправленное восприятие	Частотность (ipm)
hear	слышать	256,1	слушать	239,5
see	видеть	818,2	смотреть	667,2

Таблица 5. Частотности глаголов восприятия в русском языке (по Ляшевская, Шаров 2009).

Единица стословника	Ранг	Глагол целенаправленного восприятия	Ранг
hear	198	listen	619
see	58	look	87

Таблица 6. Частотности глаголов восприятия в английском языке (ранги по Davies, Gardner 2010).

Интересно сопоставить данные о частотности и устойчивости слов с результатами, полученными в рамках исследования детской речи. Для оценки речевого развития ребёнка используется так называемый Макартуровский опросник (версия для русского языка – Елисеева и др. 2017). Там перечислены слова, которые нормально развивающийся ребёнок, скорее всего, будет знать к полутора и к трём годам. Разумеется, рамки нормального развития достаточно широки и для того, чтобы ребёнок не получил диагноза «задержка развития», надо знать не все слова, а лишь большую их часть (так, 50% девочек в полтора года понимают 263 слова, 50% мальчиков – 248 слов), но сами списки примечательны. В список, большую часть которого ребёнок должен освоить до трёх лет, входит 716 слов (включая звукоподражания и имена близких). В это число попадает более 2/3 стословного списка Сводеша – 71 слово; из них более 50 надо освоить до полутора лет. Большая часть этих слов входит в первую тысячу по частотности (см. Табл. 7). Из тех слов, которые надо знать до полутора лет, в первую тысячу входят 40 (71%), во вторую – 11 (20%), в третью 4 (7%), и лишь одно (2%) не входит даже в третью. Из слов, которые надо выучить до трёх лет, 9 (60%) входят в первую тысячу, 3 (20%) во вторую, 2 (13%) в третью, одно (7%) является ещё менее частотным. Среди слов, которые надлежит выучить после трёх лет, процент низкочастотных выше: 9 из 29 этих слов (31%) не входят даже в третью тысячу по частотности.

Слово из списка Сводеша	Русский эквивалент (по Елисеева и др. 2017)	Возраст освоения	Ранг частотности в русском языке (по Ляшевская, Шаров 2009)	Ранг устойчивости по Е. В. Коровиной
not	не*	1,5	3	37
I	я	1,5	5	1
what	что	1,5	9	18

Таблица 7. Рано осваиваемые единицы сводешевского списка (в русском языке).

Таблица 7. Рано осваиваемые единицы сводешевского списка (в русском языке) (продолжение)

Слово из списка Сводеша	Русский эквивалент (по: Елисеева и др. 2017)	Возраст освоения	Ранг частотности в русском языке (по: Ляшевская, Шаров 2009)	Ранг устойчивости по Е. В. Коровиной
this	этот	1,5	14	42
we	мы	1,5	18	5
thou	тебе**	1,5	33	2
all	все	1,5	35	102
person	человек	1,5	39	79
who	кто	1,5	67	22
hand	рука	1,5	74	24
walk (go)	идти	1,5	95	95
big	большой	1,5	96	109
eye	глазки**	1,5	110	4
head	голова	1,5	132	70
give	дать	1,5	157	29
sit	сидеть	1,5	169	76
water	вода	1,5	191	23
good	хороший	1,5	199	99
foot	нога	1,5	205	49
stand	стоять	1,5	224	35
small	маленький	1,5	229	107
white	белый	1,5	290	81
black	чёрный	1,5	293	60
tongue	язык	1,5	306	7
lie	лежать	1,5	312	97
red	красный	1,5	442	87
sleep	спать	1,5	486	75
drink	пить	1,5	551	43
tree	дерево	1,5	659	46
sun	солнце	1,5	690	36
hair	волосы**	1,5	842	38
ear	ушки**	1,5	860	28
nose	нос	1,5	863	33
mouth	рот	1,5	901	88
dog	собака	1,5	909	11
stone	камень	1,5	910	30
green	зелёный	1,5	958	98
star	звезда	1,5	975	26
cold	холодный	1,5	1008	104
tooth	зубы**	1,5	1032	10
mountain	горка**	1,5	1040	100
warm	тёплый	1,5	1164	52

Таблица 7. Рано осваиваемые единицы сводешевского списка (в русском языке) (продолжение)

Слово из списка Сводеша	Русский эквивалент (по: Елисеева и др. 2017)	Возраст освоения	Ранг частотности в русском языке (по: Ляшевская, Шаров 2009)	Ранг устойчивости по Е. В. Коровиной
eat	есть	1,5	1290	17
bird	птичка**	1,5	1294	62
fish	рыба	1,5	1436	39
neck	шея	1,5	1449	91
rain	дождь	1,5	1490	44
yellow	жёлтый	1,5	1582	101
meat	мясо	1,5	1691	73
belly	живот	1,5	1831	106
sand	песок	1,5	2150	108
moon	луна	1,5	2525	34
egg	яйцо	1,5	2802	67
bite	кусать	1,5	9101	89
that	тот	3	36	74
say	сказать***	3	42	84
one	один	3	48	8
two	два	3	70	3
new	новый	3	73	31
see	видеть	3	113	90
night	ночь****	3	236	71
hear	слышать / слушать*****	1,5	411	51
full	полный	3	250	15
road	дорога	3	300	61
long	длинный	3	667	48
knee	колени	3	1178	69
dry	сухой	3	1493	56
fly (v.)	лететь	3	1495	68
cloud	облако	3	2643	94
swim	плыть / плавать*****	3	2668	54
claw(nail)	ногти*	3	4112	9

\* Только в составе выражений *не хочу, не буду*.

\*\* Слово либо не в начальной грамматической форме, либо с уменьшительным суффиксом.

\*\*\* Входит в пассивный словарный запас: уже от ребёнка до полутора лет требуется понимать слово *скажи*, хотя и не требуется уметь его произносить. В число слов, осваиваемых до трёх лет, входит *говорить*; *сказать* появляется ещё позже.

\*\*\*\* До полутора лет осваивается *ночью*; *ночь* как существительное осваивается в возрасте от полутора до трёх лет.

\*\*\*\*\* Авторы русской версии опросника считают данные глаголы за одну лексическую единицу: знание их (либо одного, либо другого) отмечается в одной общей клетке.



В этот список не включены позиции 'man' и 'woman', поскольку, хотя ребёнок уже в полтора года умеет называть отдельными словами мужчин и женщин, в русском языке для этого используются слова «языка нянь» (так называется «языковая подсистема, которую речевой коллектив считает пригодной в основном для общения с маленькими детьми» Елисеева и др. 2017: 7) – *дядя* и *тётя*. Возможно, при сборе стословника имеет смысл специально оговорить на всякий случай, что слова из «языка нянь» собирать не следует (хотя на практике вероятность этого, скорее всего, невелика).

Если сравнить те значения, которые должен выучить русскоязычный ребёнок, с частотностью соответствующих слов в других языках, выясняется, что в любом из них около 60% слов, входящих в первую тысячу по частотности, обозначают те реалии, названия которых мы осваиваем в раннем возрасте. Естественно, в разных языках в число рано осваиваемых будут входить разные слова (например, в английском языке слова *come* и *name* осваиваются гораздо раньше, чем их русские эквиваленты *прийти* и *имя*), так что подсчёт является лишь приблизительным, но, как кажется, результаты не должны расходиться слишком сильно.

При этом корреляции между ранним освоением и устойчивостью нет: если взять ранжирование по Е. В. Коровиной, то из рано осваиваемых слов половина относится к верхней половине списка, половина – к нижней; другие варианты ранжирования по устойчивости дают сходные результаты. Вероятно, это обусловлено тем, что раннее освоение, как и частотность (по современным словарям) связаны с нынешним уровнем развития культуры и типа хозяйства, тогда как устойчивость в большей степени ориентирована на прошлое.

Макартуровский опросник охватывает раннюю часть чувствительного периода – т. е. того периода, когда ребёнок овладевает родным языком (целиком он продолжается в среднем примерно до шести лет). И вероятность слова перейти к следующим поколениям зависит от того, в каком возрасте человек с ним встречается: чем раньше человек выучил слово, тем вероятнее, что он его не забудет, не станет образовывать обозначение для соответствующего понятия от других корней, а при контактах – заменять на заимствование. Передача лексики происходит через контексты: слыша незнакомое слово в контексте понятной ситуации, ребёнок соотносит это слово с каким-то её элементом. В этой связи кажется оправданным несколько изменить рекомендуемые контексты для сбора лексемы 'hand' – с чисто анатомических на функциональные, типа «возьми в руку», «держит ложку/палку в руке», «дай руку» или т. п. Интересно, что для ноги приведены функциональные контексты (Kassian et al. 2010: 61), а для руки, хоть и дана отсылка к лексеме 'foot', контексты исключительно анатомические (Kassian et al. 2010: 63).

Изменения частотности слов в разных языках происходят независимо. Если бы это было не так, в парах близко родственных языков или языков, вступавших в интенсивные контакты, слова с одинаковым значением имели бы более сходную частотность. Но имеющиеся данные показывают, вероятность слова войти в первую тысячу в языке 1 при условии, что слово с этим же значением входит в первую тысячу в языке 2, зависит лишь от того, сколько слов в том и в другом языке входят в первую тысячу (см. Табл. 8–10).

Испанский и португальский языки близкородственны, тем не менее, для каждого из них найдётся по 7 слов, которые в одном языке входят в первую тысячу по частотности, а в другом нет. Но во французском языке, отстоящем от них несколько дальше, нет слов, входящих в первую тысячу по частотности, которые бы в испанском или португальском оказались менее частотными. Причина этого – в том, что во французском языке стословных слов, входящих в первую тысячу по частотности, меньше, чем в испанском или португальском.

не входят в первую тысячу слов среди входящих в первую тысячу в языке:	польский	чешский по: Jelínek et al. 1961	чешский по: Čermák, Křen 2011	немецкий
польский (51 слово)		1 (оно не входит во вторую тысячу)	3 (все входят во вторую тысячу)	6 (1 из них не входит во вторую тысячу)
чешский по: Jelínek et al. 1961 (64 слова)	14 (2 из них не входят во вторую тысячу)			15 (3 из них не входят во вторую тысячу)
чешский по: Čermák, Křen 2011 (57 слов)	9 (1 из них не входит во вторую тысячу)			7 (2 из них не входят во вторую тысячу)
немецкий (54 слова)	9 (2 из них не входят во вторую тысячу)	5 (2 из них не входят во вторую тысячу)	5 (1 из них не входит во вторую тысячу)	

Таблица 8.

не входят в первую тысячу слов среди входящих в первую тысячу в языке:	немецкий	английский	французский
немецкий (54 слова)		5 (1 из них не входит во вторую тысячу)	14 (6 из них не входят во вторую тысячу)
английский (57 слов)	8 (все входят во вторую тысячу)		15 (6 из них не входят во вторую тысячу)
французский (45 слов)	5 (1 из них не входит во вторую тысячу)	3 (1 из них не входит во вторую тысячу)	

Таблица 9.

не входят в первую тысячу слов среди входящих в первую тысячу в языке:	французский	испанский	португальский
французский (45 слов)		0	0
испанский (63 слова)	16 (7 из них не входят во вторую тысячу)		7 (1 из них не входит во вторую тысячу)
португальский (63 слова)	18 (7 из них не входят во вторую тысячу)	7 (2 из них не входят во вторую тысячу)	

Таблица 10.

В более новом частотном словаре чешского языка (Čermák, Křen 2011) стословных слов, входящих в первую тысячу по частотности, меньше, чем в прежнем (Jelínek et al. 1961) – соответственно, уменьшаются и расхождения в частотности с польским и немецким.

Итак, частотность слов с одним и тем же значением – разная в разных языках и даже в одном языке в разное время. И это, видимо, универсальный закон: если бы было иначе,

слова из списка не могли бы выпадать (потому что их бы часто употребляли, не забывали и не заменяли). Сам же список базисных значений – каков бы он ни был – имеет вероятностную природу: какие именно значения войдут в первые тысячи по частотности в данном конкретном языке в заданный момент времени – непредсказуемо, но в каждый момент в любом языке бóльшая часть стословника в них входит.

### Литература

- Арапов, М. В., М. М. Херц. 1974. *Математические методы в исторической лингвистике*. Москва: Наука.
- Елисеева, М. Б., Е. А. Вершинина, В. Л. Рыскина. 2017. *Макартуровский опросник: русская версия. Оценка речевого и коммуникативного развития детей раннего возраста. Нормы развития. Образцы анализа. Комментарии*. Изд. 2-е, испр. и доп. Иваново: ЛИСТОС.
- Коровина, Е. В. 2019. Ранжирование базисной лексики С. А. Старостина: материалы к улучшению. В: Н. Н. Казанский (ред.). *Индоевропейское языкознание и классическая филология – XXIII. Материалы чтений, посвященных памяти профессора Иосифа Моисеевича Тронского. Первый полумтом: 546–556*. Санкт-Петербург: Наука.
- Ляшевская, О. Н., С. А. Шаров. 2009. *Частотный словарь современного русского языка (на материалах Национального корпуса русского языка)*. Москва: Азбуковник.
- НКРЯ – *Национальный корпус русского языка*. Онлайн-версия: <https://ruscorpora.ru>.
- Поздняков, К. И. 2014. О пороге родства и индексе стабильности в базисной лексике при массовом сравнении: атлантические языки. *Вопросы языкового родства* 11: 187–225.
- Саенко, М. Н. 2014. *Общие инновации в базисной лексике как аргумент в дискуссии о балто-славянском единстве*. Дисс. на соискание ученой степени канд. филол. наук. Москва: Институт языкознания РАН.
- Старостин, С. А. 1989. *Сравнительно-историческое языкознание и лексикостатистика*. В: И. Ф. Вардуль (ред.). *Лингвистическая реконструкция и древнейшая история Востока (тезисы докладов)*. Ч. 1: 3–39. Москва: Наука.
- Старостин, С. А. 2007. *Определение устойчивости базисной лексики*. В: С. А. Старостин. *Труды по языкознанию: 580–590*. Москва: Языки славянских культур.
- Засорина, Л. Н. (ред.). 1977. *Частотный словарь русского языка*. Москва: Русский язык.
- Шмальцель Г., Г. Нозадзе. 2012. *Грузинско-азербайджанско-армянско-русский частотный словарь*. Тбилиси: Чешское землячество в Грузии «Злата Прага».

### References

- Aksan, Yeşim, Mustafa Aksan, Ümit Mersinli, Umut Ufuk Demirhan. 2017. *A Frequency Dictionary of Turkish. Core Vocabulary for Learners*. London: Routledge.
- Anon. *SketchEngine: Igbo corpus (igWaC)*. 2015–2017. Available online at: [www.sketchengine.eu/igtenten-igbo-corpus](http://www.sketchengine.eu/igtenten-igbo-corpus).
- Arapov, Mikhail V., Maisa M. Kherts. 1974. *Matematicheskiye metody v istoricheskoy lingvistike*. Moskva: Nauka.
- Brown, Nicholas J. 2003. *Russian learners' dictionary*. London: Routledge.
- Buckwalter, Tim, Dilworth Parkinson. 2011. *A Frequency Dictionary of Arabic. Core vocabulary for learners*. London: Routledge.
- Čermák, František, Michal Křen. 2011. *A frequency dictionary of Czech: Core vocabulary for learners*. London: Routledge.
- Davies, Mark, Dee Gardner. 2010. *A Frequency Dictionary of Contemporary American English. Word sketches, collocates, and thematic lists*. London: Routledge.
- Davies, Mark, Kathy Hayward Davies. 2018. *A frequency dictionary of Spanish: Core vocabulary for learners. 2nd edition*. London: Routledge.
- Davies, Mark, Ana M. R. Preto-Bay. 2008. *A frequency dictionary of Portuguese: Core vocabulary for learners*. London: Routledge.
- Hebb, Donald O. 1949. *The organization of behavior: a neuropsychological theory*. New York: John Wiley and Sons.

- Holman, Eric W., Søren Wichmann, Cecil H. Brown, Viveka Velupillai, André Müller, Dik Bakker. 2008. Explorations in automated language classification. *Folia Linguistica* 42 (3-4): 331–354.
- Jelínek, Jaroslav, Josef V. Bečka, Marie Těšitelová. 1961. *Frekvence slov, slovních druhů a tvarů v českém jazyce*. Praha: Státní Pedagogické Nakladatelství.
- Jeliseeva, Marina B., Elena A. Vershinina, Viktorija L. Ryskina. 2017. *Makarturovskij oprosnik: russkaja versija*. Ivanovo: LISTOS.
- Jones, Randall L., Erwin Tschirner. 2006. *A frequency dictionary of German: Core vocabulary for learners*. London: Routledge.
- Kassian, Alexei, George Starostin, Anna Dybo, Vasiliy Chernov. 2010. The Swadesh wordlist. An attempt at semantic specification *Journal of Language Relationship* 4: 46–89.
- Korovina, Evgenija V. 2019. Ranzhirovanije bazisnoj leksiki S. A. Starostina: materialy k uluchsheniju. In: Nikolaj N. Kazanskij N. (ed.). *Indoevropskoje jazykoznanije i klassičeskaja filologija XXIII*: 546–556. Sankt-Petersburg: Nauka.
- Kurcz, Ida, Andrzej Lewicki, Jadwiga Sambor, Krzysztof Szafran, Jerzy Woronczak. 1990. *Słownik frekwencyjny polszczyzny współczesnej. T. 1–2*. Kraków: Instytut Języka Polskiego.
- Lonsdale, Deryle, Yvon Le Bras. 2009. *A frequency dictionary of French: Core vocabulary for learners*. London: Routledge.
- Lyashevskaja, Ol'ga N., Sergey A. Sharov. 2009. *Chastotnyj slovar' sovremennogo russkogo jazyka (na materialax Nacional'nogo korpusa russkogo jazyka)*. Moskva: Azbukovnik.
- Mowat, Farley. 1958. *Coppermine Journey: An Account of a Great Adventure – Selected from the Journals of Samuel Hearne*. Boston: Litle, Brown & Co.
- Pozdniakov, Konstantin I. 2014. O poroge rodstva i indekse stabil'nosti v bazisnoj leksike pri massovom sravnenii: atlantičeskije jazyki. *Journal of Language Relationship* 11: 187–225.
- Saenko, Mikhail N. 2014. *Obščije innovacii v bazisnoj leksike kak argument v diskussii o balto-slav'anskom jedinstve*. Dissertation Ms. Moskva: Institut jazykoznaia RAN.
- Shmal'cel', Garol'd, Givi Nozadze. 2012. *Gruzinsko-azerbajdzhansko-arm'ansko-russkij chastotnyj slovar'*. Tbilisi: European Centre for Minority Issues.
- Sharoff, Serge, Elena Umanskaya, James Wilson. 2013. *A frequency dictionary of Russian: Core vocabulary for learners*. London: Routledge.
- Starostin, Sergei A. 1989. Sravnitel'no-istoričeskoje jazykoznanije i leksikostatistika. In: I. F. Vardul' (ed.). *Lingvističeskaja rekonstrukcija i drevnejšaja istorija Vostoka. Part 1*: 3–39. Moskva: Nauka.
- Starostin, Sergei A. 2007. Opredelenije ustojčivosti bazisnoj leksiki. In: Serget A. Starostin. *Trudy po jazykoznaniju*: 580–590. Moskva: Jazyki slav'anskix kul'tur.
- Tadmor, Uri. 2009. Chapter III. Loanwords in the World's Languages: Findings and results. In: Haspelmath, Martin, Uri Tadmor (eds.). *Loanwords in the World's Languages. A comparative handbook*: 55–75. Berlin: Mouton de Gruyter.
- Tiberius, Carole, Tanneke Schoonheim. 2014. *A frequency dictionary of Dutch: Core vocabulary for learners. 2nd edition*. London: Routledge.
- Tono, Yukio, Makoto Yamazaki, Kikuo Maekawa. 2013. *A Frequency Dictionary of Japanese. Core Vocabulary for Learners*. London: Routledge.
- Xiao, Richard, Paul Rayson, Tony McEnery. 2009. *A frequency dictionary of Mandarin Chinese: Core vocabulary for learners*. London: Routledge.
- Zasorina, Lidija N. (ed.). 1977. *Chastotnyj slovar' russkogo jazyka*. Moskva: Russkij jazyk.

*Svetlana Burlak*. Stability and frequency: is there a correlation?

Words belonging to the Swadesh list of basic vocabulary are not only stable but also frequent. For each word, frequency varies across different languages and periods of time, but in general, more than half of the Swadesh items belong to the first thousand of the most frequent words, and ca. 80% of Swadesh items belong to the first three thousand, irrespective of language. It seems a priori probable that a meaning would belong to the basic vocabulary if it were easier for the speakers to remember an already existing word than to replace it with

a new derivative or borrowing. Given the construction of the human brain, things that are more frequent tend to be more strongly memorized (since the corresponding neuron circuits are strengthened, according to Hebb's postulate). For a preliminary testing of this hypothesis, we analyzed the Swadesh wordlist items in frequency dictionaries of 15 languages belonging to different families and linguistic areas. In all of these languages, 21 meanings were expressed by words belonging to the first thousand of the most frequent words. In most languages more than 60 words belong to the first two thousand and, typically, more than 80 words belong to the first three thousand of the most frequent words.

There is a rather strong correlation between frequency ranks of words with the same meaning in different languages. There are a few easily explicable counterexamples to this rule: e.g., if a word has several meanings, it will have a higher frequency. Languages of different families inherit different polysemy models from their protolanguages, which is why a universal wordlist containing a representative amount of words but not influenced by polysemy cannot be generated. Although frequency may vary depending on the speakers' way of life, a considerable amount of items on the Swadesh wordlist retain high frequency under any circumstances. Our data show that even if languages are closely related or involved in intensive contact, frequency changes within them are independent.

Observations on frequency ratings of different lexemes make it possible to provide some refinements to the method of compiling Swadesh wordlists. Thus, for terminative verbs, a perfective form should be chosen, while for non-terminative verbs, an imperfective form would be more appropriate. For verbs of movement, forms denoting a single directed movement should be chosen. For verbs of physical perception, forms denoting unintentional perception should be preferred. For the word 'hand', one should rely on functional rather than anatomical contexts.

Swadesh wordlist items are learned very early: thus, in Russian, children are expected to know ca. 70 Swadesh items before they are 3 years old. Words are acquired through their contexts, which is why the method of compiling Swadesh wordlist via diagnostic contexts is the most appropriate. The degree of stability of a word with a certain meaning differs in different languages; however, different samples of basic vocabulary show almost the same correlation with frequency. Though there is no direct dependence between frequency and stability of words, the relation between belonging to the subset of basic vocabulary and the subset of frequently used words is not random.

There is a correlation between early acquisition and frequency (but not stability). This is probably due to the fact that early acquisition and modern day frequency are more deeply connected with modern day lifestyle, while stability reflects previous epochs. Frequency of the words with the same meaning differs in different languages (and even in the same language in different times). This is, apparently, a universal principle: if it were not so, words would never have an opportunity to be eliminated from the Swadesh wordlist, since they would be so frequent that nobody could forget them and replace them with derivatives, borrowings, or other words. However, the list itself has a probabilistic nature: it cannot be predicted which meanings would gain higher frequency in a certain language at a certain period of time, but the majority of items on the Swadesh wordlist still belongs to it in any language at any period of time.

*Keywords:* basic vocabulary; Swadesh wordlist; lexical stability; lexical frequency; language acquisition.

Juan Luis García Alonso

University of Salamanca, Spain; jlga@usal.es

Alejandro G. Sinner, Javier Velaza (eds.).  
Palaeohispanic Languages and Epigraphies.  
Oxford: Oxford University Press, 2019. — 483 + xx p.

This book is a recent overview of the most important aspects of what we currently know about the ancient languages and epigraphic materials of the Iberian Peninsula and Southern France. These languages and texts received the denomination of ‘Palaeohispanic’ a few years ago and the name seems to have been generally accepted. The book, written by some of the most active and respected researchers in the field, is a collection of 14 essays covering all the regions as well as all the pertinent questions of the current state of our knowledge.

Besides Phoenician, Greek and Latin we have epigraphic evidence of several Palaeohispanic languages (Tartessian, Iberian, Celtiberian, Lusitanian) in a really sizable corpus of over 3000 inscriptions. These epigraphic texts and the languages in which they are written are relevant for many reasons, among them the history and spread of writing in the Mediterranean context, and the different substrata of the modern languages of the region.

In such a relatively young field of research as this, with often very disparate positions, the different authors were asked by the editors to be meticulous when dealing with the frequently opposing views, making an effort to differentiate “between those hypotheses that can count on broad acceptance and those that are still currently under evaluation and debate” (p. vi), something that was achieved throughout.

### 1. Method and methods: Studying Palaeohispanic languages as a discipline (pp. 1–24)

Javier de Hoz, who passed away before the book came out, is the author of two chapters of this global overview of Palaeohispanic languages and epigraphies. His ample knowledge of many different aspects of *Palaeohispanistics* made him an appropriate candidate for this methodological introduction.

After defining this field of studies in temporal, linguistic and geographical terms, he summarizes the linguistic and epigraphic situation of Hispania and the history of the research since the Renaissance, with special attention to the role of Gómez-Moreno.

After a section describing the methods of decipherment of scripts and languages, he urges to be wary of risky constructs impossible to prove, and advocates for what he calls “Palaeohispanic Philology”: to go beyond a basic understanding of scripts and languages and arrive at a cultural appreciation of the context of the texts in all its complexity.

The author points out an aspect that may be easily overlooked: the temporal depth of Palaeohispanic texts. A few final comments are added on language contacts, as well as on the diachronic and synchronic variability within each language area.

### 2. The Iberian Peninsula in pre-Roman times: An archaeological and ethnographical survey (pp. 25–55)

Alberto Lorrio and Joan Sanmartí summarize the archaeological and ethnographical data of Hispania during this period, focusing on the difficulties that this research entails, given the uneven information available (coins, Greek and Latin sources, epigraphy or archaeology).

They include three sections: Iberians, the South-Western people(s) and what they call “Hispano-Celts”, the latter being subdivided in Celtiberians, Vaccaei, Vettones (with the Celtic peoples of the West and North) and the group of Turmodigi, Autrigones, Berones, Caristi and Varduli.

It is striking, for a linguist, not to see a specific heading dedicated to the speakers of the Indo-European but clearly non-Celtic language that we today call Lusitanian. The authors seem to endorse the possibility that the Lusitani are “perhaps” Celtic (p. 39), something I cannot agree with. To me, the description of Lusitanian on the following page as “an archaic Indo-European language (...) with elements in common with the Celtic sub-family” is misleading. Lusitanian does share some isoglosses with Celtic, but also with other sub-families within Indo-European, such as the Italic or Germanic languages. The presence of the Celtic place names in *-briga* across central and western Hispania does not imply anything about the Celtic nature of Lusitanian, but rather that the area was *also* in-

habited by speakers of Celtic, the difficulty being how to explain this co-existence.

### 3. Phoenician epigraphy in the Iberian Peninsula (pp. 56–77)

J. A. Zamora begins with a history of the epigraphic finds, in southern Hispania, which for a long time were really scarce, although nowadays the number of texts is already significant (around 400 or so). He then overviews the geographical distribution and the dating of the finds and analyses the main characteristics and issues concerning the preserved material. He makes an important point that the Phoenicians most likely used to write on perishable materials, which helps to understand the irregular distribution of texts.

He then summarizes the uses of writing amongst the Phoenicians. Modern research is able to confirm a much more widespread writing practice than what the preserved inscriptions would lead us to believe. Examples of the uses of writing belong in the economic, funerary, ritual (offerings and apotropaic practices) and official spheres.

### 4. Palaeohispanic writing systems: Classification, origin, and development (pp. 78–108)

J. Ferrer and N. Moncunill undertake a study of all the varieties of said writing systems, how they originated and how they reached their geographic locations. The internal relation between all the scripts is not yet ascertained. In a more limited way both the Latin and Greek alphabets were used to represent native languages (Lusitanian, Celtiberian and Iberian).

The semi-syllabaries can be divided into Northern and Southern scripts. The first type includes North-eastern Iberian, the Celtiberian varieties and a possible third script, not yet clearly defined, attested in the lands of the *Vascones*. The second type includes South-eastern Iberian, Southwestern (Tartessian) script and the Espanca script. Additionally, there are a few southern texts that seem to indicate there was at least one more southern script. The authors remark on the important discovery of the ‘dual’ varieties where differentiating voiceless and voiced plosives was possible. Page 81 offers a complete representation of all the varieties and values of each sign.

The authors then provide a more detailed overview of the different scripts, finally offering their hypothesis on the genealogy of all the scripts, which I see as sensible but highly uncertain or difficult to prove at this stage. Two separate adaptations from a Phoenician source feel unlikely to me.

### 5. The epigraphic and linguistic situation in the south-west of the Iberian Peninsula (pp. 109–137)

J. A. Correa and A. Guerra discuss the linguistic situation in this area, including the typology and classification of the language which appears in this epigraphic corpus and is often associated with the Tartessian tradition. The script is not fully decoded. We are not certain about the phonetic value of many of the signs, most of the texts (around 100) are very brief and in most cases we do not know how to separate the words.

They then state that this script was based upon the Phoenician alphabet, most likely with no participation of the Greek alphabet (De Hoz 2010: 495–500) and finally describe this semi-syllabary, pointing out the many questions on which researchers have not reached consensus. The authors tackle the challenge of the historical context, dating these texts to the 6<sup>th</sup> and 5<sup>th</sup> centuries BCE (rather than the 8<sup>th</sup> century as previously believed).

Concerning the language of the script, the authors make a general description and remind the reader that certain features (personal names in particular) were tentatively labelled by Correa (1985) as possibly Celtic or Indo-European, a hypothesis he later discarded. Untermann (MLH IV), Rodríguez Ramos (2002) and De Hoz (2010: 371–402) have also explicitly rejected this possibility.

As for Koch’s more recent (2009, 2011 or 2014) hypothesis of the language as Celtic, Correa and Guerra notice that, on top of having met with opposition on Celtic linguistic grounds, in their opinion, this is far from proven in sheer epigraphic terms as well, something that I agree with.

### 6. The linguistic situation in the territory of Andalusia (pp. 138–159)

This second chapter by J. de Hoz covers the linguistic landscape of the different regions of modern Andalusia in ancient times. In each of these areas there were peoples known from ancient sources:

- the *Turdetani*, descendants of the Tartessians, occupied the lower and middle Guadalquivir valley. Probably akin to them were the neighbouring *Turduli*, who lived on the coastal region of Huelva and Cádiz;
- there was, similarly, another pair of peoples with a parallel double denomination, *Bastetani* and *Bastuli*, with a similar ethnic base;
- to the North the *Oretani* occupied areas both within modern Andalusia and in the southern Meseta, and to the South the Phoenicians were located in and around their city-states by the Mediterranean coast;

- finally, in Roman times, in western Andalusia and southern Portugal, we know of the presence of Celtic-speaking peoples.

The author reviews the linguistic material available through onomastics and native inscriptions, dividing it into three parts: Turdetanian for the lower and middle Guadalquivir, Iberian for the Oretanian and Bastetanian regions of the upper Guadalquivir and Celtic for western Andalusia and adjacent areas of southern Portugal.

### 7. Iberian writing and language (pp. 160–197)

After stressing the main methodological, geographical and chronological questions faced by modern research on the subject, Javier Velaza recognizes that the difficulties do not lie in the scarcity of texts (there are 2250 extant inscriptions), but rather in the hermetic character of a language with no known close relatives.

He then refers to the surprisingly homogeneous aspect of the language given the large epigraphic extension of Iberian, especially despite the fact that many different peoples inhabited the area and that the archaeological evidence is varied as well. He mentions De Hoz's hypothesis that Iberian was not vernacular throughout but acted as a *lingua franca*, being only vernacular in the Southeastern area, but remains unconvinced; as for the place of origin of the language, he would rather favour the Northeast.

Velaza then lists an overall inventory of the inscriptions written in the different varieties of scripts used for this language, after which he discusses the evolution of different usages of writing among the Iberians.

Finally, there is a section dedicated to a prudent and sensible description of the little that we know for certain about the language on different levels: onomastics, phonetics and phonology, nominal morphology, vocabulary, numerals, the verb, and syntax. From a typological perspective, the author believes Iberian to have been an agglutinative and, possibly, ergative language.

### 8. Cultural and linguistic contacts in southern Gaul (pp. 198–218)

This chapter is an introduction to the linguistic situation in Southern Gaul, which, even though geographically not a part of Hispania, is still included in *Palaeohispanistics* because of the interactions between Iberian, Gaulish, Greek and Phoenician.

A. Mullen and C. Ruiz Darasse explore the complexities of the language contacts in the area, divided into two epigraphic zones limited by the river Hérault, one to the west, using the Northeastern Iberian script, and another one to the east, using the Greek alphabet.

They overview the early contacts of this area with the Greeks, Etruscans and Phoenicians. The pre-Roman Greek epigraphic presence does not show any linguistic interaction with other communities (local or Iberian). Evidence of Greek-Etruscan (even Roman) contacts in the area earlier are clear, as are contacts between Etruscan and Gaulish/Celtic populations. The chapter continues with a revision of the contacts with the Iberian world.

The final section deals with the contacts in Roman times; the development of the so-called Gallo-Greek epigraphy; the less frequent epigraphic examples of direct Greek-Iberian interactions; the few local cases (Haute-Garonne) of a threefold Iberian, Latin and Gaulish interaction; and the final stages where Latin took over with an immediate disappearance of Iberian names and a more gradual elimination of Gaulish ones.

### 9. The Vasco-Iberian theory (pp. 219–239)

This old controversial question is discussed in detail in this chapter by E. Orduña, who had first put on the table the (at least apparently) surprising similarities of the entire numeric systems of Basque and Iberian.

The author begins by summarizing the historical development of the hypothesis since the sixteenth century, even though until Gómez-Moreno's contributions from the 1920s onwards (1922, 1925, 1942, 1943, 1945, 1949, 1962) the readings of the Palaeohispanic scripts were not correct. After him the panorama changed, with most of the important scholars<sup>1</sup> rejecting the idea of a genetic relation between the two languages. Michelena's opinions "are still dominant" (p. 221) today, as Orduña says.

After reviewing the traditional reasons supporting the theory, Orduña analyses different arguments that cannot be discarded easily, at various levels: phonology, noun morphology, pronouns, verbal morphology, syntax, onomastics, and lexicon.

The chapter ends with a few sensible words concerning the difficulty of closing this discussion and the need to know much more about Iberian and about ancient Basque to be certain about the degree of closeness between the two.

### 10. Writing and language in Celtiberia (pp. 240–303)

In the context of ancient Hispania, the best known language is Celtiberian. F. Beltrán and C. Jordán overview the language and analyze different epigraphic issues.

<sup>1</sup> Caro Baroja (1942, 1943), Bähr (1948), Tovar (1954, 1961) or Michelena (1961, 1975, 1985).



After a historical and geographical introduction to the peoples inhabiting this region, the authors point out that this corpus of over 200 inscriptions constitutes the only native epigraphic area in Indo-European Hispania, with the exception of half a dozen Lusitanian texts. All the lands to the West and Northwest of Celtiberia, with the exception of Lusitanian territory, must have been inhabited by speakers of dialects of Hispano-Celtic<sup>2</sup>. The authors mention that modern researchers debate whether Lusitanian is one of them; I would add that most linguists do not think it is.

After reviewing Celtiberian phonetics and phonology<sup>3</sup>, nominal morphology, adjectival formation, verbs, pronouns, numerals, adverbs, prepositions, conjunctions and particles, prefixes and preverbs, there is a section dedicated to personal onomastics, followed by another one dealing with the Celtiberian epigraphic culture.

### 11. Language and writing among the Lusitanians (pp. 304–334)

Eugenio Luján overviews the use of writing among this people. They did not use a Palaeohispanic script, but would rather recur (sporadically) to the Latin alphabet in order to represent their language, whose linguistic classification is also discussed here. Lusitanian is the name of a language known from half a dozen native inscriptions<sup>4</sup> coming from the territory of the Lusitani. In the same area, further linguistic (onomastic) elements, of varied nature (some akin to these texts, some clearly Celtic) have surfaced. Moreover, we have onomastic evidence of Lusitanian beyond the lands of the Lusitani, to the North, in the territory of the Gallaeci.

The chapter follows with a description of the language, based first in the native language texts, but also taking into account the remaining linguistic informa-

<sup>2</sup> “Celtiberian is the only pre-Roman language directly attested in the Iberian Peninsula which is undoubtedly Celtic” (p. 247); however, Celtic was spoken in other regions beyond Celtiberia. The term Hispano-Celtic should be applied to those other Celtic dialects which we cannot call Celtiberian with confidence (leaving Lusitanian out, of course).

<sup>3</sup> In relation to the evolution of \*p > φ > Ø, this “fricativization and loss of \*p in the initial position before a vowel or continuant and intervocalic position” (p. 149), I do not think that Cib. **konbouto** [Mon.74] is a good example, since the voiceless plosive is not found here in the right context. I do not think we really need a reading of [kom-φlouto]; we would not expect the disappearance of p- in that context (after a nasal sound).

<sup>4</sup> Either 6 or 7, depending on whether the lost inscription of Arroyo de la Luz, with its two pieces, is counted as a single one broken in two or as two separate texts.

tion obtained by the onomastic and morphological details seen in Latin inscriptions.

Luján then discusses the classification of Lusitanian within Indo-European, mentioning the two main positions about its possible Celtic character, starting with Tovar (1966) (against) and Untermann (1984, 1987) (in favour). The author’s main conclusion is that Lusitanian cannot be considered a Celtic language and I agree with his arguments, all of which seem quite reasonable.

### 12. The parts of Hispania without epigraphy (pp. 335–364)

In this section, J. Gorrochategui and J. M. Vallejo overview the linguistic situation in the areas of Hispania that did not have native epigraphic testimonies of the local languages.

The authors introduce alternative sources to epigraphics, identifying distinct linguistic areas by means of information gathered from classical literary sources or from epigraphic texts in Latin which have some native onomastics embedded in them.

After a few reasonable comments on the nature of toponyms, anthroponyms and theonyms, the authors overview the linguistic areas of the regions of Hispania without a native epigraphy:

- Western Hispania, where Lusitanian was spoken, shows anthroponyms that are congruent with it, but also a few in common with its northern neighbours;
- the Celtiberian area, well known through its native epigraphics, with specific characteristics extending to other peoples to their West and Northwest;
- Southern Hispania, with native texts from different periods, which was romanized much earlier than Northern and Western Hispania and shows three layers of onomastic evidence — Celtic in the West, Turdetanian in the Guadalquivir basin and Iberian in the East;
- the Iberian area;
- the Basque-Aquitanian area, with onomastics akin to the Basque language.

### 13. Coin evidence for Palaeohispanic languages (pp. 365–395)

As pointed out by P. P. Ripollés and A. G. Sinner in this section, numismatics was the first discipline to face the challenge of analysing certain native texts in an unknown type of writing coming from different areas of Hispania. Already during the 16th century this was a matter of concern among specialists. This chapter starts with a history of this research until our times.

The authors offer a special section dedicated to the classification of native coins, associating each of them with specific dates: Greek (6<sup>th</sup>-2<sup>nd</sup> centuries BCE),

Punic (4<sup>th</sup> century BCE – imperial period), Northeastern Iberian (3<sup>rd</sup> – 1<sup>st</sup> centuries BCE), Southeastern Iberian (3<sup>rd</sup> – 1<sup>st</sup> centuries BCE), Celtiberian (2<sup>nd</sup> – 1<sup>st</sup> centuries BCE), unidentified Southern script (2<sup>nd</sup>–1<sup>st</sup> centuries BCE).

The final part of this section describes and classifies the coins that were minted in Hispania, in every language and in every context, in some more detail, and trying to obtain all kind of historical, economic or social information from each of the areas, cultural milieus, and linguistic surroundings.

#### 14. Writing, colonization, and Latinization in the Iberian Peninsula (pp. 396–416)

Both the Palaeohispanic native writing systems and the native ancient languages would die out as a final consequence of the long process of Latinization. B. Díaz Ariño, M. J. Estarán and I. Simón overview the introduction of the Roman epigraphic culture in Hispania. There is a section dealing with the first written texts (3<sup>rd</sup> century BCE), related to the trade between Italy and the Mediterranean ports of Hispania, followed by another describing the epigraphic activity of the Roman administration until the civil wars, after which the authors define a phase in the development of Latin epigraphy due to the arrival of Italic immigrants. The final step is linked to the epigraphic use of Latin by the local elites.

The authors then offer a reflection on linguistic contacts, as perceived by the presence of bilingual inscriptions, and on the effects they had on the whole Latinization process, with a separate analysis for each area (Iberian, Celtiberian and Lusitanian).

The chapter ends with an analysis of the disappearance of Palaeohispanic languages and scripts. In some areas both events took place almost simultaneously, while in others the languages survived much longer. Local writing systems ceased to be used completely during the first decades of the 1<sup>st</sup> century CE. The Iberian and Turdetanian languages died out almost at the same time. However, Indo-European Hispania was very different: Romanization started later, the immigrants were far fewer and concentrated in a few urban centers, and the languages could have survived until the end of antiquity and, in the case of Basque, till today.

#### Final pages

The book ends with a very complete (although it admits than including 100% of the pertinent publications is not possible) and useful bibliographic selection (pp. 417–472), followed by a concordance of the inscriptions in *BDHesp* and *MLH* and a detailed index of sources.

#### References

- Bähr, Gerhard. 1948. *Baskisch und Iberisch*. Bayonne: Impr. Daracq.
- Caro Baroja, Julio. 1942. Observaciones sobre la hipótesis del vascoiberismo considerada desde el punto de vista histórico. I. (Exposición de textos). *Emerita* 10: 236–286.
- Caro Baroja, Julio. 1943. Observaciones sobre la hipótesis del vascoiberismo considerada desde el punto de vista histórico. II. (Materiales epigráficos). *Emerita* 11: 1–59.
- Correa, José Antonio. 1985. Consideraciones sobre las inscripciones tartesias. In: J. de Hoz (ed.). *Actas del III Coloquio sobre lenguas y culturas paleohispánicas* (Lisboa 1980): 377–395. Universidad de Salamanca.
- Koch, John T. 2009. *Tartessian: Celtic in the South-west at the dawn of history*. Aberystwyth: Celtic Studies Publications.
- Koch, John T. 2011. *Tartessian 2. The inscription of Mesas do Castelinho. Ro and the verbal complex. Preliminaries to historical phonology*. Aberystwyth: Celtic Studies Publications.
- Koch, John T. 2014. On the debate over the classification of the language of the South-Western (SW) Inscriptions, also known as Tartessian. *Journal of Indo-European Studies* 42: 335–427.
- de Hoz, Javier. 2010. *Historia lingüística de la península ibérica en la Antigüedad. I: Preliminares y mundo meridional prerromano*. Madrid: Editorial CSIC.
- Gómez-Moreno, Manuel. 1925. Sobre los iberos: el bronce de Áscoli. In: *Homenaje a D. Ramón Menéndez Pidal, t. III*: 475–499. Madrid. (Reprinted in: Gómez-Moreno 1949: 233–256).
- Gómez-Moreno, Manuel. 1942. *Las lenguas hispánicas*. Madrid: Imp. Gráficas Alpinas.
- Gómez-Moreno, Manuel. 1943. La escritura ibérica y su lenguaje. *Boletín de la Real Academia de la Historia* 112: 251–278.
- Gómez-Moreno, Manuel. 1945. Digresiones ibéricas, escritura, lengua. *Boletín de la Real Academia Española* 24: 275–288.
- Gómez-Moreno, Manuel. 1949. *Misceláneas. Historia, Arte, Arqueología. Primera serie, la antigüedad*. Madrid: S. Aguirre.
- Gómez-Moreno, Manuel. 1962. *La escritura bástulo-turdetana (primitiva hispánica)*. Madrid: Editorial CSIC.
- Michelena, Luis. 1961. *Fonética Histórica Vasca*. San Sebastián: Impr. de la Diputación de Guipúzcoa.
- Michelena, Luis. 1975. Guillaume de Humboldt et la Langue Basque. *Lingua e stile* 8: 107–125.
- Michelena, Luis. 1985. Sobre la posición lingüística del ibérico. In: Luis Michelena. *Lingua e Historia*: 334–340. Madrid: Paraninfo.
- MLH = Untermann, Jürgen, Dagmar S. Wodtko. 1975–2000. *Monumenta Linguarum Hispanicarum*. Wiesbaden: L. Reichert.
- Rodríguez Ramos, Jesús. 2002. Las inscripciones sudluso-tartesias. Su función, lengua y contexto socio-económico. *Complutum* 13: 85–95.
- Tovar, Antonio. 1954. Sobre el planteamiento del problema vasco-ibérico. *Archivum* 4: 220–231.
- Tovar, Antonio. 1961. *The ancient languages of Spain and Portugal*. New York: S. F. Vanni.
- Tovar, Antonio. 1966. L'inscription du Cabeço das Fraguas et la langue des lusitaniens. *Études Celtiques* 11.2: 237–268.
- Untermann, Jürgen. 1984. Los celtiberos y sus vecinos occidentales. *Lletres asturianas* 13: 6–26.
- Untermann, Jürgen. 1987. Lusitanisch, keltiberisch, keltisch. In: J. Gorrochategui, J. L. Melena, J. Santos (eds.) *Studia Palaeohispánica. Actas del IV coloquio sobre lenguas y culturas paleohispánicas. Vitoria-Gasteiz, 6–10 mayo 1985*: 57–76. Vitoria: Instituto de Ciencias de la Antigüedad, Universidad del País Vasco.