

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<sup>o</sup> 21, issue 1–2



Moscow 2023

Российский государственный гуманитарный университет  
Институт языкознания Российской Академии наук

# Вопросы языкового родства

Международный научный журнал

№21, часть 1—2

**ВЯР**

Москва 2023

*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., 2023. — № 21/1—2. — xiv + 155 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, 2023. — № 21/1—2. — xiv + 155 p.

ISSN 2219-3820

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

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

Подписано в печать 03.11.2023. Выход в свет 10.11.2023  
Формат 60×90/8. Уч.-изд. л. 14,0. Усл. печ. л. 20,8  
Тираж 1050 экз. Свободная цена  
Заказ № 1829

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

# Table of Contents / Содержание

|   |     |
|---|-----|
| Contributors / Сведения об авторах . . . . .                                    | ix  |
| Note for Contributors / Будущим авторам . . . . .                               | x   |
| От редакции / From the editors . . . . .  | xi  |
| <i>Mate Karović. Vladimir Antonovič Dybo (1931–2023): In memoriam</i> . . . . . | xii |
| [М. Капович. Памяти Владимира Антоновича Дыбо (1931–2023)]                      |     |

## Issue 1 / Часть 1

### Articles / Статьи

|  |    |
|--|----|
| <i>Andrey Nikulin. Lexical evidence for the Macro-Jê–Tupian hypothesis</i> . . . . . | 3  |
| [А. В. Никулин. Лексические свидетельства в пользу макро-же–тупийской гипотезы]      |    |
| <i>Mikhail Zhivlov. Studies in Yukaghir etymology II</i> . . . . .                   | 57 |
| [М. А. Живлов. Исследования в области юкагирской этимологии II]                      |    |

## Issue 2 / Часть 2

### Articles / Статьи

|   |     |
|---|-----|
| <i>Alexander Militarev. Hadza as Afrasian?</i> . . . . .  | 71  |
| [А. Ю. Милитарев. Хадза — афразийский язык?]  |     |
| <i>Bonny Sands, Andrew Harvey, Maarten Mous, Mauro Tosco.</i><br>Why Hadza is (probably) not Afroasiatic:<br>a discussion of Militarev’s “Hadza as Afrasian?” . . . . . | 91  |
| [Бонни Сэндс, Маартен Моус, Мауро Тоско, Эндрю Харви. Почему хадза (скорее всего) не афразийский язык: ответ на статью А. Ю. Милитарева «Хадза — афразийский язык?»]    |     |
| <i>George Starostin. How could we show that Hadza is Afroasiatic:<br/>a response to Militarev’s “Hadza as Afrasian?”</i> . . . . .                                      | 110 |
| [Г. С. Старостин. Как можно было бы показать, что хадза — афразийский язык: ответ на статью А. Ю. Милитарева «Хадза — афразийский язык?»]                               |     |
| <i>М. М. Лоренц. Классификация восточной ветви диалектов группы Минь<br/>и реконструкция правосточноминьского 100-словного списка</i> . . . . .                         | 117 |
| [Marina Lorentz. Classification of Eastern Min dialects and reconstruction of the 100-item wordlist for Proto-Eastern Min]  |     |





## Contributors

*Andrew Harvey* — Ph.D., junior professor, University of Bayreuth, andrew.harvey@uni-bayreuth.de

*Marina Lorentz* — graduate student, Institute of Linguistics, RSUH (Moscow), lorentzmarina8@gmail.com

*Alexander Militarev* — doctor of sciences (Philology); external professor, Institute of Linguistics, Russian Academy of Sciences (Moscow), amilitarev@gmail.com

*Maarten Mous* — Professor Emeritus, Centre for Linguistics, Leiden University, m.mous@hum.leidenuniv.nl

*Andrey Nikulin* — Ph.D., faculty member, Núcleo Takinahakỹ de Formação Superior Indígena, Federal University of Goiás, andre.n.guzman@gmail.com

*Bonny Sands* — Ph.D., faculty member, Northern Arizona University, Bonny.Sands@nau.edu

*George Starostin* — candidate of sciences (Philology); chief research fellow, Institute for Oriental and Classical Studies, Higher School of Economics (Moscow); external professor, Santa Fe Institute (New Mexico, USA), gstarst1@gmail.com

*Mauro Tosco* — Ph.D., faculty member, Department of Human Studies, University of Turin; mauro.tosco@unito.it

*Mikhail Zhivlov* — candidate of sciences (Philology); lecturer, Institute for Oriental and Classical Studies, Russian State University for the Humanities; senior researcher, Institute for Oriental and Classical Studies, HSE (Moscow), zhivlov@gmail.com

## Сведения об авторах

*Живлов, Михаил Александрович* — канд. филол. наук; доцент Сектора компаративистики Института восточных культур и античности РГГУ, старший научный сотрудник Института классического Востока и античности ВШЭ (Москва), zhivlov@gmail.com

*Лоренц, Марина Максимовна* — аспирант Института лингвистики РГГУ (Москва), lorentzmarina8@gmail.com

*Милитарев, Александр Юрьевич* — доктор филол. наук; внештатный сотрудник Института языкознания РАН (Москва), amilitarev@gmail.com

*Моус, Маартен* — заслуженный профессор Лейденского университета (Нидерланды), m.mous@hum.leidenuniv.nl

*Никудин, Андрей Владимирович* — Ph.D., сотрудник федерального университета штата Гояс, andre.n.guzman@gmail.com

*Старостин, Георгий Сергеевич* — канд. филол. наук; ведущий научный сотрудник Института классического Востока и античности ВШЭ (Москва), внешний сотрудник Института Санта-Фе (Нью-Мексико, США), gstarst1@gmail.com

*Сэндс, Бонни* — Ph.D., сотрудник Университета Северной Аризоны (США), Bonny.Sands@nau.edu

*Тоско, Мауро* — Ph.D., сотрудник отдела гуманитарных исследований Туринского университета (Италия); mauro.tosco@unito.it

*Харви, Эндрю* — Ph.D., младший профессор Байройтского университета (Германия), andrew.harvey@uni-bayreuth.de

## 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 прикладываются 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»

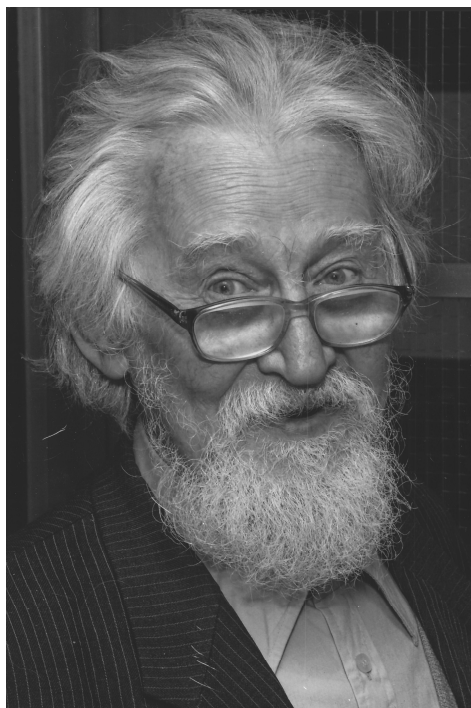
## От редакции

Уже в процессе подготовки материалов к текущему номеру, 7 мая 2023 г. ушел из жизни Владимир Антонович Дыбо, главный редактор журнала «Вопросы языкового родства» с момента его основания в 2009 г. Редакционный совет и редакционная коллегия журнала (для многих из членов которых Владимир Антонович был больше чем просто коллега — учитель, вдохновитель, почти идеальный образец лингвиста-компаративиста) приняли совместное решение почтить память выдающегося ученого специальным памятным номером, выпуск которого планируется организовать в 2024 г.; здесь же мы, с разрешения автора, публикуем некролог, написанный одним из благодарных учеников Владимира Антоновича — профессором Загребского университета Мате Каповичем.

## From the editors

As the first materials for this issue of the Journal were being assembled, the sad news came in about the passing, on May 7, 2023, of Vladimir Antonovich Dybo, the editor-in-chief of the Journal of Language Relationship since its founding in 2009. The Advisory Board and the Editorial Staff of the Journal, many of whose members saw Vladimir Dybo as so much more than just a colleague — a teacher, an influencer, a near-perfect model of a classic scholar of historical linguistics — have decided to commemorate the passing of this eminent researcher with a special issue in his memory, to be published some time in 2024. In the meantime, with the grateful permission of the author, we offer an obituary written by one of the (informal) disciples of Vladimir Antonovich, professor Mate Kapović of the University of Zagreb.

## Vladimir Antonovič Dybo (1931–2023): In memoriam



I first heard of Vladimir A. Dybo as a young student of linguistics at the University of Zagreb, interested in historical linguistics, sometimes in the early 2000's, together with the fascinating concept of accentual paradigms. In those still early days of the Internet the now ubiquitous book PDFs were still not a thing and it was often very difficult to obtain certain foreign books, especially if you were just a student. Still, I managed to get a copy of Dybo's famous 1981 monograph<sup>1</sup> from a professor and became instantly enthralled with it. The wealth of data from various Slavic languages and old manuscripts together with reconstructions, footnotes, small print and tables were captivating.

In 2005, now a young research assistant at the Department of Linguistics, I organized a conference called "International Workshop on Balto-Slavic Accentology" (IWoBA) in Zagreb. Dybo was one of the accentologists to show up at the conference and undoubtedly a star of the whole gathering, though it was hard to describe him as star due to his humble and unassertive demeanor. During the conference, a trip to the Trakošćan castle was organized, which is situated in the area of Croatia where the famous Bednja Kajkavian local dialect<sup>2</sup> (at the same time very archaic when it comes to accentuation and very innovative when it comes to vocalism) was spoken. Marc Greenberg and I even managed to record some dialectal material from a native Bednja speaker who lived just below the castle during the excursion. Upon hearing of this, Dybo, who himself used the Bednja accentological material in his works very often and was highly familiar with it, asked if he could listen to our recordings on the bus on the way back to Zagreb.

Next time I saw Dybo in person was in 2006 at the second IWoBA organized at the University of Copenhagen (and we continued to see each other regularly at numerous following IWoBA conferences in various European cities). I gave him one of my early papers on accentology (back then, authors still got article offprints) and he was kind enough to inform me the very next day that he had already read it and praised the article. I also remember asking him about the Moscow accentological school "new approach"<sup>3</sup> in a coffee shop and he gave me a lengthy and patient explanation of his views on the matter.

---

<sup>1</sup> Дыбо, Владимир А. 1981. *Славянская акцентология. Опыт реконструкции системы акцентных парадигм в праславянском* [Slavic accentology. An attempt to reconstruct the system of accentual paradigms in Proto-Slavic]. Москва: Издательство «Наука».

<sup>2</sup> Jedvaj, Josip. 1956. Bednjanski govor [Bednja local dialect]. *Hrvatski dijalektološki zbornik* 1: 279–330.

<sup>3</sup> Cf. e.g. Hendriks, Pepijn. 2003. A Note on Stang's Law in Moscow Accentology. *Dutch Contributions to the Thirteenth International Congress of Slavists, Ljubljana. Linguistics* (SSGL 30): 107–123.

In 2006, on the recommendation of Dybo, I was approached by his former student and long-time collaborator Sergei L. Nikolaev, to go on a dialectological-accentological excursion to the famous Old Štokavian Posavina region<sup>4</sup> in Croatia. We went on three such excursions (in 2006, 2007 and 2010) and Dybo himself, though already advanced in age, took part in 2007 and 2010 (together with other colleagues: Martina Peraić, Marfa N. Tolstaja, Mikhail V. Oslon and Aleksandra V. Ter-Avanesova). All three excursions were highly memorable and we managed to record a lot of new material from Posavina (some of which is still unpublished, unfortunately).

I was roommates with Dybo during one of our excursions and I have fond memories of that. Dybo was a man of amazing energy — I remember one time seeing him around midnight by his laptop reading something in the pdf of Pedersen’s comparative Celtic grammar. Though energetic in this regard he was always calm — while Nikolaev and I, both temperamental, would get into occasional heated discussions and even shouting matches about accentology, Dybo remained calm at all times and offered his opinions serenely. He was also not what the Germans would call *Fachidiot* — I recall a discussion at a balcony in Slavonski Brod, where he, at my astonishment, tackled in detail various aspects of the value theory of labor.

Dybo was the most prominent scholar of the Moscow accentological school (abbreviated as MAS). Dybo’s (and the MAS) methodology is first and foremost based on a thorough analysis of data from a wide array of primary sources — including different Slavic languages, dialects and old manuscripts with accentual markings<sup>5</sup>. Dybo’s works are thus typified by putting forward huge amounts of evidence from a large number of Slavic sources — thus, even if one would not always agree with Dybo, one could not only check the basis for his reconstructions but also use the material to form one’s own conclusions. The reconstruction of accentual paradigm distribution for specific words (i.e. ascribing the original accentual paradigm to every reconstructed Slavic word) was also an important part of the said methodology. Slavic accent was always analyzed within the frame of morphology<sup>6</sup> and word-formation, which was very important for the reconstruction of the Balto-Slavic accentual system and for Dybo’s influential and ingenious valence theory<sup>7</sup>.

The already mentioned 1981 monography by Dybo is nowadays the classic reconstruction of the Proto-Slavic (or Common Slavic — depending on one’s terminology) accentual system, accepted by almost all researchers and probably the most cited work on Slavic accentology. Most western accentologists, regardless on what they think about the earlier stages of

<sup>4</sup> Cf. Ivšić, Stjepan. 1913. Današnji posavski govor [The present-day Posavina dialect]. *Rad JAZU* 196 (I): 124–254, 197 (II): 9–138. The dialectal material from this work was also often used by Dybo in his articles and books.

<sup>5</sup> In 2010 in Slavonia (Croatia), Dybo told a couple of us about his adventures in 1965, when he went from the USSR to Yugoslavia (which was not always easy at that time) to look for Petretić’s Old Kajkavian manuscripts. Later, he used that material in his works — cf. e.g. pp. 564–565 in his 2000 monograph (Дыбо, Владимир А. 2000. *Морфонологизированные парадигматические акцентные системы. Типология и генезис, Том I* [Morphologization of a paradigmatic accentual system. Typology and genesis. Volume I]. Москва: Языки русской культуры).

<sup>6</sup> Cf. the indicative name of Lehfeldt’s monograph on the MAS approach to Slavic accentology: Lehfeldt, Werner. 1993<sup>1</sup>/2001<sup>2</sup>. *Einführung in die morphologische Konzeption der slavischen Akzentologie* [Introduction to the morphological approach to Slavic accentology], München: Otto Sagner. In 2006, during the Copenhagen IWoBA, Dybo provided all the participants with the draft of his overview of Lehfeldt’s monograph: Дыбо, Владимир А. 2006. Сравнительно-историческая акцентология, новый взгляд: по поводу книги В. Лефельдта 'Введение в морфологическую концепцию славянской акцентологии' [Comparative-historical accentology, the new look: on occasion of the book of W. Lehfeldt “Introduction to the morphological approach to Slavic accentology”], *Вопросы языкознания* 2: 3–27.

<sup>7</sup> Cf. also Dybo, V. A., S. L. Nikolayev & S. A. Starostin. 1978. A tonological hypothesis on the origin of paradigmatic accent systems. *Estonian papers in phonetics* 1978: 16–20.

(Balto-)Slavic and Indo-European accentuation, in effect take Dybo's reconstructions as a given (almost as if they were themselves attested — like Old Church Slavic) and proceed from there to try to connect the Slavic accentual system with the wider Indo-European frame. Dybo's other monographs<sup>8</sup> also remain very important and frequently referenced to, while the sheer number of his papers on accentology is huge<sup>9</sup>. In a time when more and more linguists read only in English, even the fact that almost all of his works were in Russian (except for a few translations) was not a major hinderance.

The death of Vladimir A. Dybo is without a doubt a blow for the whole field of Slavic accentology. After Stang's (1900—1977) revolution, which started with Stang 1957<sup>10</sup>, and the early and important work by Dybo's friend and colleague Vladislav M. Illič-Svityč (1934—1966) in 1963<sup>11</sup>, it was Dybo who was undisputedly a towering figure in the field of (Balto-)Slavic accentology for more than 60 years in the second half of the 20<sup>th</sup> and the beginning of the 21<sup>st</sup> century (Dybo's first article was published in 1958 and the last two in 2019). Without Vladimir A. Dybo, the modern field of Balto-Slavic (but also Indo-European) accentology would not look as it looks today. He was one of the founders of the field, he will be remembered dearly and his works will continue to be a valuable source and inspiration for further generations of Slavic historical accentologists and linguists.

---

<sup>8</sup> Besides the already mentioned 1981 and 2000 monograph, there are two more done in coauthorship: Дыбо, В. А., Г. И. Замятина & С. Л. Николаев. 1990, *Основы славянской акцентологии* [Basics of Slavic accentology], Москва: Издательство «Наука» and Дыбо, В. А., Г. И. Замятина & С. Л. Николаев. 1993. *Основы славянской акцентологии. Словарь* [Basics of Slavic accentology. Dictionary]; Москва: Издательство «Наука».

<sup>9</sup> A full bibliography was put together by Mikhail Oslon: <https://tromanes.org/pub/Dybo/Библиография%20работ%20В.А.Дыбо.pdf>.

<sup>10</sup> Stang, Christian S. 1957. *Slavonic accentuation*, Oslo: I kommisjon hos H. Aschehoug & Co. (W. Nygaard).

<sup>11</sup> Иллич-Свитыч, Владислав М. 1963. *Именная акцентуация в балтийском и славянском. Судьба акцентуационных парадигм* [Nominal accentuation in Baltic and Slavic. The fate of accentual paradigms], Москва: Издательство Академии наук СССР (English translation: Illich-Svitych, Vladislav M. 1979. *Nominal Accentuation in Baltic and Slavic*, Cambridge / London: The MIT Press).

**Issue 1 / Часть 1**

**Articles / Статьи**





## Lexical evidence for the Macro-Jê–Tupian hypothesis

The Macro-Jê and Tupian language families of Eastern South America have long been thought to be distantly related, mainly based on morphological evidence. This article assembles lexical evidence for the Macro-Jê–Tupian hypothesis. Reconstructed Proto-Macro-Jê and Proto-Tupian forms are compared, with special attention to the distribution of the etyma in each family, morphosyntactic behavior of the comparanda, and semantic and phonological plausibility of the proposed etymologies. Although the total number of possible cognates is very limited, the fact that they show recurrent sound correspondences renders the Macro-Jê–Tupian hypothesis promising and worthy of further research.

**Keywords:** Macro-Jê; Tupian; comparative method; South American indigenous languages.

The goal of this contribution is to present lexical evidence for the hypothesis whereby the Macro-Jê and Tupian languages are considered to be distantly related. Macro-Jê and Tupian are two major language families of Eastern South America, whose geographic spread coincides to a great extent. Both are present south of the Amazon River in what is now Brazil and Eastern Bolivia. Northern Argentina, Paraguay, and (formerly) Uruguay are home to a few peoples that speak Tupian languages of the Guaranian branch, though in the past two Macro-Jê languages—Ingain and Kaingang—were spoken there, too. In addition, due to post-Columbian migrations a few Tupian languages—Wajãpi, Teko, Zo'e, and Nheengatu—, are now spoken north of the Amazon River in French Guiana, Brazil, and Venezuela.

Sections 1 and 2 present the Macro-Jê and Tupian families, respectively, with an emphasis on the state-of-the-art reconstructions of the respective protolanguages. Section 3 surveys the extant scholarship on the Macro-Jê–Tupian hypothesis. The potential cognates are discussed in section 4, and the respective sound correspondences are dealt with in section 5. Section 6 concludes the paper. The provenance of linguistic data is indicated at the end of the paper, before the list of abbreviations and the acknowledgments.

Throughout this article, I employ the Macro-Jê Alphabet (Nikulin 2020: 50–53) for reconstructed forms of Proto-Macro-Jê and other Macro-Jê (proto)languages that lack an established practical orthography. For Proto-Tupian and Proto-Cariban, the International Phonetic Alphabet is used, except that \*/ɾ/, \*/β/, \*/ε/ are written as \*r, \*β, \*e. Practical orthography is used for contemporary or historically attested languages, when possible. In Mundurukú, Yudja, and Mondé forms, tone is indicated despite being unmarked in the respective practical orthographies (´ for high tone, ` for creaky voice; cf. Pinheiro et al. 2020); in Tuparí, the stress position is likewise indicated by means of an acute accent.

### 1. Macro-Jê

Approximately 35 languages are classified as Macro-Jê, though only 12 of these (Karajá, Maxakalí, Laklãnõ, Kaingang, Akwẽ-Xerente, Xavante, Panará, Pykobjê–Křĩkatí, Canela–Krahô, Apinajé, Mẽbêngôkre, and Khĩsêtjê) currently serve as the main means of communication in the respective communities. Their classification is shown in Figure 1. Note that **Southern Ka-**

**makã** is a cluster composed of three dialects or closely related languages (Menien, Kotoxó = Mongoyó, and Kamakã proper); **Core Maxakalian** includes at least six varieties (Maxakalí = Tikmũ'ũn, Ritual Maxakalí, Makoní, Pataxó, Pataxó-Hãhãhãe, and Koropó); **Southern Jê** includes two languages, Kaingang (with its at least five dialects) and Laklãnõ = Xokleng; **Akuwẽ** includes four languages (Xavante, Akwẽ-Xerente, Xakriabá = Krêká, and Akroá); **Northern Jê** includes ca. 7 languages (Kajkwakhrattxi, Khĩsêtjê, Mëbêngôkre with its two extant dialects, Apinajé, Parkatêjê, Pykobjê–Krĩkatí with its two dialects, and Canela–Krahô with its no less than three dialects); **Karajá** has four dialects (Southern Iny, Northern Iny, Javaé, and Xambioá = Ixỹbiòwa); **Chiquitano** is composed of three dialects, or maybe three closely related languages (Bésiro, Migueleño, and Eastern).

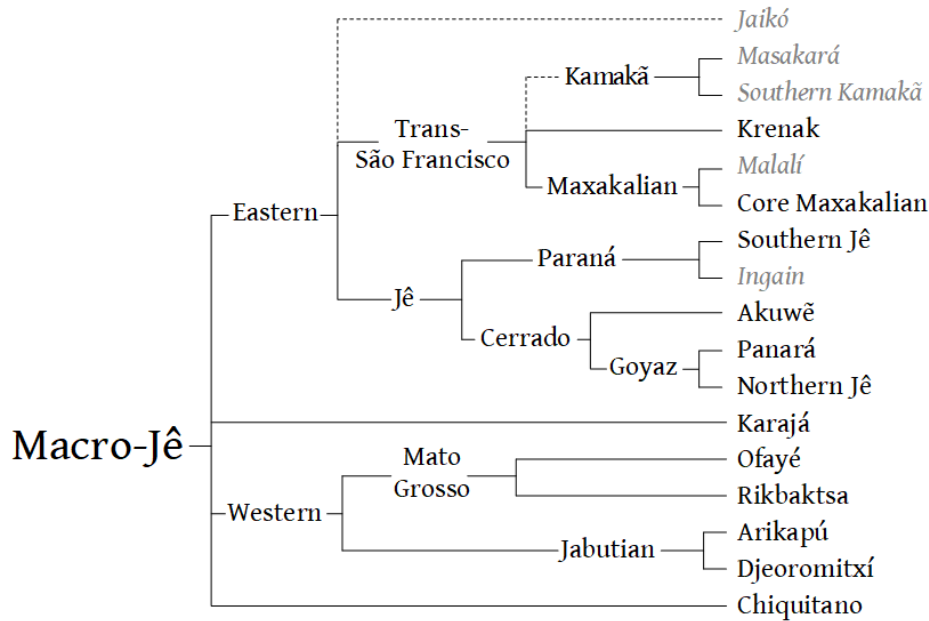


Figure 1. Macro-Jê *Stammbaum* (adapted from Nikulin 2020: 178)<sup>1</sup>

The only extant study that deals with the reconstruction of Proto-Macro-Jê phonology, lexicon, and morphology is Nikulin 2020. In that proposal, 11 consonants (\**/p m w t n r c ñ j k ŋ/*) and at least 16 vowels (\**/a â ə ã õ y ÿ o ô u ã e ê i ĩ/*)—and possibly even more, as indicated by subscript digits—are reconstructed for Proto-Macro-Jê. The maximal syllable was \*/CrVC<sup>o</sup>/, where /<sup>o</sup>/ stands for the so-called echo vowel<sup>2</sup>. Complex onsets were composed of a peripheral (labial or velar) non-continuant and a rhotic: \*/pr mr kr ŋr/. Underlying nasal onsets surfaced as postoralized preceding an oral nucleus: \*/m mr n ñ ŋ ŋr/ were thus pronounced as \*[mb mbr nd ŋ ŋ ŋr] before oral vowels. For example, PMJ \*/mi<sub>1</sub>n<sup>o</sup>/ ‘water’ was likely pronounced as \*[mbini]. In Nikulin’s (2020) PMJ reconstructions, these allophonic realizations are represented by means of the combinations \**mb*, \**mbr*, \**nd*, \**nŋ*, \**ŋg*, \**ŋgr*, as in \**mbi<sub>1</sub>n<sup>o</sup>*. Likewise,

<sup>1</sup> The classification presented here differs from Nikulin 2020 in that Chiquitano is considered here a branch of Macro-Jê rather than an outgroup. This change is motivated by the absence of clear innovations that would define non-Chiquitano Macro-Jê languages as a clade. The labels in gray italics refer to scantily attested languages.

<sup>2</sup> A reviewer has inquired whether postnuclear consonants followed by an echo vowel are syllabified as codas or onsets. The answer depends on the level of analysis. On the surface, the echo vowel is indeed realized as a regular segment, with the preceding consonant syllabified as its onset (at least in some daughter languages). However, the underlying status of the echo vowels is less clear. It may be argued that their occurrence is best represented by a timing-related feature, whereby the release of the nucleus gesture is delayed until the release of the coda gesture.

underlying \*/j/ surfaced as \*[ɲ] preceding a nasal nucleus, as in the genitive adposition PMJ \*/-jũk/ \*[-ɲũk]. This allophone is represented as \*ñ in Nikulin’s (2020) PMJ reconstructions, as in \*-ñũk.

Proto-Macro-Jê was a head-final language. An important fact about its morphosyntax is that PMJ stems were subdivided into two classes, known as relational and absolute stems. Relational stems required their internal argument to be expressed immediately to the left of the stem, either as a noun phrase or as a person index of the so-called internal series (one of \*a- 2, \*i- 3NCRF, \*ta- 3CRF). Note that the internal series lacked dedicated first-person indices, and pronouns were employed instead for expressing first-person internal arguments. Conversely, absolute stems lacked an internal argument and thus were not capable of taking person indices. Nouns, verbs, and adverbs/adpositions were lexically specified either as absolute or as relational; relational stems are indicated by means of a hyphen before the stem.

Another important division, which cross-cuts all relational stems, is whether their initial segment was the thematic consonant \*/j/ followed by a vowel (class II stems) or not (class I stems). The thematic consonant \*/j/ in class II stems was deleted upon the accretion of a person index; the person indices, in turn, had special allomorphs in class II stems: \*∅- 2, \*c- 3NCRF, \*t- 3CRF). Class I stems started with consonants other than \*/j/. It is tempting to analyze class II stems as underlyingly vowel-initial (cf. Rodrigues 2012), but Salanova (2011) shows that the thematic consonant \*/j/ is best understood as a part of the stem in at least some Macro-Jê languages.

Proto-Macro-Jê roots are commonly monosyllabic, though some disyllabic roots can be reconstructed as well. A frequent evolution pathway, especially common in Jê and Chiquitano, is the fossilization of prefixes or incorporated roots, whose semantics cannot be identified with precision at all times, at the left margin of stems, especially verbal ones. These fossilized elements have been variously labeled as formatives (Oliveira 2005: 82) or transitivity prefixes (Nikulin & Salanova 2019: 539–540) in Jê studies, and as classifiers in Chiquitano studies (Ciucci 2020).

One outstanding aspect of Proto-Macro-Jê phonology is the frequent occurrence of stem-final consonants, which may be followed or not by an echo vowel. These consonants were often lost in many contemporary languages. Nikulin & Silva (2020) establish that three branches of Macro-Jê are useful for reconstructing PMJ codas. **Maxakalí** (alongside other Maxakalian languages) is particularly conservative regarding the place of articulation of Proto-Macro-Jê codas, but not their manner of articulation or the echo vowels; synchronically, the language distinguishes between four codas, symbolized as /P T C K/, which are underspecified for features other than place of articulation (Silva 2015, 2020). **Krenak** is conservative in that it preserves stops as stops and nasals as nasals in the coda position, but erstwhile alveolar codas merge with velar ones (\*-t, \*-n > -k, -ŋ), and erstwhile palatal codas become alveolar (\*-c, \*-ñ > -t, -n); echo vowels are not preserved. **Proto-Jê** preserves most PMJ codas intact, but many of them are lost in individual Jê branches, sometimes leaving traces such as vowel lengthening<sup>3</sup> or morphophonological alternations. For the development of codas in other branches of Macro-Jê, see Nikulin (2020: 158sqq.). Taking PMJ codas into account is crucial for any attempts at exploring the external connections of Macro-Jê, especially given the fact that most

<sup>3</sup> This is a novel finding, not described in Nikulin & Silva (2020) or Nikulin (2020). More specifically, the nasal codas \*-n and \*-ñ followed by echo vowels are deleted in the Goyaz branch with compensatory lengthening, as in PCerr \*pryn ‘road’, \*-mbyn ‘tail’, \*-mbên ‘liquid’, \*-jarên ‘root’, \*-ŋgôn ‘louse’, \*mãñ ‘greater ema’, \*-jwañ ‘tooth’, \*-kwañ ‘hole’, \*-ŋgoñ ‘wet’, \*-ŋgrôñ ‘embers’ > Proto-Goyaz \*pry:, \*-mby: ‘penis’ (cf. \*-jamby: ‘tail’), \*-mbê:, \*-jarê:, \*-ŋgô:, \*-mã:, \*-jwa:, \*-kwa:, \*-ŋgo:, \*-ŋgrô:). Note that in the Macro-Jê Alphabet echo vowels are unmarked in PCerr reconstructions (by contrast, their absence is marked by means of an apostrophe). The vowel length is most consistently reflected in Pykobjê–Krikatí as documented by Pries (2008).

Proto-Macro-Jê roots are monosyllabic. Evidently matches involving  $*C(r)VC$  structures are more reliable for demonstrating common origin of languages than those involving  $*C(r)V$  structures.

As shown in Figure 1 above, the Macro-Jê family is currently thought to include four first-level branches. The **Eastern** branch includes Jê, Maxakalian, Krenak, and possibly the poorly known Jaikó and Kamakã languages. These languages, except for a few Jê languages, are spoken east of the Araguaia River. The **Western** branch includes several languages spoken west of the Araguaia River: Rikbaktsa, Ofayé, and the Jabutian languages Djeoromitxí and Arikapú. **Karajá**, spoken along the Araguaia River, and **Chiquitano**, spoken mostly in the Chiquitano Dry Forest region in Bolivia and adjacent areas of Brazil, do not appear to form a clade with any other Macro-Jê group. Therefore, I consider that a given form can be reconstructed for Proto-Macro-Jê if its reflexes are present in at least two major subdivisions of Macro-Jê (Eastern, Western, Karajá, or Chiquitano).

## 2. Proto-Tupian

The Tupian language family includes approximately 70 languages, of which ca. 45 serve as primary means of communication in the respective communities. The subgrouping of Tupian is shown in Figure 2.

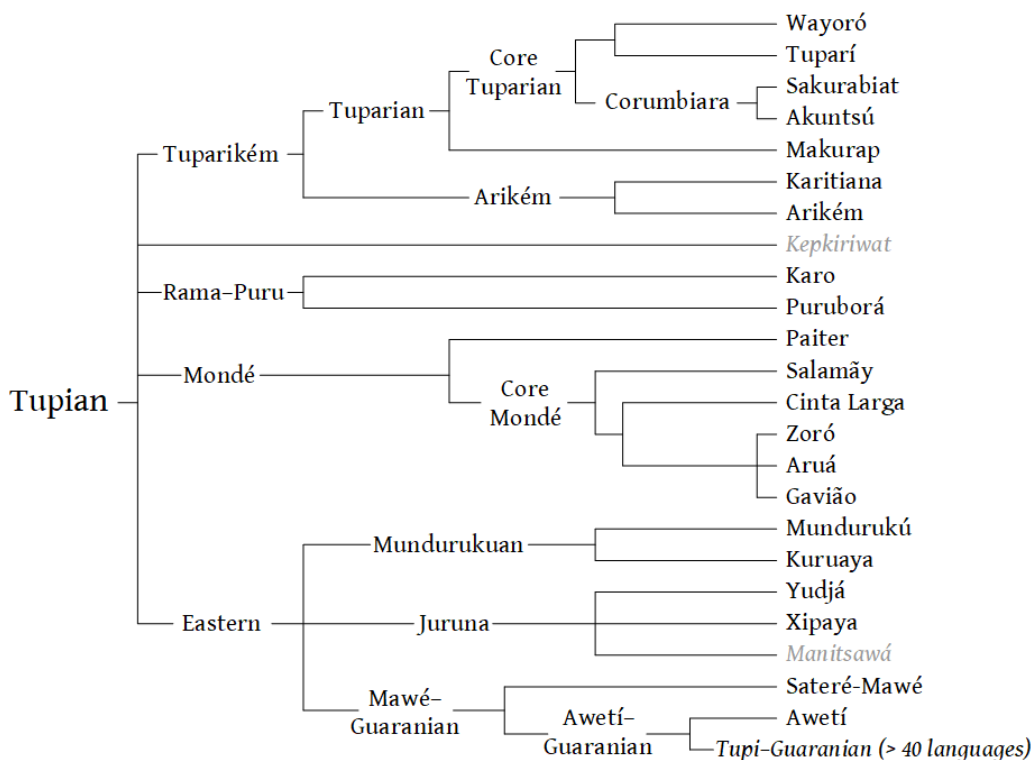


Figure 2. Tupian Stammbaum (based on Nikulin & Carvalho 2022: 20–21)<sup>4</sup>

<sup>4</sup> Nikulin and Carvalho’s (2022) proposal differs from a more conservative proposal by Galucio et al. (2015) in that it posits a clade consisting of Tuparian and Arikém (based on three shared innovations involving Proto-Tupian  $*i$  and  $*ə$ ), dubbed “Tuparikém”, and reinstates the so-called Eastern clade, originally proposed by Rodrigues (2005).

For a significant period of time, the only attempt at a phonological and lexical reconstruction of Proto-Tupian had remained that of Aryon Dall’Igna Rodrigues, with an early version thereof found already in Hanke et al. 1958. Its elements are presented in a significant number of publications by Rodrigues and his students, with Rodrigues 2005, 2007 and Corrêa-da-Silva 2010 being the most complete sources. Rodrigues’ proposal has been criticized for failing to follow the principles of bottom-up reconstruction; for his overreliance on data of one single branch, Tupi–Guaranian, and especially the Old Tupí language; for misrepresentation of phonological facts of specific languages; and for positing typologically implausible developments (Meira & Drude 2015: 290–291; Singerman 2018: 390–392; Nikulin & Carvalho 2019: 276–278, 2022, among others). Moreover, recent years have seen considerable progress in phonological and lexical reconstruction of the protolanguages of individual Tupian branches, such as Proto-Tupi–Guaranian (Carvalho 2022, 2023, *forthc.*; Carvalho & Birchall 2022); Proto-Mawé–Guaranian (Meira & Drude 2015), Proto-Mundurukuan (Picanço 2019), Proto-Juruna (Fargetti & Rodrigues 2008, 2021; Carvalho 2019), Proto-Tuparian (Galucio & Nogueira 2012; Nogueira et al. 2019; Nikulin & Andrade 2020), Proto-Tuparikém (Nikulin *forthc.*). Taking into account recent progress in comparative studies of Tupian, Nikulin and Carvalho (2022) proposed an updated reconstruction of the sound system of Proto-Tupian, with an emphasis on the vowel system, though the reconstruction of Proto-Tupian consonants was also updated with respect to Rodrigues’ (2007) proposal.

The inventory of Proto-Tupian onsets posited by Nikulin and Carvalho (2022) includes \*/p m β w t n ð r tʰ c j k kʰ k̟ ʔ/. Of these, the consonant \*/tʰ/ is rare but well-supported, while the reconstruction of \*/β/ and \*/ð/ is more dubious. The phonological and phonetic properties of \*/k kʰ k̟/ are a matter of speculation. PT \*/k/ yields velar reflexes in all branches; \*/kʰ/ yields velar reflexes in all branches except Tuparian and Kepkiriwat, which reflects it as \*/ʔ/ or zero; \*/k̟/ yields velar reflexes in Tuparian and Kepkiriwat, but \*/ʔ/ or zero in other branches. Since it is unclear whether \*/kʰ/ was actually articulated as a palatalized velar stop, I will henceforth employ the *ad hoc* character \*/k̟̃/ for the character in question; I also replace \*/k̟/ with the *ad hoc* character \*/k̟̃/ so as to avoid unwanted associations with ejective or uvular stops. Similarly to Proto-Macro-Jê, the underlying nasals \*/m n ŋ/ were articulated as postalized \*[mb nd ŋ] before oral vowels; for example, \*/mæC/ ‘snake’ was likely articulated as \*[mbæC̃]. This is represented in Nikulin & Carvalho’s (2022) Proto-Tupian reconstructions (as in \*mbæC), following Moore and Galucio’s (1994: 124) representation conventions for Tuparian. No complex onsets can be reconstructed for Proto-Tupian.

The inventory of Proto-Tupian codas includes only four possibilities: \*/P T C K/. The use of small caps signals that these codas were underspecified for features other than place of articulation, just like in Maxakalí (Silva 2015, 2020). This is still the case in many daughter languages, such as Gavião (Moore 1984: 230), Proto-Tuparian (Moore & Galucio 1994: 123), Sakurabiat (Galucio 1994: 998–992), Puruborá (Galucio 2005: 170–171), Awetí (Drude 2009), Tuparí (Singerman 2016), and many other languages for which such an analysis has never been proposed but is certainly possible. Major deviations from this pattern are found in Juruna, where erstwhile codas are now syllabified as onsets of the following syllables, and in Karo and Mundurukuan, where codas now contrast for nasality. In both Karo and Mundurukuan, codas are usually oral after oral vowels, and nasal after nasal vowels. However, nasal codas also occur after oral vowels, mostly at morpheme boundaries (as is the case with two homonymous suffixes in Mundurukú, *-m* ‘instrumental’ and *-m* ‘imperfective’; Picanço 2005: 158–163), as a result of morphophonological processes (such as /-t-t/ → /-n/ in Karo; Gabas Jr 1999: 58–59), or due to late vowel denasalization, as in Proto-Tupian \*/-jĩ:k > Proto-Mundurukuan \*/-ðiŋ ‘smoke’ (Picanço 2019: 139). Karo is unique among Tupian languages in allowing oral codas after nasal

vowels, as in *-jakōp* ‘warm’ or *-pāt* ‘beautiful’ (Gabas Jr 1999: 49), a fact unaccounted for by Nikulin and Carvalho (2022).

Seven vowel qualities are reconstructed for Proto-Tupian: \*/a ə i e i o u/. Each of them had a nasal counterpart. This proposal differs from the traditional reconstruction in Rodrigues 2005, which posited only six vowel qualities (\*/a i e i o u/), in having \*/ə/ instead of his \*/o/ (and sometimes \*/e/), \*/i u/ instead of his \*/i/, and \*/o/ instead of his \*/u/. There is evidence that vowel length may have been contrastive; it is best preserved in the Tuparikém branch and possibly in Sateré-Mawé and Mondé.

Mundurukuan, Juruna, Mondé, and Karo languages are tonal, and there is evidence that pitch accent may be contrastive in Makurap; in addition, lexically specified stress has been described for Tuparí and Akuntsú (see Nikulin & Andrade 2020: 286). This suggests that Proto-Tupian may have also been a tonal language, but no attempts have been made at reconstructing its prosody.

A typical Proto-Tupian morpheme had one or two syllables, and morpheme-internal codas appear to have been rare (though existent, as in *\*jaCjo* ‘armadillo’). Just like in Proto-Macro-Jê, stems were subdivided into relational and absolute, with relational stems obligatorily taking a complement immediately to its left (signaled by means of a hyphen before the stem), and absolute stems disallowing them<sup>5</sup>. The class of relational stems was further subdivided into two classes. Class I stems started with consonants, took the allomorph *\*i-* of the third-person index, and did not require any thematic element when their internal argument was expressed by means of a noun phrase. Class II stems, conversely, took the allomorph *\*c-* of the third-person index, and occurred with the thematic consonant *\*j-* when their internal argument was expressed by means of a noun phrase (or a person index other than the third-person one). The original configuration is most faithfully preserved in Makurap, Mundurukuan, and Sateré-Mawé.

Although the differences between Rodrigues’ (2005, 2007) and Nikulin and Carvalho’s (2022) proposals are significant, the consequences of preferring one proposal over the other are minimal for the purposes of establishing long-range connections with other families. This is so because most differences are related to the manner of articulation of the consonants and to specific vowel features, but the reconstructed forms are still quite similar across proposals, as shown in Table 1.

|                    | ‘to grind’ = ‘larva’ | ‘arrow’              | ‘leaf’           | ‘to seize’ | ‘door’              | ‘armadillo’ |
|--------------------|----------------------|----------------------|------------------|------------|---------------------|-------------|
| Rodrigues          | *ček <sup>w</sup>    | *ek <sup>wə</sup> ip | *ep <sup>w</sup> | *pičik     | *ek <sup>w</sup> en | *tajtu      |
| Nikulin & Carvalho | *-təK                | *əkup / *-jəkup      | *əP / *-jəP      | *-pitik    | *ək-ēT / *-jək-ēT   | *jaCjo      |

Table 1. Rodrigues’ (2005, 2007) and Nikulin & Carvalho’s (2022) Proto-Tupian reconstructions

As shown in Figure 2 above, the Tupian family is currently maintained to include no less than five first-level branches. The **Eastern** branch includes the Juruna, Mundurukuan, and Mawé-Guaranian groups (the latter is further subdivided into Sateré-Mawé and Awetí-Guaranian, and Awetí-Guaranian is in turn subdivided into Awetí and Tupi-Guaranian). This branch reaches its highest diversity between the lower Madeira and the lower Iriri Rivers. The **Tuparikém** branch includes the Tuparian and Arikém groups, which includes seven lan-

<sup>5</sup> There were also pairs of relational and absolute stems, which some authors have analyzed as constituting an inflectional paradigm. Examples include *\*-pi* ‘foot (rel.)’ / *\*mbi* ‘foot’ (abs.), *\*-ja:pe* ‘path (rel.)’ / *\*pe* ‘path’ (abs.), and *\*-jəK* ‘house (rel.)’ / *\*ək* ‘house’ (abs.). I prefer envisaging such pairs as derivational.

guages spoken in what is now the Brazilian state of Rondônia. The **Mondé** branch includes a handful of languages spoken in Rondônia and in adjacent areas of the Mato Grosso state. The **Rama-Puru** branch includes two languages, Karo and Puruborá, both spoken in Rondônia. The extinct **Kepkiriwat** language was also spoken in Rondônia. The languages of the latter four branches are therefore spoken in the same area, which facilitates lexical diffusion. I consider that a given form can be securely reconstructed for Proto-Tupian if its reflexes are present in the Eastern branch and at least one of the Rondonian branches (Tuparikém, Mondé, Rama-Puru, and Kepkiriwat). If the Eastern branch lacks a cognate, the requirement is that reflexes be present in at least three branches. Cognate sets whose distribution is restricted to two Rondonian branches (say, Tuparikém and Mondé) are likely to involve horizontal transmission. For example, one could technically reconstruct PT *\*-aka:T* or *\*-áka:T* ‘to bite’ based on Karitiana *-okoot*, Paiter *-ákar*, and Salamãý *-áka:l*, but given the absence of cognates outside the Tuparikém and Mondé branches, this verb is unlikely to have existed in Proto-Tupian.

### 3. Macro-Jê–Tupian hypothesis

Possible external relations of Macro-Jê and Tupian are still debated. Both families have figured in a number of partially overlapping long-range proposals, and even the limits of the Macro-Jê family are not universally agreed upon. Macro-Jê languages have been linked to, or claimed to include as a constituent branch, language groups such as Bororoan, Yaathê, Karirian, Purian, Guató, and Otí (Guérios 1939; Davis 1968; Greenberg 1987; Rodrigues 1993, 1999; Ribeiro 2002, 2011; Ribeiro & Voort 2010; Martins 2009, 2011; Nikulin & Carvalho 2018; Silva *forthc.*). Other long-range proposals have connected Macro-Jê to language families such as Cariban (Rodrigues 2000, 2009; Meira et al. 2010: 512–515; Nikulin & Carvalho 2018); Chibchan (Pache 2023); Mapudungun and Katukina–Harakmbut (Adelaar 2008: 11); Mataguyan and Guaicuruan (Viegas Barros 2005; Nikulin & Carvalho 2018), Payaguá and Guachí (Viegas Barros 2005), Zamucoan (Nikulin & Carvalho 2018), and even the putative Nostratic macrofamily (Aikhenvald-Angenot & Angenot 1992). Tupian has been most notably compared with Cariban (Rodrigues 1985, 2000, 2009; Meira et al. 2010: 512–515; Nikulin & Carvalho 2018), but also Bororoan (Nikulin & Carvalho 2018), Yaathê (Silva *forthc.*), Karirian (Ribeiro 2002; Nikulin & Carvalho 2018), Mataguyan, Guaicuruan, and Zamucoan (Nikulin & Carvalho 2018). In addition, the aforementioned families were thought by Greenberg (1987) to be part of a much larger Amerind macrofamily, with Macro-Jê classified as a member of the so-called Ge–Pano–Carib branch, and Tupian as a member of the so-called Equatorial subgroup of the Andean–Equatorial branch.

This study, however, focuses on one specific proposal, whereby Macro-Jê and Tupian are considered to be related to each other (though possibly also to other language families). Although some lexical lookalikes have been identified already by Davis (1968: 47), the most widely known claim on the possible relation of these two families is found in Rodrigues (2000, 2009), who proposes that Macro-Jê, Tupian, and Cariban are all ultimately related (note that in Rodrigues’ definition the Macro-Jê family encompasses language groups such as Bororoan, Purian, Karirian, Yaathê, and Guató, whose inclusion is not supported by Nikulin’s 2020 study). The proposal has had a moderately positive reception in the scholarly community (cf. Meira et al. 2010: 512–515; Ribeiro 2002: 41–42, 2011: 107–109; Nikulin & Carvalho 2018) and sometimes goes by the label “TuKaJê”.

The evidence that substantiates the TuKaJê hypothesis is largely morphological and morphophonological in nature. Most notably, Macro-Jê, Tupian, and Cariban share a pattern

whereby stems capable of taking an internal argument — directly possessable nouns, postpositions, and at least some classes of verbs in at least some constructions — are subdivided into two large classes, commonly referred to as “class I” (which typically includes consonant-initial stems) and “class II” (vowel-initial stems). Class I stems do not undergo any alternations in their paradigm, and they combine with the allomorph *\*i-* of the third-person index in Proto-Macro-Jê, Proto-Tupian, and Proto-Cariban. By contrast, class II stems are preceded by the element *\*j-* when they take an internal argument expressed by a noun phrase in its canonical position (i.e., immediately to the left from the head), again in Proto-Macro-Jê, Proto-Tupian, and Proto-Cariban. This element has been variously analyzed as a so-called “contiguity relational prefix” (in works by Rodrigues and his students), as a “thematic consonant” (Nikulin 2020), or as the initial segment of the stem (Salanova 2011; Meira & Drude 2013, 2015). When the internal argument is expressed by a third-person index, the latter takes the allomorph *\*c-* in Proto-Macro-Jê and Proto-Tupian (*\*∅-* in Proto-Cariban), and the element *\*j-* is not present. With other person indices, *\*j-* may be present or absent depending on the language family and the person. This is illustrated below in example (1) (Proto-Macro-Jê and Proto-Tupian reconstructions are mine; the Proto-Cariban paradigm is from Meira et al. 2010). Note the outstanding similarities in the person indices themselves, which are particularly strong between Tupian and Cariban.

| (1)      | Proto-Macro-Jê                 | Proto-Tupian     |                     | Proto-Cariban      |
|----------|--------------------------------|------------------|---------------------|--------------------|
| class I  | *NP <i>ηgyn</i> <sup>◦</sup>   | *NP <i>ηguP</i>  | ‘NP’s louse’        | *NP C...           |
|          | * <i>i-ηgyn</i> <sup>◦</sup>   | * <i>i-ηguP</i>  | ‘her/his louse’     | * <i>i-C</i> ...   |
|          | * <i>(∅-ηgyn)</i> <sup>◦</sup> | * <i>o-ηguP</i>  | ‘my louse’          | * <i>u-C</i> ...   |
|          | * <i>a-ηgyn</i> <sup>◦</sup>   | * <i>e-ηguP</i>  | ‘your louse’        | * <i>a-C</i> ...   |
|          | * <i>ta-ηgyn</i> <sup>◦</sup>  | * <i>tə-ηguP</i> | ‘her/his own louse’ | * <i>ti-C</i> ...  |
|          | * <i>u-ηgyn</i> <sup>◦</sup>   | —                | ‘our (INCL) lice’   | * <i>ki-C</i> ...  |
| class II | *NP <i>j-uñ</i> <sup>◦</sup>   | *NP <i>j-ãC</i>  | ‘NP’s tooth’        | *NP <i>j-V</i> ... |
|          | * <i>c-uñ</i> <sup>◦</sup>     | * <i>c-ãC</i>    | ‘her/his tooth’     | * <i>∅-V</i> ...   |
|          | * <i>(∅-)j-uñ</i> <sup>◦</sup> | * <i>o-j-ãC</i>  | ‘my tooth’          | * <i>u-j-V</i> ... |
|          | * <i>∅-uñ</i> <sup>◦</sup>     | * <i>e-j-ãC</i>  | ‘your tooth’        | * <i>a-j-V</i> ... |
|          | * <i>t-uñ</i> <sup>◦</sup>     | * <i>tə-j-ãC</i> | ‘her/his own tooth’ | * <i>t-V</i> ...   |
|          | * <i>u-j-uñ</i> <sup>◦</sup>   | —                | ‘our (INCL) teeth’  | * <i>k-V</i> ...   |

Another morphological similarity, identified by Ribeiro (2002: 41–42), involves the morphology employed for converting absolute (unpossessable) nouns to relational (possessable) ones in a subset of Macro-Jê and Tupian languages. In a few languages belonging to the Cerrado branch of the Jê group — Xavante, Akwẽ-Xerente, and possibly Panará — this is attained by means of a prefix or an adposition whose Proto-Cerrado form may be reconstructed as *\*-ñĩm-* (> Xavante *-nhim-/nhib-/nhi-*, Akwẽ-Xerente *-ñĩm*, Panará *-jĩ-*), as shown in (2).

- (2) a. Xavante < Akuwẽ < Cerrado < Jê < Macro-Jê (Estevam 2011: 163)  
*dzeru* → *wa-nhib-dzeru-wawẽ*  
 money 1SG-PSSD-money-AUG  
 ‘money’ ‘our plentiful money’
- b. Akwẽ-Xerente < Akuwẽ < Cerrado < Jê < Macro-Jê (Xerente 2019: 77)  
*tka* → *ĩ-ñĩm= tka*  
 land 1SG-PSSD=land  
 ‘land’ ‘my land’



- c. Panará < Cerrado < Jê < Macro-Jê (Dourado 2001: 72)<sup>6</sup>  
*inkwa* → *kjẽ-mẽra jĩ kwa*  
 house I-PL PSSD house  
 ‘house’ ‘our house’

As for the Tupian family, a likely cognate prefix, dubbed ‘indirect possession mediator’ in Rodrigues et al. 2006: 23, is found in three major branches: Tuparian (Makurap *-xep-* ‘alienable possession marker’), Mundurukuan (Mundurukú *-e-*, bearing high tone after noun phrases and low tone after person indices), and Mawé–Guaranian (Sateré-Mawé *-e-*, or *-he-* after some person indices; Awetí *te-* / *-e-*; PTG *te(p)-* / *-re(p)-*). I follow Rodrigues et al. 2006: 23 in reconstructing its Proto-Tupian form as *\*-ep-*. Its final *\*p* is preserved in Makurap as well as in the TG relational stem for ‘container’ (3i). It is deleted before consonant-initial roots in TG, and before all roots in Mundurukú, Sateré-Mawé, and Awetí. The Makurap, Awetí, and Tupi-Guaranian reflexes suggest the reconstruction *\*jep-* instead of *\*ep-*. One may surmise that reflexes of *\*j-* in the latter set of languages were inserted due to the fact that vowel-initial possessable (relational) stems are otherwise uncommon in Tupian. Some examples follow in 3.

- (3) Makurap < Tuparian < Tupian (Braga 2005: 42–43)
- a. *-pia-t* → *o=xepia-t*  
 -liver-PSSD 1SG=ALZ-liver-PSSD  
 ‘liver’ ‘my liver (an animal’s liver belonging to me)’
- b. *xau* → *o-xep-xau-t*  
 flour-PSSD 1SG=ALZ-flour-PSSD  
 ‘flour’ ‘my flour’

Mundurukú < Mundurukuan < Tupian (Picanço 2005: 259)

- c. *kòbé* → *ayácát é-kòbé*  
 canoe woman PSSD-canoe  
 ‘canoe’ ‘woman’s canoe’
- d. *nobánó* → *wuy-e-nobánó*  
 rifle 1+2-PSSD-rifle  
 ‘rifle’ ‘our (INCL) rifle’

Sateré-Mawé < Mawé–Guaranian < Tupian (Ribeiro 2010: 67, 85, 90, 91)

- e. *kui’a* → *uru-e-kui’a*  
 bowl 1+2-PSSD-bowl  
 ‘calabash bowl’ ‘our (INCL) calabash bowl’
- f. *sokpe* → *u-he-sokpe*  
 clothes 1-PSSD-clothes  
 ‘clothes’ ‘my clothes’

<sup>6</sup> Dourado (2001: 71–72) claims that *-jĩ* is only found in elders’ speech, and that the more common genitive postposition (or rather a genitive noun in her analysis) is *-jõ*, with cognates all across Macro-Jê (Ribeiro 2002, 2009) that reflect Proto-Macro-Jê *\*-ñũk* (Nikulín 2020: 404). However, the very existence of Panará *-jĩ* is doubtful: all instances of this form in the cited word are accompanied with the noun transcribed as *koa* by Dourado (2001: 71–72, 77), whose form is attested as *inkwa* /*ɲwa*/ [ɲkwa] in more recent works (Bardagil-Mas 2018: 51). It is thus possible that the combination *-jĩ kwa* in Dourado (2001) is simply a phonetic variant, or even a mistranscription, of *-jõ inkwa*. More recent sources on Panará do not report the existence of *-jĩ* either.

Awetí < Awetí–Guaranian < Mawé–Guaranian < Tupian (Drude 2011: 178)

g. *ky* → *Mopot e-ky*  
 ax Mopot PSSD-ax  
 ‘ax’ ‘Mopot’s ax’

Old Tupí < TG < Awetí–Guaranian < Mawé–Guaranian < Tupian (Barbosa 1956: 110–111)

h. *mbetar-a* → *te-mbetar-a*  
 tembetá-REF PSSD-tembetá-REF  
 ‘tembetá’ ‘one’s tembetá’

i. *uru-∅* → *abá rep-uru-∅*  
 container-REF person PSSD-container-REF  
 ‘container’ ‘indigenous person’s container’

Despite the morphosyntactic, semantic, and phonological similarity between the aforementioned morphemes, Ribeiro’s (2002: 41–42) hypothesis is rendered less plausible than it could have been by the very limited distribution of *\*-ñĩm-* on the Macro-Jê side of the comparison: its reflexes are only found in the Cerrado branch of the Jê group (or, if the alleged Panará reflex is shown to be a ghost morpheme—as suggested in footnote 6—, in its Akuwẽ subbranch), and an entirely different postposition *\*-ñũk* is reconstructed in the same meaning for Proto-Macro-Jê.

Finally, Rodrigues (2000: 101), Ribeiro (2002: 42), and Rodrigues et al. (2006: 34–35) point out the similarity between morphemes of similar shape in some Macro-Jê languages (Xavante *-nhimi-*, Akwẽ-Xerente *-nmĩ-* < Proto-Akuwẽ *\*-ñĩmĩ-*) and in some Tupian languages (PTG unpossessable *\*mbi-*, possessable *\*te-mbi-* / *\*-re-mbi-*, with cognates in Awetí, Sateré-Mawé, and possibly other branches, such as Tuparian), whose function has been variously described as a patient nominalizer or an antipassive nominalizer. In both language families, it attaches to transitive verbs (more specifically, to their nonfinite forms in the case of the Macro-Jê languages), and takes a possessor encoding the notional agent of the verb. This is unusual, since in both language families it is typically the absolutive participant — and not the ergative one — that shares the coding strategy with possessors of nouns. Cf. the illustrations in (4):

(4) a. Xavante < Akuwẽ < Cerrado < Jê < Macro-Jê (Estevam 2011: 330)

*romhu-ri* → *wa-nhimi-romhu-ri*  
 work-NF 1PL-NMLZ.ANTP-work-NF  
 ‘work.NF’ ‘our work’

b. Sateré-Mawé < Mawé–Guaranian < Tupian (Ribeiro 2010: 67, 71)

*-koi* → *mi-koi*  
 plant<sub>v</sub> NMLZ<sub>p</sub>-plant<sub>v</sub>  
 ‘to plant’ ‘plant (noun)’

c. Apyãwa < TG < Awetí–Guaranian < Mawé–Guaranian < Tupian (Almeida et al. 1983: 35)

*-’o* → *xe-re-mi-’o-∅*  
 eat 1SG-PSSD-NMLZ<sub>p</sub>-eat-REF  
 ‘to eat’ ‘my food’

Once again, the Macro-Jê–Tupian comparison is undermined by the distribution of the alleged cognates on the Macro-Jê side, with reflexes being restricted to the Akuwẽ subbranch of

the Cerrado branch of the Jê group. Prefixes with similar properties in other Macro-Jê languages, such as the Proto-Chiquitano inverse voice marker *\*-ij-*, bear no formal similarity to Proto-Akuwẽ *\*-ñĩmĩ-* or to the Tupian forms.<sup>7</sup>

Regarding the lexical evidence, as mentioned above, Davis (1968: 47) identified ten similarities between his own Proto-Jê reconstructed forms and Proto-Tupian reconstructions extracted from Hanke et al. 1958 as well as forms representing Guajajara, a Tupi–Guaranian language of the Tupian family. Five of his cognate pairs — ‘liver’, ‘husband’, ‘foot’, ‘to eat’, ‘hand’ / ‘arm’ — are accepted as valid in this study, and are discussed in 4.1. The remaining five comparisons are rejected because of a mismatch in stem-final consonants, not always reconstructed by Davis (1968) for Proto-Jê but identified in later comparative work (‘water’, ‘louse’, ‘head’, ‘path’, ‘one’). This is summarized in Table 2. The updated reconstructions are from Nikulin 2020 for the Macro-Jê languages, whereas for the Tupian languages they are based on Nikulin & Carvalho 2022. Guajajara forms have been checked against Harrison & Harrison’s (2013) dictionary.

Rodrigues & Cabral (2010) make another attempt at identifying lookalikes involving Macro-Jê and Tupian languages. These authors take it for granted that languages such as Bororoan, Karirian, Purian, Yaathê, and Guató are part of the Macro-Jê family (cf. Rodrigues 1999), a position not confirmed by more recent studies; as a consequence, multiple proposed cognate sets do not include data of languages classified as Macro-Jê beyond reasonable doubt. A serious shortcoming of Rodrigues & Cabral’s (2010) study is that they consider data of contemporary Macro-Jê languages rather than reconstructed forms. Once the phonological history of individual languages is taken into account, some problems in Rodrigues & Cabral’s (2010) proposal become apparent. A case in point is their comparison of several Kaingang forms containing *f* /*ɸ*/ with Old Tupí forms containing *\*p*. Kaingang *-fa* ‘leg’, *-fór* ‘full’, *-for* ‘thrown away’, *-fo* ‘pus’ (whence *-fó-m* ‘to suppurate’), *-fyr* ‘extremity’, *-fár* ‘skin, bark’, *-fi* ‘to give, to lay’ are thus compared to Old Tupí *-py* ‘foot’, *-por-* ‘full’, *-por-* ‘to jump’, *-peu* ‘pus’, *-apyr-* ‘tip’, *-pir-* ‘skin’, *t-epy* ‘payment, price’. However, as observed by Ribeiro (2004a: 94, fn. 3), Kaingang *f* /*ɸ*/ is known to go back to a coronal consonant, reconstructed in Nikulin 2020 as PSJ *\*θ* < PJ *\*c* < PMJ *\*c*, which entails that the Kaingang–Tupian lookalikes are fortuitous.

The absence of a phonological reconstruction of Proto-Macro-Jê has for long remained a major obstacle in further entertaining the Macro-Jê–Tupian hypothesis. This gap has now been filled (Nikulin 2020), as discussed in section 1. Furthermore, Nikulin and Carvalho (2022) proposed a revision of the reconstruction of Proto-Tupian, as stated in section 2. We are therefore now in position to compare reconstructed Proto-Macro-Jê and Proto-Tupian forms.

---

<sup>7</sup> The Proto-Chiquitano inverse voice marker *\*-ij-* almost certainly goes back to a patient nominalizer, since the Chiquitano verbs in the inverse voice encode their notional patient by means of suffixal person indices, reminiscent of those used for nominal predication (and not found elsewhere in the verbal paradigm), whereas their notional agent is encoded by means of prefixes of the absolutive/genitive series. That way, I surmise that the Chiquitano inverse construction (as in *\*a-ij-arapá-ta* 2SG-INV-pour-F.3SGP ‘you pour it’) originated as a nominal predication (‘it is your poured thing’). This is formally and functionally similar to the evolution of patient nominalizations in the Tupian languages, which currently employ erstwhile patient nominalizations in the object focus construction (Galucio & Nogueira 2018). I do not discard the possibility that Proto-Chiquitano *\*-ij-* is cognate with the Tupian (erstwhile) patient nominalizers: Wayoro, Akuntsu, Sakurabiat *-i-*, Tuparí *-y(’)-*, Makurap *-yĩ-* (the correspondences between these forms are not entirely regular, making it difficult to unambiguously reconstruct the Proto-Tupian form).

| gloss             | Proto-Jê<br>(Davis 1968)       | updated<br>reconstruction   | Proto-Tupian<br>or Guajajara<br>(as in Davis 1968) | updated form<br>or reconstruction   | comparison status  |
|-------------------|--------------------------------|---|--|---|--|
| ‘liver’           | * <i>ma</i>                    | PJ *- <i>mba</i> <<br>PMJ *- <i>mbâ</i>                                   | PT * <i>pia</i>                                    | PT *- <i>piʔa</i> ,<br>abs. * <i>mbiʔa</i>  | accepted   |
| ‘husband’         | * <i>mzen</i>                  | PJ *- <i>mbê₂n</i> <<br>PMJ *- <i>mbi₂n</i>                               | PT * <i>men</i>                                    | PT *- <i>mẽT</i>  | accepted   |
| ‘water’           | * <i>ŋo</i>                    | PCerr * <i>ŋgôj</i> ’,<br>PSJ * <i>ŋgôj</i> (irregular<br>correspondence) | PT * <i>igi</i>                                    | PT * <i>ʔu</i> ‘water’; *- <i>ŋgu</i><br>‘liquid’ (Tuparikém<br>branch only)  | rejected<br>(coda mismatch +<br>poor distribution)                   |
| ‘louse’           | * <i>ŋo</i>                    | PJ *- <i>ŋgâ₁n</i> <<br>PMJ *- <i>ŋgy₁n</i> °                             | PT * <i>ŋkiv</i>                                   | PT *- <i>ŋguP</i>   | rejected<br>(coda mismatch)  |
| ‘arm’ /<br>‘hand’ | * <i>pa</i> ‘arm’              | PJ *- <i>pa</i> ‘arm,<br>branch’ < PMJ *- <i>pa</i> ⁸                     | PT * <i>po</i> ‘hand’                              | PT *- <i>pə</i> ,<br>abs. * <i>mbə</i> ‘hand’   | accepted   |
| ‘foot’            | * <i>par</i>                   | PJ *- <i>par</i> <<br>PMJ *- <i>pâr</i> °                                 | PT * <i>pi</i>                                     | PT *- <i>pi</i> , abs. * <i>mbi</i>   | accepted   |
| ‘head’            | * <i>krã</i> , * <i>krãñ</i>   | PJ *- <i>krj̃j</i> ’ <<br>PMJ *- <i>krj̃ñ</i> °                           | Guaj. <i>kaŋ-</i>                                  | Guaj. - <i>àkàg</i> ‘head’ <<br>PTG *- <i>ʔa-kāk</i> ‘head’ <<br>PT *- <i>ʔa</i> ‘head’ +<br>PTG *- <i>kāk</i> ‘bone’ | rejected<br>(coda mismatch +<br>wrong morphological<br>segmentation) |
| ‘to eat’          | * <i>ku</i> , * <i>kur</i>     | PJ *- <i>ku₂</i> < PMJ *- <i>ko₂</i>                                      | Guaj. - <i>ʔu</i>                                  | Guaj. - <i>ʔu</i> <<br>PTG *- <i>ʔu</i> < PT *- <i>ko</i>   | accepted   |
| ‘path’            | * <i>pr̃i</i>                  | PJ * <i>pr̃yn</i> <<br>PMJ * <i>pr̃an</i> °                               | Guaj. <i>pε</i>                                    | Guaj. <i>pe</i> / - <i>rape</i> <<br>PTG * <i>pe</i> / *- <i>rape</i> <<br>PT * <i>pe</i> / *- <i>ja:pe</i>           | rejected<br>(coda mismatch)  |
| ‘one’             | * <i>pici</i> , * <i>picit</i> | PCerr *- <i>pʲji</i> <<br>PMJ *- <i>p(V)jet</i>                           | Guaj. <i>pitci</i>                                 | Guaj. <i>pitài</i> ~ <i>pitei</i> ~<br><i>mitài</i> ~ <i>pitáz</i> ~ <i>petei</i> <<br>* <i>pe-tẽ-C</i> ⁹             | rejected<br>(multiple issues)  |

Table 2. Davis’ (1968) Jê–Tupian etymologies and their current status

#### 4. Possible cognates

This section presents the lexical evidence for the Macro-Jê–Tupian hypothesis. Seeking to reduce the number of false positives, I adopt a stringent approach to cognate identification. In order to qualify as a likely cognate set, the reconstructed Proto-Macro-Jê and Proto-Tupian morphemes must show a full match between all consonants with respect to the place of articulation (i.e., Proto-Macro-Jê labial consonants can only correspond to Proto-Tupian labial consonants, and so on), whereas back vowels in one protolanguage are not allowed to correspond

⁸ Nikulin (2020: 369) reconstructs a palatal coda in this word (PMJ \*-*paj* ~ \*-*paj*°), based on the Akuwẽ reflexes: Xavante -*pai-hi* ‘arm’, -*pa-nõ* [-pa:n:õ] /-*paj-dõ*/ ‘arm’, Akwẽ-Xerente -*pai-nõ* ‘arm’. Note, however, that the palatal coda does not show up in Akwẽ-Xerente -*pa-krta* // -*pa-krda* ‘arm’, nor is it visible in Xavante *pa* ‘creek’, -*pa* or *wede-pa* ‘branch, root’; Akwẽ-Xerente -*pa* or *wdê-pa* ‘root’. Therefore, the grounds for reconstructing a palatal coda in PMJ are rather weak.

⁹ The reconstruction \**petẽC* is shallower than Proto-Tupi–Guaranian: the reflexes of this form are only found in a few Tupi–Guaranian languages, such as Tapiete *pente*, Mbyá *petẽ* ~ *teĩ*, Ka’apor *petẽ*. This is likely a fossilized derivative from the PTG root \**pe* (followed by \*-(e)*te* ‘true’ and by a diminutive suffix). Most Tupi–Guaranian languages reflect different derivatives of \**pe*, such as \**o-je-pe* (with a 3<sup>rd</sup> person active prefix \**o-* and the reflexive prefix \**-je-*); see Schleicher 1998: 12–13.

to front vowels in another protolanguage. Some slight deviations from this principle are duly justified. At this stage, semantically shifted cognates are not considered. In what follows, I list the Macro-Jê–Tupian lookalikes that satisfy the aforementioned criteria separated into four groups: etyma that are clearly reconstructible both to Proto-Macro-Jê and Proto-Tupian (4.1), etyma that are clearly reconstructible to Proto-Macro-Jê but have a deficient distribution in Tupian (4.2), etyma that are clearly reconstructible to Proto-Tupian but have a deficient distribution in Macro-Jê (4.3), and etyma that have a limited distribution both in Macro-Jê and Tupian (4.4). 4.5 lists some lookalikes that are best interpreted as loans or accidental resemblances.

In what follows, rather than citing reflexes in all daughter languages to support the reconstructed forms, I provide data from representative languages of each branch: typically Bésiro for Chiquitano, Djeoromitxí for Jabutian, Maxakalí for Maxakalian, Xavante for Akuwẽ, Khĩsêjtjê for Goyaz, Kaingang for Southern Jê, Makurap or Wayoró for Tuparian, Karitiana for Arikém, Paiter for Mondé, Yudja for Juruna, Mundurukú for Mundurukuan, Sateré-Mawé and Apyãwa for Mawé–Guaranian.

#### 4.1. Good distribution in both families

##### **3NCRF prefix:** PMJ \**i-* / \**c-* : PT \**i-* / \**c-*

The Proto-Macro-Jê reconstructions are from Nikulin (2020: 393, 423), who claims that \**i-* was used with class I stems, and \**c-* with class II stems. The reflex of \**i-* are found in all Macro-Jê branches, including Chiquitano (Bésiro *i-*), Western (Rikbaktsa *i-*, Arikapú *i-*, possibly also Ofayé *ã-*), Karajá (*i-*), and Eastern (Xavante *ĩ-*, Maxakalí *ũ-*). The reflexes of \**c-* are found in at least one Western language (Ofayé *h-*), in Karajá (*t-/tx-*), and in several Eastern languages (Khĩsêjtjê *s-*, Xavante *ts-*), but possibly also in Chiquitano (Bésiro *θ-*). The original distribution is still clearly preserved in Karajá and the Akuwẽ languages,<sup>10</sup> possibly also in Ofayé and Chiquitano.

The Proto-Tupian prefixes \**i-* and \**c-* are likewise used with class I and class II stems, respectively. The original distribution is most clearly seen in the Mundurukuan and Mawé–Guaranian languages of the Eastern branch (Mundurukú *i-/y-* and *t-*, Sateré-Mawé *i-* and *h-*, Apyãwa *i-* and *h-/θ-*) and in one Tuparikém language (Makurap *θ-/y-* and *t-*). In the Tuparikém branch, the prefix \**i-* is mostly preserved in all languages, with special reflexes before vowel-initial roots in Tuparian (Makurap and Wayoró *y-*, Tuparí *s-/y-*, Akuntsú *t-/ɲ-*, Sakurabiat *s-*); in Makurap, it was irregularly lost before consonants, thus yielding forms such as *θ-tur-et* ‘her/his spade’ or *θ-kar-et* ‘her/his body’ (Braga 2005: 51) instead of the expected \**i-tur-et*, \**i-kar-et*. However, in all Tuparikém languages except Makurap the prefix \**i-* was also extended to erstwhile class II stems, replacing \**c-* altogether. It is possible that the prefixes \**i-* and \**c-* are also reflected in the Mondé languages, but I am unaware of a coherent account of their evolution in that particular branch of Tupian.

The reflexes of this person index are opposed to those of PMJ \**ta-*, PT \**tə-* in some Macro-Jê branches (Karajá, Western) and in some Tupian languages (Tuparikém branch, Sateré-Mawé) in that the indexed argument has a disjoint reference with some other participant (typically the subject).

<sup>10</sup> The Akuwẽ languages have innovated by extending the prefix *ĩ-* (originally used with class I stems) to most class II stems, resulting in the allomorphs Xavante *ĩts-*, Akwẽ-Xerente *ĩs-* (instead of *ts-/s-*). The conservative allomorphs *ts-/s-* are found, for example, in the perlativ postposition (Xavante *-dzô*, Akwẽ-Xerente *-zô*; third-person form *ts-ô/s-ô*).

**‘meat, flesh’**: PMJ \**ĩt* / \**-ñĩt* : PT \**ẽt* / \**-jẽt*

The Proto-Macro-Jê reconstruction \**-ñĩt* is from Nikulin 2020: 407. The root is preserved in all first-level branches of Macro-Jê, including Chiquitano (Bésiro *n-{a}ñé-se*), Western (Djeoromitxí *-nĩ*, Rikbaktsa *-ni*), Karajá (*dèè*), and Eastern (Maxakalí *-yĩn*, Khîsêjtjê *-nhi*, Xavante *-nhi*, Kaingang *-nĩ*). The reconstruction of the coda \**-t* is based on the evidence from the Trans-São Francisco languages, where Krenak *-ñiik* preserves its manner of articulation (with the regular change from an alveolar to a velar), and Maxakalí *-yĩn* preserves its place of articulation. The correspondences are regular. As for the absolute form \**ĩt*, it is preserved only in the Maxakalí compound *ĩn-mõ-xa* ‘the Ìnmõxa monster’, analyzed in Silva 2020: 184 as ‘the flesh going out’; it must be an archaism, since \**-ñĩt* ‘meat’ has extra morphology — the ancient relationalizing prefix \*/-j-/ — compared to it.

The Proto-Tupian reconstruction is based on Proto-Tuparian \**-jẽtʔã* (Nikulin & Andrade 2020: 296) and Proto-Mundurukuan \**-ẽn* (Picanço 2019: 137), with reflexes present in all languages of the respective branches (Wayoró *-yẽra*, Mundurukú *-ẽn*, etc.); see Galucio et al. 2015: 253 for a selection. Proto-Tuparian shows a fossilized formative \**-ʔa* (originally a classifier for spherical objects, but found in other Proto-Tuparian terms as well) and the relationalizing prefix \*/-j-/, which surfaces as \**-j-* before a nasal vowel. The correspondences are otherwise regular.

In Macro-Jê, the reflexes of \**-ñĩt* belong to class II in the languages of the Goyaz branch of the Jê group, but to class I in the languages of the Akuwẽ branch of Jê (see Estevam 2011: 138 for Xavante) and in Karajá (see Ribeiro 2012a: 216 for an example). It must have originally belonged to the less productive class II. In Tupian, \**-jẽt* must have originally belonged to class II, as attested for Makurap by Braga (2005: 208; note that she uses the label “class I” for my class II); other Tuparian languages have lost the distinction. Mundurukuan has apparently reanalyzed the erstwhile absolute stem \**ẽt* ‘meat/flesh (unpossessed)’ as relational.

**‘to stand’**: PMJ \**ja* (nonfinite \**-ja-m*) : PT \**-ja* or \**-ʔãP*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 398. The etymon is preserved in the Western (Ofayé *-xe:ʔ ~ -he:ʔ*, possibly Rikbaktsa *-sa* ‘to start walking /a developmental milestone/’), Karajá (nonfinite *-lma*), and Eastern (singular only: Maxakalí *-xip*, irrealis *xihip*; Khîsêjtjê *ta*, nonfinite *-tãm*; Xavante *dza*, nonfinite *-dzam*; Kaingang *jẽ*, nonfinite *jẽg*) branches. The correspondences are regular. The original finite form was lost in Maxakalí, whose irrealis form has been remodeled after the realis form (*-xip* < \**-ja-m*; the expected irrealis form would be \**xihi* < \**ja*), and in Karajá, which now shows a suppletive finite form *-yĩ* of unknown origin.

The Proto-Tupian reconstructions correspond to two distinct etyma, which could be compared to PMJ finite and nonfinite forms, respectively.

Proto-Tupian \**-ja* is preserved in two Rondonian branches, Tuparikém (as an auxiliary only: Karitiana *ty-so* ‘IMPF:standing’, Sakurabiat *ta-t* ‘PRG.PRS:standing’, *ta-a* ‘PRG.PST:standing’) and Rama-Puru (Karo *-ja* ‘to stand’, with a possible cognate in Puruborá; Galucio et al. 2015: 258). The correspondences are regular. Note that Proto-Tuparikém can be reconstructed as having a series of no less than three auxiliaries contrasting for position only, as shown in Table 3. These correspond to lexical verbs for ‘to sit’ and ‘to stand’ in Rama-Puru or other Tupian languages; the term for ‘to lie’ is noncognate in Rama-Puru (\**-mbop* > Karo *-mbop*, Puruborá *-bop-a*), but clear cognates are found elsewhere in Tupian, as in Old Tupí *tub-/rub-* ‘to lie.NF’ (Barbosa 1956: 305).

Proto-Tupian \**-ʔãP* is preserved in at least one Tuparikém language (Akuntsu *-ãP*) and in most Mawé–Guaranian languages (Eastern branch), such as Sateré-Mawé *-’am* ‘to go up’, Old Tupí *-am* ‘to stand’, Kamayurá *-’am* ‘to stand’, among many other cognates. The verb for ‘to stand up’ in Mawé–Guaranian languages is evidently derived from this root: Sateré-Mawé *-poi’am* ‘to stand up’, Old Tupí *-puam* ‘to stand’, Kamayurá *-uhwam* ‘to stand’, among others. The correspondences are regular.

|          | Proto-Tuparikém | Sakurabiat       | Karitiana       | Proto-Rama-Puru | Karo  |
|----------|-----------------|------------------|-----------------|-----------------|---|
|          | AUX             | PRG (PRS / PST)  | IMPF            | lexical verb    | lexical verb                                    |
|          |                 | Galucio 2001: 58 | Rocha 2022: 239 |                 | Gabas Jr 1989: 16, Galucio et al. 2015: 257–258 |
| lying    | *joP            | to(o)p-Ø / to-a  | ty-syp          | (*-mbop)        | (-mbop)   |
| sitting  | *jē             | yē-t / y-ã       | ty-ja           | *-jō            | -yã   |
| standing | *ja             | ta-t / ta-a      | ty-so           | *-ja            | -ya   |

Table 3. Tuparikém auxiliaries and Rama-Puru lexical verbs

In Macro-Jê, the finite stem is reconstructed as absolute (uninflectable), and its nonfinite counterpart is a class II relational stem. This is clearly seen in the Khîsêtjê reflex: the finite stem *ta* is absolute, and the nonfinite stem *-tām* takes the full set of the person prefixes (1 *i-tām*, 2 *a-tām*, 3 *s-ām*), where *-t-* is a thematic consonant. In Tupian, the morphosyntactic behavior of *\*-ja* and *\*-ʔāP* cannot be reconstructed with certainty. The former is reflected as an auxiliary in the Tuparikém languages, where it combines with other morphemes (such as *-t* ‘present’ and *-a* ‘past’ in Sakurabiat; *ty-* ‘imperfective’ in Karitiana), whereas the Karo and Puruborá reflexes are only marginally attested in the available data. The latter is mostly known from Mawé–Guaranian languages, where the reflexes are active class I intransitive verbs. Therefore, there is a class membership mismatch between the PMJ class II stem *\*-ja-m* and the PT class I stem *\*-ʔāP*.

**‘name’:** PMJ *\*-jet* : PT *\*-jeT*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 405. The root is preserved in Karajá (*nii*) and in many languages of the Eastern branch (Maxakalí *-xu-xet-’ax/-ã-xet-’ax*, Khîsêtjê *-nhinti*, Xavante *-nhitsi*, Kaingang *-jiji*). In all these languages, the root is preceded by a prefix whose PMJ shape is difficult to reconstruct: Karajá and the Cerrado languages point to PMJ *\*-ñi-jet*, the Southern Jê languages to *\*-ji-jet* or maybe *\*-jy-jet*, and Maxakalí shows an alternating pair of prefixes, whose choice depends on the syntactic context. The coda *\*-t* is reconstructed based on the correspondence between Maxakalí */-t/* and Jê zero. The semantic equivalents in Chiquitano (*\*-tsiri / \*-iri*), Ofayé (*-xirê?*), and Krenak (*-unjak*) show some superficial resemblance to PMJ *\*-jet*, but are unlikely to be cognate due to lack of regular sound correspondences.

Proto-Tupian *\*-jeT* is preserved in most Tupian languages, including the Tuparikém (Makurap *-xet*, Karitiana *-sat*), Mondé (Paite *-léd*), and Eastern (Sateré-Mawé *-set*, Apyãwa *ter-a / -rer-a*) branches; see Galucio et al. 2015: 261 for a selection of reflexes. The PT reconstruction is based on the intermediate reconstructions, such as Proto-Mawé–Guaranian *\*-t’et*, or—in my notation—*\*-ceT* (Meira & Drude 2015: 294) and Proto-Tupian *\*-jeT* (Nikulin & Andrade 2020: 295). The correspondences are regular, with the possible exception of probable exceptions in the Juruna languages, such as Yudja *-zá* (the regular reflex of PT *\*-T* is Yudja *l /ɓ/*, not *z*).

In Tupian, the root is reconstructed as a class II relational stem. In Macro-Jê, it is always accompanied with derivational prefixes, and the inflectional properties of the bare root are thus not recoverable.

**‘father’:** PMJ *\*-jo<sub>2</sub>m* : PT *\*-joP*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 401. The root is preserved in most branches, including Western (Djeoromitxí *ho{txi}* ‘father’, *-ro* ‘father; male’, Rikbaktsa *-zo*, Ofayé *-xəw ~ -xôw ~ -xew*), Karajá (3<sup>rd</sup> person *t-by*), and Eastern (Canela–Krahô *-xūm* ‘male’,

3<sup>rd</sup> person *h-ũm* ‘father’, Kaingang *-jóg*). The Correspondences are regular, except that the Ofayé reflex shows irregular vowels in Eduardo Ribeiro’s (*ə*) and Sarah G. Gudschinsky’s (*e*) attestations. In addition, no traces of the coda *\*-m* are seen in the alleged cognates in the Northern Jê languages of the Trans-Tocantins subgroup: Apinajé *-xũr*, Mëbêngôkre *djũnũ* or *djũn-wa* ‘father (voc.)’, Kajkwakhrattxi and Khîsêtjê *turê* ‘father (voc.)’, though the Mëbêngôkre term for male — 3<sup>rd</sup> person *θ-ũm-ti-re* — does show the expected *-m*. In Karajá and in the Northern Jê languages of the Timbira branch, only the third-person form (PMJ *\*c-o<sub>2</sub>m*) is used in the meaning ‘father’, though the Timbira languages preserve the uninflected form *\*-jo<sub>2</sub>m* in the meaning ‘male’.

Proto-Tupian *\*-joP* ‘father’ is preserved in most Tupian languages, including Kepkiriwat (⟨*xuá*⟩), Tuparikém (Wayoró *-ndop*, Karitiana *-syp* ‘father of a woman’), Mondé (Paiter *-lob*), and Eastern (Yudja *-pá*, Kuruaya *-lop*, Awetí *tup/-up*, Apyãwa *tow-a/-row-a*). The reconstructed form is based on the intermediate reconstructions, such as Proto-Mawé–Guaranian *\*-t<sup>j</sup>up*, or — in my notation — *\*-cuP* (Meira & Drude 2015: 293) and Proto-Tuparian *\*-joP* (Nikulin & Andrade 2020: 295). The correspondences are regular; Alves (2004: 180) documents Tuparí *-hòp*, with an unexpected long vowel (symbolized by means of a grave accent in the practical orthography), but the expected form with a short vowel is attested in Singerman 2018: 50. There is also a homonymous stem PT *\*-joP* ‘fish roe, pus’, whose reflexes have at times been claimed to belong to the same etymology as *\*-joP* ‘father’ (cf. Meira & Drude 2015: 293); its reflexes are found in the Rama-Puru (Karo *-xop* ‘dirt’, Puruborá *-tɔP* ‘fish roe’), Mondé (Paiter *-lób* ‘pus’), and Eastern (Sateré-Mawé *-sup* ‘sperm’, *win sup* ‘fly maggots’, Apyãwa *ipira-ow-a* ‘fish roe’) branches.

In both language families, the root is reconstructed as a class II stem, with the following provisos. In Macro-Jê, it appears to have shifted to class I in Ofayé (3<sup>rd</sup> person *ã-xəw* ~ *ã-xôw* instead of the expected *\*h-xəw* ~ *\*h-ôw*; Oliveira 2006: 97; Ribeiro n/d). In Tupi–Guaranian, *\*tuP*/*\*-ruP* belongs to the so-called subclass IIb, which includes a handful of kinship terms; it differs from other class II subtypes in having a third-person form identical to the absolute one (*\*tuP* ‘her/his father’). It thus contrasts with nouns such as *\*-ruP* ‘fish roe’, whose third-person form is reconstructed in my proposal as *\*θ-uP* (> Apyãwa *h-ow-a*).

**‘pus’**: PMJ *\*-jo<sub>2</sub>w<sup>o</sup>* : PT *\*-joP* ‘fish roe, pus’

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 401. The root is preserved in the Western (Djeoromitxí *-ro* ~ *-ro{o}* ‘sap, pus, mucus’) and Eastern (Maxakalí *-xapa*, Khîsêtjê *-tu{ru}*, Xavante *-dzub{rui}* // *-dzub{ru}*, Kaingang *{f}jo* ‘pus’, *{f}jó-m* ‘to suppurate’) branches. The correspondences are mostly regular, except that the Akuwê reflexes show an unexpected palatal coda in the utterance-medial allomorph *\*-jubruj*. In addition, the PMJ coda *\*-w<sup>o</sup>* is reconstructed exclusively in order to account for the correspondence PJ *\*-P* ~ Maxakalí *-pV*. If the Maxakalí datum turns out to be noncognate, the PMJ reconstruction can be updated to *\*-jo<sub>2</sub>p*. The Cerrado languages reflect a derived form, *\*-jup-r*, which can be interpreted as an erstwhile nonfinite form of the verb ‘to suppurate’. The Southern Jê languages reanalyzed the third-person index *\*c-* > *\*θ-* as a part of the stem (Ribeiro 2004a: 95).

Proto-Tupian *\*-joP* ‘fish roe, pus’ is reflected in the Rama-Puru (Karo *-xop* ‘dirt’, Puruborá *-tɔP* ‘fish roe’), Mondé (Paiter *-lób* ‘pus’), and Eastern (Sateré-Mawé *-sup* ‘sperm’, *win sup* ‘fly maggots’, Apyãwa *ipira-ow-a* ‘fish roe’) branches. The correspondences are regular. At least in Paiter, it contrasts with the nearly homonymous term for ‘father’ in having high tone (Bontkes 1978: 5), suggesting that the tonal contrast was also present in PT. In this study, I do not make an attempt at reconstructing PT tone.

In both language families, the root is reconstructed as a class II stem, as evidenced by the third-person forms such as Khîsêtjê *s-u{ru}* and Apyãwa *h-ow-a*.



**‘tooth’**: PMJ \**-juũ*<sup>o</sup> : PT \**-jãC*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 403. The etymon is preserved in all Macro-Jê branches, including Chiquitano (e.g. Bésiro *-só’o*), Western (Djeoromitxí *hüi*, Ofayé *-xe?*), Karajá (*juu*), and Eastern (Khîsêjtjê *-twa*, Xavante *-’wa*, Kaingang *-jã*, Maxakalí *-xox*). The reflexes in all daughter languages are regular, except that Xavante *-’wa* and Akwê-Xerente *-kwa* unexpectedly lack utterance-medial allomorphs with a palatal coda (Xavante \**-’wai*, Akwê-Xerente \**-kwai*), or at least such allomorphs have not been attested in the literature. Possibly the utterance-final allomorph, which regularly loses the underlying palatal coda, has been generalized in the history of the Akuwê languages (see Nikulin 2017: 155–158 on utterance-medial and utterance-final allomorphs in Akuwê). The palatal nasal coda followed by an echo vowel is reconstructed based on the reflexes in the Maxakalian languages (Maxakalí *-xox* and Pataxó-Hãhãhãe *<-tei>*, *<-tóy>*, *<-t<sup>h</sup>ui>*, *<-txüi>* point to a palatal coda), as well as in Krenak (*-jun*, with *n* clearly going back to PMJ \**-ñ* or \**-ñ<sup>o</sup>*), Pykobjê–Krîkatí (*-xwaa*, with the long vowel suggesting an erstwhile \**-n<sup>o</sup>* or \**-ñ<sup>o</sup>*), and Ofayé (*-xe?*, with the plural and diminutive forms attested in Oliveira 2006: 79 strongly suggesting the presence of an underlying nasal coda).

Proto-Tupian \**-jãC* is preserved in all branches of Tupian, including Kepkiriwat (*<-nhain>*, *<-nhai->*), Tuparikém (Makurap *-yãiy*, Karitiana *-jõj*), Rama-Puru (Karo *-yãiy*), Mondé (Zoró *-jêêj*), and Eastern (Munduruku *-nũy*, Apyãwa *tÿj-a/-rÿj-a*); see Galucio et al. 2015: 254 for a selection of reflexes. The correspondences are completely regular, except that those Mondé languages that preserve this etymon — Aruá, Gavião, and Zoró — unexpectedly show a long front vowel /ẽ:/ as the reflex of PT \**-ã*.

In both families the stem is reconstructed as relational, class II. This is clearly seen in the third-person (singular) forms, with no thematic consonant: Bésiro *Ø-o’ó-xi*, Karajá *tx-uu*, Khîsêjtjê *s-wa* < PMJ \**-c-uũ<sup>o</sup>*; Makurap *t-ãiy*, Munduruku *t-ũy*, Apyãwa *h-ÿj-a* < PT \**-c-ãC* (Ribeiro 2012a: 119; Santos 1997: 39; Braga 2005: 50; Picanço 2005: 262; Almeida et al. 1983: 26–27).

**‘to ingest’ = ‘to eat/drink’**: PMJ \**-ko<sub>2</sub>* : PT \**-ko*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 410. Reflexes are found in most Macro-Jê branches: Western (Djeoromitxí *-ko* ‘to eat’, Ofayé *-hô* ‘to eat something solid’), Karajá (*-ky* ‘to eat grains’), and Eastern (Khîsêjtjê *-khu* ‘to eat.PL’, Xavante *-hu* ‘to ingest.PL’, Kaingang *-ko* ‘to eat, to use’). The correspondences are regular. Rikbaktsa *-ku* ‘to drink’ is viewed as a reflex of PMJ \**-ko<sub>2</sub>* in Nikulin 2020, but it could be alternatively considered cognate with Proto-Goyaz \**ij-kô* (nonfinite \**-kô-m*) ‘to drink’ (> Khîsêjtjê *i-khõ*, *-khõm*).

The Proto-Tupian reconstruction is mentioned *in passim* in Nikulin & Carvalho 2022: 16; see Galucio et al. 2015: 256 for a selection of reflexes. The root is preserved in most branches of Tupian, including Kepkiriwat (*<-qu->*), Tuparikém (Tuparí *-ko*, Karitiana *-’y*), Rama-Puru (Karo *-’o*, Puruborá *-’o*), and Eastern (Munduruku *-’o*, Apyãwa *-’o*). The correspondences are regular.

In both language families, the root is a class I stem. In Old Tupí and possibly some other TG languages, this verb is unusual in that it does not take the third-person accusative prefix *îo-* when finite (Barbosa 1956: 305). In the languages of the Cerrado branch of the Jê group, the verb \**-ku* takes indices of the accusative series when finite (just like all monosyllabic canonical transitives), whereas its nonfinite form is \**-ku-r’*. Note that in almost all Tupian languages the reflexes cover the entire semantic domain of eating and drinking; in Macro-Jê, this is synchronically the case in the Akuwê languages (compare Xavante *-hu* ‘to eat.PL’ and *õ-hu* ‘to drink.PL’).

**‘tree, tree-like object (leg, horn, bone)’**: PMJ \**(-)ky<sub>1</sub>m<sup>o</sup>* : PT \**(-)k<sub>1</sub>uP*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 409. The root is preserved in all first-level branches of Macro-Jê, including Chiquitano (Bésiro *-{tápa}ki* ‘horn’), Western (Djeoro-

mitxí *ku* ‘tree’, {*me*}*ku* ‘horn’, Ofayé *həy* ‘tree’, *-həy* ‘horn’), Karajá (female speech *kòd*, male speech *òd* ‘wood, horn’), and Eastern (Maxakalí *-kup* ‘stick, bone, leg’, *-ptox-kup* ‘horn’, Khĩsètjê *khô* ‘club’, *-khô* ‘grove’, Xavante *-ômo* // *-u* ‘horn’, Kaingang *ka* ‘tree’, *-{nĩ}ka* ‘horn’). The reconstruction of a labial nasal coda followed by an echo vowel is based on the evidence from the Akuwẽ languages (Proto-Akuwẽ *\*-kômô* // *\*-ku* ‘horn’) and corroborated by Maxakalian, which preserves its place of articulation. The correspondences are regular.

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 31; see Galucio et al. 2015: 252 for a selection of reflexes. The root is preserved in all first-level branches of Tupian, including Kepkiriwat (<quêp>/<queb> ‘tree, wood’, <nécubá> ‘elbow’, <cü-ümarã> ‘leg garter’, <ócüpe> ‘stud’, <mbátoquêp> ‘index finger’, <jaácupe> ‘maize cob’), Tuparikém (Makurap *kup* ‘tree’, *-api-kup* ‘horn’, Karitiana *’ep* ‘tree’, *-’ep* ‘bone’), Rama-Puru (Karo *ma-’úp* ‘tree’, Puruborá *?iP* ‘tree’), Mondé (Paiter *ihb* ‘tree’), and Eastern (Yudja *epá* ‘stick’, Mundurukú *íp* ‘tree’, *-’ip* ‘tree/wood (classifier)’, Sateré-Mawé *aria-’yp* ‘tree’, *-’yp* ‘tree (of a concrete species), handle’, Apyãwa *-’yw-a* ‘leg, handle, tree (of a concrete species)’); see Galucio et al. 2015: 252 for a selection of reflexes. The correspondences are regular.

In both language families, the root occurs both as a class I relational stem and as an absolute stem. It is thus reconstructed as relationally labile (i.e., the possessor is optional). Note the closely matching semantics of the reflexes in Macro-Jê and Tupian: ‘tree’ is the most recurring meaning, but ‘leg’, ‘horn’, and ‘bone’ are also attested across both families.

**‘liver’:** PMJ *\*-mbâ* : PT *\*-pi(-)ʔa* / *\*mbi(-)ʔa*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 375. The root is preserved at least in the Western (Djeoromitxí *-bä*, Rikbaktsa *-py*, Ofayé *-fa(h)*), Karajá (*maa*), and Eastern (Khĩsètjê *-mba*, Xavante *-pa*, Kaingang *-tj-mê*) branches. The correspondences are regular.

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 32, where the element *\*-ʔa* is given as a part of the root. It is semantically and morphologically plausible to analyze *\*-ʔa* as a formative for spherical objects. The root is preserved in most first-level branches, including Kepkiriwat (<piá>), Tuparikém (Makurap *-pia*), Rama-Puru (Karo *-pía*, Puruborá *-bia*), and Eastern (Yudja *-bi’á*, Mundurukú *-psà*, Sateré-Mawé *-py’a/my’a*, Apyãwa *-py’ã-0/my’ã-0*); see Galucio et al. 2015: 256 for a selection of reflexes. The correspondences are regular.

In Macro-Jê, the reflexes of *\*-mbâ* belong to class I. In Tupian, *\*-piʔa* is reconstructed as a relational class I stem, and *\*mbiʔa* as an absolute one; this combination is also known as class Ib in Tupi–Guaranian studies. The erstwhile absolute stem *\*mbiʔa* is preserved in the Mawé–Guaranian languages but was apparently lost in all other branches.

**‘smoke’:** PMJ *\*-ñĩjãk* : PT *\*-jĩ.K*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 400. The etymon is preserved in the Paraná and Akuwẽ branches of the Jê groups (Xavante *-nhidzé*, Akwẽ-Xerente *-nĩze*, Laklãnõ *nějó*, Kaingang *nĩja*) as well as in Karajá *he-dà* (from *hèè* ‘firewood’). The reconstruction of a stem-final velar stop is supported by the Kaingang derivative *nĩjãg* ‘to produce smoke’. Karajá *hedà* (Palha 1942: 25; Ribeiro 2012a: 105) is not the main term for ‘smoke’ in the modern language, where *wàdasi* ‘smoke’ is found instead.

Proto-Tupian *\*-jĩ.K* is preserved in most Tupian languages, including Kepkiriwat (<iá-in>), Tuparikém (Wayoró *-yĩng*, Karitiana *-jing*), Mondé (Paiter *mokây-nĩg*), and Eastern (Mundurukú *-diğ*, Sateré-Mawé *y-hiğ*, Awetí *taza-ting*, Kawaiwete *tata-sing*); see Galucio et al. 2015: 259 for a selection of reflexes. The reconstructed form is based on the intermediate reconstructions, such as Proto-Mawé–Guaranian *\*-t’ij*, or—in my notation—*\*-cĩK* (Meira & Drude 2015: 294), Proto-Mundurukuan *\*-đinj* (Picanço 2019: 139), Proto-Tuparian *\*-jiĩ.K* (Nikulin & Andrade 2020:

296), with the reconstruction of a long vowel based on evidence from Tuparian languages, such as Wayoró and Sakurabiat. The correspondences between these forms are mostly regular. The denasalization of \*ĩ in Mundurukuan could be regular, as the sequence \*\*ðĩ was banned in Proto-Mundurukuan (Picanço 2005: 173). So could be the second stage of the purported development \*jĩ > \*cĩ > \*çĩ in Proto-Mawé–Guaranian.<sup>11</sup> Somewhat problematic are the alleged reflexes in the Juruna languages (such as Yudja *-xi'ã* < Proto-Juruna \**-fi-ʔã*). Nikulin and Andrade (2020: 296, fn. 30) discuss several difficulties with the reflexes in individual Tuparian languages. Finally, an irregular reflex of PTG \*-k is seen at least in the Apyãwa form *tata-xin-a* /tãtã-tçĩt-a/.

Both in Macro-Jê and Tupian, the stem is reconstructed as relational, class II. This is clearly seen in the third-person forms, with no thematic consonant: Xavante *[i]ts-idzé* < PJ \**c-ĩjã*<sub>2</sub><sup>k</sup> (Lachnitt 1987: 79); Mundurukú *t-iğ* < PT \**c-ĩ:k* (Picanço 2005: 320).

This comparison deviates from my stringent criteria in that the PMJ sequence \*jã is not matched to any PT segment. However, the fact that the PT form is reconstructed with a long vowel makes the comparison somewhat more plausible: it is easy to imagine a contraction of an \*ijV sequence into \*ĩ.

**‘feces’: PMJ \**-ñVt*<sup>o</sup> : PT \**-jVt***

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 407. The etymon is preserved in Chiquitano (Bésiro *-a'a*) and in the Eastern branch (Maxakalí *-yõn* ‘feces, to defecate’, Khîsêtjê *-nhin* // *-nhini*, Xavante *-nhana*). Note that in multiple Macro-Jê languages the reflexes of \**-ñVt*<sup>o</sup> are polysemous and can refer not only to feces, but also to bowels (Mêbêngôkre, Parkatêjê, Pykobjê–Křikatí, Canela–Krahô, Xavante); in other languages, terms for ‘bowel’ or ‘small bowel’ are derived from the respective root (Bésiro *-an-terere*, Mêbêngôkre *-nhîn kra*, Parkatêjê *-jñn-kra*, Pykobjê–Křikatí *-jêhn cra*, Canela–Krahô *-jñn kra*, Akwê-Xerente *-nnã hi-rê*). The correspondences involving the consonants are regular, except that the Xavante and Akwê-Xerente reflexes unexpectedly lack utterance-medial allomorphs with a voiceless stop coda (Xavante \**-nhatã*, Akwê-Xerente \**-ntã*), possibly due to analogy with the regular utterance-final allomorphs *-nhana* [-jñã:nã] / *-nnã* [-n<sup>h</sup>nã]. By contrast, the vowels across Macro-Jê show no regular correspondence whatsoever. Maxakalí *õ /ũ/* points to PMJ \**ũ*; Khîsêtjê *i /ĩ/* < Proto-Goyaz \**ĩ* suggests PMJ \**ĩ*; Xavante *a /ã/* < Proto-Akwê \**ã* can go back to PMJ \**ã* or \**ỹ*. The Chiquitano reflex shows a nasal vowel /ã/ in the Migueleño and Eastern varieties; the Macro-Jê origins of Proto-Chiquitano \**ã* have not been established yet, but it could technically be the regular reflex of PMJ \**ã* or \**ỹ*.

The Proto-Tupian reconstruction is based on reflexes in the Tuparikém and Eastern branches, including Wayoró *-yên* (< Proto-Tuparian \**-jêT*; Nikulin & Andrade 2020: 296),<sup>12</sup> Karitiana *-jîn* (< Proto-Arikém \**-jĩT*), Mundurukú *-nũn* (< Proto-Mundurukuan \**-ðãn*; Picanço 2019: 139), Xipaya *-súna*, Yudja *unã*, and Sateré-Mawé *-jun*. Just like in Macro-Jê, many of its reflexes either colexify ‘feces’ with ‘bowel’ (e.g. Wayoró *-yên*) or use derivatives of the root in question in the meaning ‘bowel’ (Karitiana *-jîn-py*, Mundurukú *-nũn-pú* < PT \**-jVt-pə*). In addi-

<sup>11</sup> PT \**j* normally yields Proto-Mawé–Guaranian \**c* (> Sateré-Mawé *s*, Awetí and PTG zero word-medially); the reflex \**ç* (> Sateré-Mawé *h*, Awetí *t*, PTG \**t* word-medially) is otherwise known to occur following an \**i* or a \**c* by progressive palatalization. But the sequence \**ci* is not reconstructed for any Proto-Mawé–Guaranian morpheme (at least in Meira & Drude 2015), and may have been subject to regressive palatalization in pre-Proto-Mawé–Guaranian.

<sup>12</sup> Tuparian has a similar root \**-jê:T* (also \**ki-jê:T*) ‘ashes’, which, however, must be unrelated to \**-jêT* ‘feces’ (*pace* Nikulin & Andrade 2020: 296), since its reflexes are documented with a long vowel in most daughter languages (Galucio et al. 2015: 259).

tion, PTG *\*tũT/\*-rũT* ‘black’ (*\*-ũT* in compounds, as in *\*-piθ-ũT* ‘black skin’) regularly corresponds to Sateré-Mawé *-jun* ‘feces’. It is reflected, for example, as Apyãwa *-ron*, *-pi-on*; Siriono *-rö* ‘muddy’, *-i-sö* ‘dark’, etc. Despite the semantic discrepancy, the evolution ‘feces’ > ‘dirty’ > ‘black’ seems feasible. The correspondences involving the consonants are regular. However, the vowels correspond in a unique way in this cognate set: Juruna and Mawé–Guaranian point to Proto-Tupian *\*-jōT*, Mundurukuan to *\*-jãT*, Tuparian to *\*-jēT*, and Arikém to *\*-jĩT*.

In both families the stem is reconstructed as relational, class II. This is clearly seen in the third-person forms, with no thematic consonant: Khĩsêtjê *s-ĩn* // *s-ĩni*, Xavante *ts-ãna* < PMJ *\*c-ŨT*; Mundurukú *t-ũn* < PT *\*c-ŨT* (Nonato et al. 2012: 7; Lachnitt 1987: 74; Picanço 2005: 151). The fact that the vowel correspondences are highly irregular both in Macro-Jê and Tupian can be possibly accounted for by reconstructing a low-frequency nasal vowel for both protolanguages. The colexification of the meanings ‘feces’ and ‘bowel’, found in both language families, renders the cognation hypothesis particularly plausible.

**‘earth’:** PMJ *\*ŋgyN°* : PT *\*kuc*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 418. The etymon is preserved in the Chiquitano (Bésiro *kí-xi*), Western (Ofayé *həyẽ?*) and Eastern branches (Kaingang *ga*). The Ofayé reflex, not listed in Nikulin 2020, points to a nasal coda, as suggested by the plural form *hə:-ñe* and the allomorph *hət-*, found in compounds (Oliveira 2006: 79). PNJ *\*ŋgy°* ‘clay, mud’ could be related, but the origin of the diphthong *\*y°* is unclear; the regular reflex of PMJ *\*yN* would be PNJ *\*â:* or *\*ô:*. Karajá *sùù* (underlying /θu/) does not appear to be cognate with the aforementioned forms, since PMJ *\*y* is normally reflected as *ð* /*ɔ*/ in Karajá; the reflex of PMJ *\*ŋg* in Karajá is presently unknown (but *\*ŋgr* is indeed regularly reflected as *s* /*θ*/ in Karajá).

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 31. Its reflexes are found in the Kepkiriwat (<cuitá-á>, <queitaá>), Tuparikém (Makurap *kux*, Wayoró *kuy*, Karitiana *’ej*), Rama-Puru (Puruborá *?iC*), and Eastern branches (Yudja *etá* ‘sand, beach’, Sateré-Mawé *’yi*, Apyãwa *yj-a*).

In both Macro-Jê and Tupian, the reflexes of PMJ *\*ŋgyN°* and PT *\*kuc* are typically absolute (unpossessable) nouns, though in some languages they are optionally possessed and behave as class I relational stems, as in Bésiro *n-í-kĩ ma-monkó-ka* ‘Chiquitanía’ (literally ‘the land of the Monkóxi nation’).

**‘arm’:** PMJ *\*-pa* ‘arm, branch’ :

PT *\*-pə* / *\*mbə* ‘hand, vine-like’, *\*-pə-ʔa* / *\*mbə-ʔa* ‘arm’

The Proto-Macro-Jê reconstruction adopted here differs slightly from the one in Nikulin 2020: 369, where it is reconstructed as *\*-paj* ~ *\*-paj°*. Reflexes are found in most branches, including Chiquitano (Bésiro *-pa* ‘arm, wing’), Western (Djeoromitxí *{ha}pa* ‘arm’, *ku-{ra}pa* ‘branch’, Rikbaktsa *-pa-* ‘arm (in compounds)’, *-{tsi}pa* ‘arm’, *-sara-pa* ‘branch’, Ofayé *-φe* ‘arm, wing’), and Eastern (Khĩsêtjê *-hwa* ‘arm, branch’, Xavante *-pai-hi* ‘arm’, *-pa-nõ* [-pa:n:õ] /-paj-dõ/ ‘arm’, Kaingang *-pẽ* ‘arm’, *ka pẽ* ‘branch’). The only reason for reconstructing a palatal coda are the reflexes in the Akuwẽ languages: Xavante *-pai-hi* ‘arm’, *-pa-nõ* [-pa:n:õ] /-paj-dõ/ ‘arm’, Akwẽ-Xerente *-pai-nõ* ‘arm’. However, no palatal coda is found in Xavante *pa* ‘creek’, *-pa* or *wede-pa* ‘branch, root’; Akwẽ-Xerente *-pa-krta* // *-pa-krda* ‘arm’, *-pa* or *wdê-pa* ‘root’. Therefore, the grounds for reconstructing a palatal coda in PMJ are rather weak. The meanings ‘arm’ and ‘branch’ were probably colexified in PMJ *\*-pa*, as shown by evidence from Jabutian, Rikbaktsa, and Jê. The meaning ‘wing’, seen in Chiquitano and Ofayé, is probably innovative, since a distinct root for ‘wing, armpit’ is otherwise reconstructed (PMJ *\*-jar°*; Nikulin 2020: 399).

The Proto-Tupian reconstruction  $*-pə / *mbə$  is from Nikulin & Carvalho 2022: 31. The correspondences are regular. Reflexes are found in all branches of Tupian, including Kepkiriwat ( $-mbo$  ‘CL:long’:  $\langle umbó \rangle$  ‘my guts’,  $\langle uhembó \rangle$  ‘my neck’,  $\langle boi uarumbó \rangle$  ‘anaconda’), Tuparikém (Wayoró  $mbo / -wo$  ‘hand’, Karitiana  $-py$  ‘hand’), Rama-Puru (Karo  $=pũ$  ‘CL:cylindrical+small’, Puruborá  $-bə$  ‘CL:vine-like’), Mondé (Paiteir  $-pá-be$ ), and Eastern (Yudja  $-wá$  ‘hand’ < Proto-Juruna  $*-bu-á$ , Mundurukú  $-pu$  ‘hand, finger; CL:vine-like’, Sateré-Mawé  $-po/mo$  ‘hand’,  $-po-’yp / mo-’yp$  ‘arm’, Apyãwa  $-pa-∅ / ma-∅$  ‘hand’). The reflexes in languages such as Kepkiriwat, Karo, Puruborá, and Mundurukú clearly show that PT  $*-pə$  occurred not only as a body part term, but also as a second element in compounds designating long, vine-like objects, such as vines (PT  $*utu-pə$ ), roots (PT  $*-ja-pə$ , Eastern branch only), and possibly threads, snakes, cords, fingers, etc. The term for ‘arm’ is reconstructed as  $*-pə-ʔa / *mbə-ʔa$ , whose second element appears to be  $*-ʔa$  ‘head, CL:spherical’; it has known reflexes in Rama-Puru (Karo  $-pá-be$  ‘hand’, Puruborá  $-ba$  ‘arm’) and Eastern branches (Mundurukú  $-pà$  ‘arm; CL:cylindrical+thick’ < Proto-Mundurukuan  $*-pə$ ; Picanço 2019: 136). The reflexes in Mundurukuan clearly point to PT  $*Vʔa$ , and the quality of the vowel that precedes the glottal stop is inferred based on the possible morphological relation to  $*-pə/*mbə$ . Unlike in Macro-Jê, Tupian shows a distinct root for ‘branch’, PT  $*-jãŋã$  (Wayoró  $kuw-angã$  ‘branch’,  $mbo-angã$  ‘wrist’; Karitiana  $-jõngõ \sim -jõngo$  ‘arm, branch’; Mundurukú  $-dákú \sim -nákú$  ‘branch’; Apyãwa  $-rakỹ-∅$  ‘branch’). There is also an alternate candidate for the main term for ‘arm’, PT  $*-nē$ , with reflexes in Kepkiriwat, Tuparian, Arikém, Mondé, and Mundurukuan (compare also PT  $*-nē-pi$  ‘armpit’).

In Macro-Jê, the reflexes of  $*-pa$  belong to class I. In Tupian,  $*-pə$  is reconstructed as a relational class I stem, and  $*mbə$  as an absolute one; this combination is also known as class Ib in Tupi–Guaranian studies. In Kepkiriwat and Tuparian, the erstwhile absolute stem  $*mbə > *mbo$  ‘hand (unpossessed)’ was apparently reanalyzed as relational, whereas Karo and possibly some other languages have lost the allomorph  $*mbə$  entirely.

**‘foot’:** PMJ  $*-pâr^\circ$  : PT  $*-pi / *mbi$

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 370. Reflexes are found in the Western (Rikbaktsa  $-pyry$ , Ofayé  $-φar$ ) and Eastern (Maxakalí  $-pata$ , Khîsêtjê  $-hwaj // -hwaji$ , Xavante  $-para$ , Kaingang  $-pên$ ) branches. The correspondences are regular.

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 32. The root is preserved in all first-level branches, including Kepkiriwat ( $\langle -mbi \rangle$  ‘leg’,  $\langle -mbitecaiã \rangle$  ‘heel’), Tuparikém (Makurap  $-mi$ , Karitiana  $-pi$ ), Rama-Puru (Karo  $-pi-be$ , Puruborá  $-fi-be$ ), Mondé (Paiteir  $-pi-pe$ ), and Eastern (Mundurukú  $-i$ , Sateré-Mawé  $-py/my$ , Apyãwa  $-py-∅/my-∅$ ); see Galucio et al. 2015: 255 for a selection of reflexes. The correspondences are regular.

In Macro-Jê, the reflexes of  $*-pâr^\circ$  belong to class I. In Tupian,  $*-pi$  is reconstructed as a relational class I stem, and  $*mbi$  as an absolute one; this combination is also known as class Ib in Tupi–Guaranian studies. In Kepkiriwat and Tuparian, the erstwhile absolute stem  $*mbi > *mbi$  ‘foot (unpossessed)’ was apparently reanalyzed as relational, whereas Arikém, Mundurukuan, and some other languages have lost the form  $*mbi$  entirely.

This comparison deviates from my stringent criteria in that a PMJ coda is not matched to any PT segment. However, the correspondences are otherwise recurrent, and the semantic match is perfect; the rhotic codas in PMJ are in any case infrequent.

**‘to burn, to set on fire’:** PMJ  $*(-)py_1k^\circ \sim *(-)py_1ŋ^\circ$  : PT  $*-puuk$

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 371. The etymon is preserved in two first-level branches of Macro-Jê, Western (Rikbaktsa  $-pok$  ‘to set on fire’) and Eastern (Maxakalí  $-puk$  ‘to burn (intr.)’, Canela–Krahô  $pôr$ , nonfinite  $-hpôc$  ‘to burn (intr.)’). The correspondences

are regular, including the sound change  $*-k > *-r$  in finite forms of intransitive verbs, typical of the Cerrado languages (cf. Nikulin & Salanova 2019: 544). A difference in valency between the Rikbaktsa verb and its Eastern Macro-Jê cognates is a problem for the comparison, though hardly an insurmountable one. The uncertainty between the reconstruction of  $*-k^\circ$  or  $*-ŋ^\circ$  is due to the absence of a cognate in Krenak, the only Macro-Jê language that is known to preserve the distinction (cf. Nikulin 2020: 159).

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 34. The etymon is preserved in three branches of Tupian: Tuparikém (Wayoró *-pug{a}* ‘to cook’), Rama-Puru (Karo *-pâk* ‘to burn’), and Eastern (Mundurukú *-pik* ‘to burn’).

In all said languages, the verb is a relational class I stem, except for the finite form in the Cerrado languages (finite intransitive verbs are absolute).

### **3CRF prefix:** PMJ $*ta-$ : PT $*tə-$

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 253–260, 383. The etymon is preserved in Karajá (*ta-* with class I stems, *t-* with class II stems) and in two languages of the Western branch (Rikbaktsa *ta-*, Arikapú *ta-*). The correspondences are regular.

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 383. PT  $*tə-$  is preserved in the Tuparikém branch (Wayoró *te-*, Karitiana *ta-*) and in at least one Eastern language (Sateré-Mawé *to-*); a possible reflex with an unexpected vowel is also seen in the Rama-Puru branch (Karo *to-*). The correspondences are otherwise regular. In addition, Mondé and Awetí–Guaranian languages have 3CRF indices that point to PT  $*ə-$  rather than  $*tə-$  (Awetí *o-/w-*, Gavião *a-*; Sabino 2016: 71–72, 146; Moore 1984: 30), a fact I am presently unable to account for.

In both language families, the morpheme in question is a third-person index which signals coreferentiality with another participant (typically the subject). In all languages where it occurs, it can encode the possessor of a relational noun, but in some languages it can also encode the patient of a transitive verb or a complement of an adposition (as in Rikbaktsa), or else the subject of a intransitive verb (as in Wayoró), or of a subclass of intransitive verbs (as in Arikapú). In the latter use, the person index is taken to be coreferential with the noun phrase expressing the subject.

### **‘to give’:** PMJ $*-ũp$ : PT $*-õp$

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 422. Reflexes are found in most Macro-Jê branches: Western (Djeoromitxí  $-õ$ ), Karajá ( $-õ$ ), and Eastern (Khîsêtjê  $-ngõ$ , Maxakalí  $-hõm$ ). The labial stop in the coda position is reconstructed based on evidence from Maxakalí ( $-hõm$  /  $-hũp$ ), Krenak ( $-um$ ), and the nonfinite form found in the Akuwẽ languages (Xavante  $-nh-om-ri$ ). The correspondences are mostly regular, including the consonantal epenthesis in onsetless stressed syllables in Maxakalian and Jê; however, the origins of the voiceless nasal  $m̥$  in Krenak are unclear.

The Proto-Tupian reconstruction is mentioned *in passim* in Nikulin & Carvalho 2022: 16; see Galucio et al. 2015: 258 for a selection of reflexes. The root is preserved in the Tuparikém (Tuparí  $-om$ ) and Eastern (Sateré-Mawé  $-um$ ) branches. The Mundurukú reflex  $-ũm$  /  $-g-ũm$  unexpectedly shows an unrounded vowel / $\tilde{i}$ / (represented as  $\tilde{u}$  orthographically), but the rounded reflex is found in the closely related Kuruaya ( $-õm$  /  $-n-õm$ ; Galucio et al. 2015: 258). The correspondences are otherwise regular.

In both language families, the root is vowel-initial, with no thematic consonant, and is thus classifiable as class I. In the languages of the Cerrado branch of the Jê group, the verb  $*-gõ$  takes indices of the accusative series when finite (just like all monosyllabic canonical transitives), whereas its nonfinite form is a class II stem  $*-ñ-õp-r'$ , with a thematic consonant and a

suffix of nonfiniteness. Since the verb typically takes an inanimate theme, it frequently occurs with a third-person index (PT *\*i-*) in Tupian languages, which typically takes a consonantal allomorph before a vowel-initial root.

**‘to go up, to rise’**: PMJ *\*-we(C)* : PT *\*-we(:)P*

The Proto-Macro-Jê reconstruction is given as *\*-wi(C)* in Nikulin 2020: 382, which is an unfortunate typo (cf. Nikulin 2020: 148): PNJ *\*i* can only go back to PMJ *\*e*. Reflexes are found in the Western (Ofayé *-wi*, possibly Djeoromitxí *{hu}wi*) and Eastern (Khîsêtjê *a-pi*, nonfinite *-tá-pi-ri*). The correspondences are regular. The Northern Jê reflexes continue PNJ *\*a:-pi*, nonfinite *\*-jə:-pi-r*; the alternating prefixes *\*a:-* (finite) and *\*jə:-* (nonfinite), found in a number of intransitive verbs, are of unclear origin, but they are clearly distinct from the antipassive prefixes *\*a-/ap-* (finite) and *\*jə-/\*ju-* (nonfinite), which has a short vowel. The absence of clear cognates in diagnostic languages, such as Maxakali, Krenak, Xavante, or Akwê-Xerente, makes it impossible to determine whether the Proto-Macro-Jê verb had a final consonant. If Maxakali *-ã-pep/-xu-pep* ‘to leave/arrive.SG’ is cognate, the PMJ reconstruction can be amended to *\*-wep ~ \*-wem<sup>o</sup>*, but the semantic discrepancy renders the comparison uncertain.

The Proto-Tupian reconstruction is based on reflexes such as Wayoró *-ngwep* (< Proto-Tuparian *\*-wep*; Nikulin & Andrade 2020: 299), Karitiana *-haap* ‘to rise (of the sun)’ (< Proto-Arikém *\*-hä:P*), Paiter *-web-á* ‘to swell’, and Awetí *-tep* (attested in Reiter 2011: 205). The correspondences are regular except for the mismatch between the short vowel in Tuparian and the long vowel in Arikém.

The class membership of Proto-Macro-Jê *\*-we(C)* is difficult to determine based on direct evidence: the Ofayé reflex is only marginally attested, whereas in other languages only a prefixed derivative was preserved. In Tupian, the verb is a class I stem.

#### 4.2. Good distribution in Macro-Jê only

**‘hole’**: PMJ *\*-kuñ<sup>o</sup>* : Proto-Mundurukuan *\*-kāj*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 411. Reflexes of the bare root are found in the Western (Djeoromitxí *-kü*) and Eastern (Maxakali *-kox*, Canela–Krahô *kwa* ‘well, spring’) branches. In derivatives, such as the terms for ‘sky’ and ‘mouth’, it is preserved in even more languages, as in Karajá (female speech *biku*, male speech *biu* ‘rain, sky’), Khîsêtjê (*-jajkhwa* ‘mouth’), or Kaingang (*jānkā* ‘door’).

The Proto-Mundurukuan form, reflected in Mundurukú as *-kũy /-kẽj/* and in Kuruaya as *-kāj*, is from Picanço 2019: 136. It lacks known cognates in other Tupian languages. If it is shown to be of Proto-Tupian original, its PT etymon must have been *\*-kãʔãC*, *\*-kãʔãC*, *\*-kãʔãC*, *\*-kãʔãC*, *\*-kãʔãC*, or *\*-kãʔãC*. Two alternative candidates for the Proto-Tupian term for ‘hole’ are *\*-kãʔã* and *\*-kãP*, but both have their distribution limited to two Rondonian branches only. The former has reflexes in Tuparikém (Tupari *-apsi-kúm’e* ‘inner ear’, *-kúm’e* ‘vagina’, Akuntsú *-api-tep-kímã* ‘inner ear’ with an irregular final vowel, Karitiana *emã ~ emmã* ‘pit’) and Mondé (Paiter *-îwã* ‘hole, buttock’). The latter has reflexes in Tuparikém (Karitiana *-’op* ‘hole, channel’ and Kepkiriwat (‘uapicáp’ /u-api-kaP/ ‘inner ear’).<sup>13</sup>

In both Macro-Jê and Mundurukuan, the roots in question are relational class I stems.

<sup>13</sup> Otherwise, each Tupian branch employs its own root(s) for the meaning ‘hole’: Makurap *pun*; Tupari *-áu’am*; Karitiana *-’op*; Karo *-xâk ~ Puruborá fEK*; Aruá *⟨ñiñap⟩*; Proto-Juruna *\*-ku(-)á* and *\*karapú*; Sateré-Mawé *-kaʔa*; Proto-Awetí–Guaranian *\*-k<sup>m</sup>aT*.

It must be noted that PMJ *\*-kuĩ*<sup>o</sup> shows similarity with yet another root, found in Mondé only: Paiter *-koy* in *ḡôy-koy* ‘pit’ (from *ḡôhy* ‘earth’). This root cannot be cognate with Proto-Mundurukuan *\*-kāj*, and could in principle be equated, at least etymologically, with the directional suffix *-koy* ‘towards’.

**‘ripe’:** PMJ *\*-ndêp*<sup>o</sup> : Tuparí *-tep*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 390. The etymon is best known for its reflexes in the Eastern branch (Maxakalí *tep-ta* ‘banana’, Khîsêtjê *-ndep-txi* ‘red’, Apinajé *-nep* ‘ripe’, Canela–Krahô *-ntep-ti* ‘ripe, red’), but a likely reflex is also found in the Western branch (Rikbaktsa *-{ne}ne* ‘ripe’).

Tuparí *-tep* is documented, for example, in Alves 2004: 257, 258. No cognates in other Tupian languages are known. It can technically go back to a variety of forms, such as PT *\*-teP*, *\*-təP*, *\*-ndeP*, *\*-ndəP*, *\*-ðeP*, or *\*-ðəP*. In many other Tupian languages, the concept ‘ripe’ is expressed by a reflex of *\*-woP* ‘red, ripe’ instead: compare Wayoró *-ngop* ‘red, ripe’ (Nogueira et al. 2021: 103), Akuntsú *-kop* ‘red, ripe’ (Aragon 2014: 104, 131), Paiter *-ób* ‘red, ripe’ (Bontkes 1978: 14), Yudja *-upa* ‘ripe’ (Fargetti 2001: 281–283), Mundurukú *-op* ‘ripe’ (Crofts 1985: 99), etc. It is possible that *\*-woP* was primarily used a color term, whereas the highly hypothetical form PT *\*-teP*, *\*-təP*, *\*-ndeP*, *\*-ndəP*, *\*-ðeP*, or *\*-ðəP* could have been a dedicated term for ‘ripe’, ousted in most daughter languages by reflexes of *\*-woP*.

Both in Jê and Tuparí the stem in question is a relational class I stem. Maxakalí *tep-ta* is an absolute stem, and Rikbaktsa *-{ne}ne* is an intransitive verb (the language no longer has a class I / class II distinction).

**‘to kill’:** PMJ *\*-wĩ* : Karo *-wĩ*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 383. Reflexes are found in the Western (Ofayé *-wĩ* ‘to shoot’), Karajá (*-wè-* ‘to sting, to penetrate, to stab’), and Eastern (Ritual Maxakalí *-mĩ-y*, irrealis *-mĩ* ‘to kill’, Maxakalí *-mĩ-y*, irrealis *-mĩ* ‘to make’, Khîsêtjê *-pĩ*, nonfinite *-pĩ-rĩ* ‘to kill.SG’, Xavante *-wĩ*, nonfinite *-wĩ-rĩ* ‘to kill.SG’). The correspondences are regular. The meaning ‘to kill’ is attested only in Ritual Maxakalí and in most languages of the Cerrado branch (singular only), whereas Ofayé, Karajá, and spoken Maxakalí all show deviant meanings. Even the Cerrado languages Canela–Krahô and Pykobjê–Křikati do not use the reflexes of PMJ *\*-wĩ* as the basic verb for ‘to kill’; instead, they are used figuratively, e.g. as ‘to extinguish a fire’, ‘to kill by drowning (of water)’, ‘to suffocate’.

The Karo verb *-wĩ* ‘to kill’ (Gabas Jr 1999: 48, 57) can technically go back to PT *\*-wĩ*, *\*-wĩC*, *\*-wĩ̃*, or *\*-wĩ̃C* (note that PT *\*ĩ* and *\*ĩ̃* merge in all Tupian languages except Juruna and Mawé–Guaranian, whereas *\*C* is deleted after a front high vowel in these languages; see the cognate set for ‘heavy’ in 4.3). It is likely related to Puruborá *-wi* ‘to kill’ (Galucio et al. 2015: 257), but the absence of vowel nasality in the putative Puruborá cognate is unaccounted for. Karitiana *-mĩ* ‘to beat’ is technically comparable with Karo *-wĩ* ‘to kill’, given that *\*m* and *\*w* merge as *m* before nasal vowels in Arikém, but it could likewise be cognate with Proto-Tuparian *\*-mĩ* (> Tupari *-mĩ* ‘to stab, to sting’, Sakurabiat and Akuntsú *-mĩ* ‘to kill’); in the latter case, the Proto-Tuparikém form must be reconstructed as *\*-mĩ̃*, thus showing no regular correspondence with Karo *-wĩ̃*. Alternatively, one could reconstruct Proto-Tupian *\*-wĩ̃*, *\*-wĩ̃C*, *\*-wĩ̃̃*, or *\*-wĩ̃̃C* based on reflexes in the Rama-Puru and Tuparikém branches and posit an irregular sound change *\*w > \*m* in Proto-Tuparikém, Proto-Tuparian, or Proto-Core Tuparian. If such a verb existed in Proto-Tupian, it was likely distinct from PT *\*-ʔaoka* ~ *\*-ʔaok̃a* ‘to kill, to beat’, with reflexes in Kepkiriwat, Mondé, Mundurukuan, and Mawé–Guaranian, in that the latter



prototypically referred to beating to death, whereas the former probably referred to killing by stabbing or piercing (e.g. with an arrow), as suggested by the Tuparí reflex.

In Macro-Jê, the root is a class I stem. In the languages of the Cerrado branch of the Jê group, its reflexes take indices of the accusative series when finite (just like all monosyllabic canonical transitives), whereas its nonfinite form is PCerr *\*-wĩ-r'*. Karo does not have a class I/class II distinction.

#### 4.3. Good distribution in Tupian only

**'bitter':** PT *\*-ðəP* : PCerr *\*-ndap* 'sour, bitter'

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 27, with reflexes in the Tuparikém (Tuparí *-tép-'a* 'bitter', *-tép-'ut* 'sour', Karitiana *-taap*) and Eastern (Mundurukú *-cúp*, Sateré-Mawé *-nop*, Awetí *-lop*, Apyãwa *-rap*) branches, as well as possibly in Mondé (Paite *-{pe}txáb*). The correspondences involving the nucleus and the coda are regular. However, the correspondence between Proto-Tuparian/Proto-Arikém *\*t* and Sateré-Mawé *n*, Awetí *l*, PTG *\*r* is unprecedented: Proto-Tuparian/Proto-Arikém *\*t* points to PT *\*t*, Sateré-Mawé *n* suggests PT *\*nd* (allophone of *\*/n/*), whereas Awetí *l* : PTG *\*r* is not known to go back to any specific PT consonant<sup>14</sup>. The Mundurukuan reflexes are uninformative, since *\*t* and *\*nd* are not otherwise distinguished in Mundurukuan. I reconstruct *\*ð* for the correspondence in question and assume that it was a low-frequency phoneme in Proto-Tupian, just like its apparent reflex /l/ in Awetí. However, other solutions are also imaginable, such as the existence of a hypothetical alternation between the allomorphs *\*-təP* (relational) and *\*-ndəP* (absolute), with the subsequent generalization of the former in Tuparikém and of the latter in Mawé–Guaranian.

The Proto-Cerrado reconstruction is from Nikulin 2020: 456. Reflexes include Akwẽ-Xerente *-{wam}t(a)pa/-{wam}tap* 'bitter, sour', Mẽbêngôkre *-nap* 'sour', Pykobjê–Křĩkatí *-ntap* 'sour, ripe', and possibly Khĩsêtjê *-ndap* // *-ndawy* 'smooth'. The correspondences are regular. The semantic reconstruction is not straightforward: there are other terms for 'sour' and 'bitter' in the Goyaz languages, which are also quite old: Proto-Goyaz *\*-jwa* 'sour' and *\*-jê* 'bitter' (< PJ *\*-jô<sup>K</sup>* 'sour, salty', *\*-jê<sup>K</sup>* 'bitter'). If PCerr *\*-ndap* is shown to be of Macro-Jê origin, its erstwhile form should be reconstructed as PMJ *\*-ndap<sup>o</sup>* or *\*-ndâp<sup>o15</sup>*.

Both PT *\*-ðəP* and Proto-Cerrado *\*-ndap* are reconstructed as relational class I stems.

**'to do, to say, to be like this':** PT *\*-ke* : PSJ *\*kê* // *\*ke*

The Tupian reconstruction is based on reflexes in the Tuparikém (Makurap *-ke*, Karitiana *-'a*), Rama-Puru (Karo *-'e*), and Eastern (Sateré-Mawé *-'e*, Apyãwa *-'ẽ/-e*) branches. The correspondences are regular, except that the Awetí–Guaranian branch has innovated some irregular inflected forms: the third-person form is reconstructed as *\*eʔi* (rather than the expected *\*\*o-ʔe*), whereas the second-person singular form is attested as *eʔi* (rather than *\*e-'e*) in Awetí and reconstructed as *\*ere* (rather than *\*\*ere-ʔe*) in PTG. Although other forms are regular (PTG 1 *\*a-ʔe*, 1+2 *\*ja-ʔe*, 1+3 *\*oro-ʔe*, 2+3 *\*peʔj-e*), some daughter languages show extra irregularities, such as

<sup>14</sup> It is conceivable that Awetí *l* and PTG *\*r* are the regular reflexes of PT *\*nd*. Unfortunately, PT *\*ndo(:)* 'hill' and *\*ndoK* 'to eat.INTR' lack known reflexes in these languages, making it difficult to determine the evolution pathways of PT *\*nd* in the Awetí–Guaranian branch.

<sup>15</sup> Other Macro-Jê languages show noncognate forms for 'sour' and 'bitter': Krenak *-rə* 'sour', *-ñãngrok* 'bitter'; Maxakalí *-xupyãg* 'sour', *-xũĩy* 'pain, sour, bitter, spicy'; Karajá 3 *tx-ubrèrè* 'sour'; Ofayé 3 *h-əfè* 'sour', 3 *ə-xahtə* 'bitter'; Rikbaktsa *-bui* 'sour', *-sikpia* ~ *-spia* 'bitter'; Proto-Jabutian *\*-jombi* 'pain; sour', *\*-wəwə* 'bile' (whence Djeoromitxí *-wäwä-rü* 'bitter') or Arikapú *-oay* ~ *-way* 'bitter'; Proto-Chiquitano *\*ókor-* 'to be sour', *\*pičar-* 'to be bitter'.

the nasalization  $*e > \tilde{e}$  in the Apyãwa forms  $\tilde{a}-\tilde{e}$ ,  $xa-\tilde{e}$ ,  $ara-\tilde{e}$  or the analogical vowel raising in Guarasugwe ( $\acute{e}ri$ ,  $\acute{i}-\tilde{?i}$  instead of  $*\acute{e}re$ ,  $*\acute{e}-\tilde{?i}$ ), Kawaiwete ( $a-\tilde{i}$  instead of  $*a-\tilde{e}$ ), or Kamayurá ( $i-\tilde{i}$  instead of  $*e-\tilde{i}$ ).

On the Macro-Jê side of the comparison, one finds only PSJ  $*kê$  //  $*ke$  (the latter allomorph appears utterance-finally due to a general process of vowel lowering), reflected, for example, in Kaingang  $ke$  //  $ké$  ‘to do, to say’. This verb lacks known cognates elsewhere in Macro-Jê. It is semantically close to its Tupian counterparts in that it is used both for actions and speech acts. However, it cannot be a Tupian loan, since the only Tupian languages that have a velar reflex of PT  $*k$  – Tuparian and Kepkiriwat – are spoken 1,500 km northwest from the Southern Jê-speaking zone. If it goes back to Proto-Macro-Jê, the protoform must have been  $*-ki(C)$ .

**‘white’:** PT  $*-k̃iɾ$  : PCerr  $*-ka:$

The Proto-Tupian reconstruction is based on reflexes in Rama-Puru (Karo  $-kûɫ$ ), Mondé (Paiteir  $-kír$ ), Eastern (Sateré-Mawé  $-kytí'i$ ,  $-kyt\{si\tilde{g}\}$ ), and apparently Tuparikém (Wayoró  $\{-y\}ir\{a\}$  ‘white’, though the main root for ‘white’ in Tuparikém is  $*-pa(:)k$ ); see Galucio et al. 2015: 260 for more reflexes.

The Proto-Cerrado form for ‘white’ is reconstructed as  $*-ka$  by Nikulin (2020: 467), who does not recognize the existence of contrastive vowel length in that protolanguage. However, it is now clear that long vowels in Pykobjê–Křikatí (and Canela–Krahô, whose long vowels are however not so thoroughly documented) correspond to long vowels in Xavante, where they are preserved utterance-medially only, as documented by McLeod & Mitchell (1977). Therefore, long vowels must have existed in Proto-Cerrado. Xavante  $-ʼa$  /-ʼa:/ ‘white’ has an underlying long vowel, as seen in the example  $tsiʼa$   $hã$   $piʼõ$  [si:ʼa: hã piʼõ] /ci:-ʼa: hã piʼõj/ ‘the chicken (lit. white bird) is female’ (McLeod & Mitchell 1977: 107), and so does Pykobjê–Křikatí  $-jaca$  /-jak<sup>h</sup>a:/ ‘white’. Other reflexes include Khīsêtjê  $-jakha$ , Měbêngôkre  $-jaka$ , and Akwē-Xerente  $-ka$ . The Northern Jê reflexes contain the element  $*-ja-$ , which could have historically been a plural prefix. The updated Proto-Cerrado reconstruction is, therefore,  $*-ka:$ . No cognates elsewhere in Macro-Jê are known, but no stronger candidates for the Proto-Macro-Jê root for ‘white’ are known either<sup>16</sup>. If  $*-ka:$  is shown to be of Macro-Jê origin, its protoform can be reconstructed as  $*-kaC^\circ$  or as  $*-kâC^\circ$ , with an unidentified coda.

In both language families, the term in question is a class I relational stem.

**‘husband’:** PT  $*-mēɾ$  : PMJ  $*-mbi_2n$  (Eastern)

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 32. Reflexes are found in the Tuparikém (Wayoró  $-mēn$ , Karitiana  $-man$ ), Rama-Puru (Puruborá  $-mēɾ$ ), Mondé (Gavião  $-met$ ), and Eastern (Yudja  $-mená$ , Sateré-Mawé  $-men$ , Apyãwa  $-men-a$ ) branches. The correspondences are regular.

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 380. Reflexes are found in the Eastern branch only and include Maxakalí  $-pit$  ‘male’, Khīsêtjê  $-mdjên$  //  $-mdjēni$  ‘husband’, Panará  $inp̃in-pjâ$  ‘husband’, and Kaingang  $-mén$  ‘husband’. The correspondences are regular.

In both language families, the noun in question is a class I relational noun.

**‘I’:** PT  $*o-$  : PCerr  $*wa$

In Proto-Tupian,  $*o-$  is reconstructed as an absolutive/genitive first-person prefix, from which the pronoun  $*ōɾ$  is derived, just like the pronoun  $*ēɾ$  is derived from the second-person prefix  $*e-$ .

<sup>16</sup> Each Macro-Jê branch employs its own root(s) for this meaning: Krenak  $-jirum$ ; PSJ  $*kupri$ ; Proto-Karajá  $*-kûrã$ ; Ofayé  $-kãē?$  and  $-gãte?$ ; Rikbaktsa  $-baraza$ ; Arikapú  $-mãw$  ~  $-mão$ ; Djeoromitxí  $-kãñõrũ$ ; Proto-Chiquitano  $*purusuβii$ .

It has reflexes in almost all Tupian languages. Before consonant-initial stems, it is reflected as Kepkiriwat <u->, Makurap *o-*, Wayoró *o-*, Karitiana *y-*, Karo *o-*, Puruborá *ɔ-*, Paiter *o-*, Yudja *u-*, Mundurukú *o-*, Sateré-Mawé *u{ij}-*, Apyãwa *w{e}-* ‘1CRF’, among many other reflexes. Before vowel-initial stems, it shows asyllabic allomorphs in some languages, such as Wayoró *m(b)-/∅-* (before rounded vowels) or Yudja *w-/∅-*. In Mawé–Guaranian, it is unexpectedly reflected as *\*u<sub>C</sub>-* rather than *\*u-*. The TG reflex is only used anaphorically, particularly when a first-person possessor on a noun or a first-person argument of a gerund of an intransitive verb is coreferential with some other participant. As for noncoreferential uses, it has been ousted by the clitic *\*ice=* in the TG languages.

The Proto-Cerrado pronoun *\*wa* ‘I’ is reflected as Xavante *wa hã*, Akwẽ-Xerente *wa (hã)*, Khĩsêtjê *pa* (topical) / *wa* (nominative), Mẽbêngôkre *ba*, Canela–Krahô *pa* (topical) / *wa* (nominative), among other reflexes. It is erroneously reconstructed as *\*waj*’ in Nikulin 2020: 451, where the palatal coda is claimed to have been present in the reconstructed form based on the Akwẽ-Xerente reflexes *waĩmẽ* ‘with me’, *waĩtê* ‘mine’, mistakenly segmented as *waĩ-mẽ*, *waĩ-tê*. Instead, the correct segmentation must be *wa=ĩ-mẽ*, *wa=ĩ-tê*, where *ĩ-* is a first-person prefix preceded by the cliticized pronoun *wa*. Compare also the second-person forms *kaimẽ* ‘with you’ and *kaitê* ‘yours’, analyzable as *ka=ai-mẽ*, *ka=ai-tê*, where *ka* is a pronoun and *ai-* is a second-person prefix. Some Northern Jê languages show distinct reflexes of *\*wa* when stressed (topical) and unstressed (nominative); at least in Khĩsêtjê this is the expected consequence of the conditioned split that affected PNJ *\*b*.

Proto-Cerrado *\*wa* has no known cognates in other Macro-Jê languages. If it is shown to go back to Macro-Jê, its original form can be stipulated to have been *\*u(C)*. Nikulin (2020: 187–193) reconstructs a case paradigm consisting of PMJ *\*iĩ* (first-person internal case pronoun) and *\*a* (first-person agentive case pronoun), but does not discard the possibility that the pronominal case paradigm included even more cases. It is possible that Proto-Cerrado *\*wa* reflects a PMJ first-person pronoun inflected for some other case, whose original function is yet to be identified.

#### **‘to wake up’**: PT *\*-pa<sub>K</sub>* : Proto-Jabutian *\*-pa*

Proto-Tupian *\*-pa<sub>K</sub>* is reconstructed based on its reflexes in the Tuparikém (Wayoró *-{e}pak*), Rama-Puru (Karo *-{pe}pak*), Mondé (Paiter *-páká-tẽ* ‘to wake smb. up’, *-pák{o}* ‘to be awake’), and Eastern (Yudja *-pak-*, Apyãwa *-pãk*) branches. The correspondences are regular. The elements *e-* in Tuparian and *pe-* in Karo are, at least etymologically, intransitivizing and impersonal passive markers, respectively.

The Proto-Jabutian reconstruction is from Nikulin 2020: 542. The root is preserved both in Arikapú and Djeoromitxí as *-pa*. It is hardly borrowed from the neighboring Tuparian languages, since all Tuparian languages show the element *e-* found in Wayoró. If the Jabutian root is shown to be of Macro-Jê origin, the protoform can be reconstructed as *\*-pa(C)*, *\*-pə(C)*, or possibly *\*-pỹ(C)*. No stronger candidates for the Proto-Macro-Jê root for ‘to wake up’ are known<sup>17</sup>.

Both in Tupian and Jabutian the verb is a relational class I stem.

#### **‘heavy’**: PT *\*-pəti<sub>C</sub>* : Maxakalí *-ptux*

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 34. Reflexes are found in most Tupian languages, including Tuparikém (Makurap *-poti*, Karitiana *-pyti*), Rama-Puru

<sup>17</sup> Each Macro-Jê branch employs its own root for this meaning: Krenak *-mrət* ~ *-mrək* (intr.), *-mrəŋ* (tr.); Maxakalí *-koxa-k*, irrealis *-koxa* (tr.); PJ *\*-rüt*° ‘to wake up, to look’ (intr.), cf. also PNJ *\*-mbra*: (nonfinite *\*-mbra:-r*) ‘to wake up’ (tr.); Proto-Karajá *\*-eθi-θã* (intr., with the reflexive prefix *\*eθi-*); Ofayé *-xêrê-ge* / *-xêhõê* ~ *-xêhõj*; Rikbaktsa *-popo* (tr.); Proto-Chiquitano *\*syto-pyr-* (intr., from *\*-sýto* ~ *\*-súto* ‘eye’).

(Karo *-pi'ti*), Mondé (Paiter *pai-ga*), and Eastern branches (Yudja *-padít-*, Sateré-Mawé *-potyi*, Apyãwa *-pooj*). The correspondences are mostly regular, except that Wayoró *-pooti* ~ *-poti* (Nogueira et al. 2021: 105) has an unexpected variant with a long vowel; Mundurukú *-poxí* (Picanço 2005: 264) has an irregular rounded vowel (unlike in the closely related Kuruaya); and Karo *-pi'ti* (Gabas Jr 1999: 15) has an unexpected vowel in the initial syllable followed by a glottal stop. In addition, the vowel of the final syllable has been attested as *e* in the Juruna branch (Yudja *-padét-*, Xipaya *-padet-*; Fargetti & Rodrigues 2008: 562), but at least the Yudja form is mistranscribed in that source. The actual Yudja form is *-padít-*, as attested elsewhere (Lima 2014: 28) and confirmed by native speakers.

On the Macro-Jê side, one finds Maxakalí *-ptux* /-ptiC/ ‘heavy’ (Silva 2020: 96), with the allomorph *-putux* occurring after consonants. It lacks known cognates elsewhere in Macro-Jê, though technically it can be quite old, given that no other Proto-Macro-Jê term for ‘heavy’ can be reconstructed<sup>18</sup>. The hypothetical PMJ form could then start with *\*pr*, *\*m(b)r*, *\*pVt*, *\*pVn(d)*, *\*pVr*, *\*mbVt*, *\*mbVn(d)*, *\*mbVr*, *\*wVt*, *\*wVn(d)*, or *\*wVr*; the nucleus could be either *\*â*, *\*y*, or *\*ÿ*; the coda could be any palatal coda, with or without an echo vowel. It is unlikely that the Maxakalí form was borrowed from Tupian. Although Maxakalí has a handful of well-known loanwords from a Tupian language, these come from Old Tupí, or from a closely related variety (Ribeiro 2012b: 91). However, Old Tupí, just like all TG languages, does not preserve Proto-Tupian *\*t* as a stop, and has the form *-posyî* /-pɔsiC/ ‘heavy’ as the reflex of PT *\*-pat̪iC*. Such a form would have been borrowed into Maxakalí as *\*-poxux* \*/-puciC/, or perhaps as *\*-pxux* \*/-pciC/ (assuming a diachronic loss of unstressed /u/, as in *-pxet* ‘one’ and *-ptox* ‘head’; see Silva & Nikulin 2021: 36). From a phonological point of view, non-TG Tupian languages would be more suitable candidates, but all of these languages are spoken thousands of kilometers west of the current Maxakalí area.

**‘to go’, ‘to come’:** PT *\*-tēP* ‘to exit’, *\*-ʔatēP* ‘to arrive’  
: PMJ *\*tē* (nonfinite *\*-tē-m* or *\*-tē-n*) ‘to go, to come’ (Eastern)

The Proto-Tupian reconstruction *\*-tēP* ‘to exit’ is based on reflexes in the Tugarikém (Karitiana *-tam* ‘to fly’) and Eastern (Mundurukú *-cēm*, Sateré-Mawé *-tem*, Awetí *-tem*, Kawaiwete *-em*) branches. The Proto-Tupian reconstruction *\*-ʔatēP* ‘to arrive’ is based on reflexes in the Tugarikém (Karitiana *-otam*), Rama-Puru (Puruborá *-anēm-ā*), and Eastern (Mundurukú *-àjēm*, Awetí *-{to}tem*, Apyãwa *-{w}aem*) branches. The erstwhile presence of PT *\*ʔ* is recoverable based on the creaky voice in Mundurukú. All TG languages show a fossilized element *\*w-*, which is likely to have originated in an active third-person prefix. The correspondences are regular for both verbs. The former appears to have split into two different verbs in the Guaranian branch of TG: *\*-θēP* (> Tapiete, Mbyá *-ē*) and *\*-cēP* (> Paraguayan Guaraní *-sē*, Mbyá *-xē* ‘to leave definitely’), with Mbyá showing reflexes of both with different meanings. This is likely a result of horizontal transmission between (pre-Proto-)Guaranian varieties. PT *\*t* is otherwise known to have two reflexes in Proto-Guaranian in the default position, which I reconstruct as *\*θ* and *\*c*, but the conditioning environments for this purported split have not been established so far<sup>19</sup>.

<sup>18</sup> Each Macro-Jê branch employs its own root for this meaning: Krenak *ṃukran* ~ *ṃukranj*; Proto-Goyaz *\*-pyt̪i*; Proto-Akuwẽ *\*-pirê*: // *\*-pirê*; PSJ *\*kuθy*; Proto-Karajá *\*-kut̪i*; Ofayé *-wencãw̃* ~ *-encãw̃*; Rikbaktsa *-tsakyrik*; Proto-Jabutian *\*-kômỹ* ~ *\*-kumỹ*; Proto-Chiquitano *\*-ũmĩ*/*\*-ũmĩ*. Despite the superficial similarity between the Maxakalí, Proto-Goyaz, and Proto-Akuwẽ forms, as well as between the PSJ and the Proto-Karajá one, none of them are conceivably cognate with each other because of lack of regular correspondences involving vowels.

<sup>19</sup> Some authors have proposed that the distinction between these two consonants is quite old, and project it to the Proto-Tupi-Guaranian (Carvalho 2022) or even Proto-Tupian (Rodrigues 2007) stage. Others assume

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 387. Reflexes are found in the Eastern branch only and include Maxakalí *-nũn* (irrealis *nũ*) ‘to come’, Krenak *-nĩŋ* (imperative *nĩ*), Khĩsêtjê *thẽ* (nonfinite *-thẽm*) ‘to go/come.SG’, Xavante *-nem* ‘to go/come.DU’ (nonfinite only), and Kaingang *tĩ* (nonfinite *-tĩg*) ‘to go/come.DU’. The correspondences are mostly regular, except that the nonfinite form in Jê points to PMJ *\*-tẽ-m*, whereas the realis/indicative forms found in the Trans-São Francisco languages Maxakalí and Krenak point to PMJ *\*-tẽ-n*. In the Jê languages, the verb in question is used as the generic movement verb (the concepts ‘to go’ and ‘to come’ are distinguished by means of centrifugal and centripetal particles), restricted to singular subjects in the Goyaz and Southern branches and to dual subjects in the Akuwẽ branch. For plural subjects, the verb *\*mũ<sub>1</sub>* ‘to go/come.PL’ is used. In the Trans-São Francisco branch, the opposition between the cognates of *\*tẽ* and *\*mũ<sub>1</sub>* is not that of number, but rather of direction: PJ *\*tẽ* ‘to go/come.SG’ corresponds to *\*nẽ-n* (irrealis *\*nẽ*) ‘to come’, whereas *\*mũ<sub>1</sub>* ‘to go/come.PL’ corresponds to *\*mũ-ŋ* (irrealis *\*mũ*) ‘to go’.

The Tupian verbs are class I verbs. In Macro-Jê, the finite stem is reconstructed as absolute (uninflectable), and its nonfinite counterpart is a class I relational stem. The proposed match is between the nonfinite stem in Macro-Jê and the invariable stem in Tupian; note that Tupian does not have a systematic finiteness distinction in verbal stems except for a handful of verbs in TG, which are usually referred to as irregular verbs (cf. Barbosa 1956: 305–309).

**‘to arrive’:** PT *\*-wũc* ‘to go out, to arrive’ : PCerr *\*wôc*, nonfinite *\*-wôc*

The Proto-Tupian reconstruction is based on reflexes in the Tuparikém (Karitiana *-hej* ‘to go away, to abandon’), Rama-Puru (Karo *-wũy* ‘to go out’, Puruborá *-wi* ‘to go out’), and Eastern (Yudja *-wĩ* ‘to arrive’) branches. The correspondence between the onset consonants and the vowels is regular. The fact that Karitiana and Karo show a palatal coda, absent in Puruborá and the Juruna languages, remains unexplained. An identical correspondence is observed in the Rama-Puru cognate set for ‘to wait’ (Karo *-pũy*, Puruborá *-bĩ*), suggesting that at least Puruborá may have regularly lost the palatal coda after an *i*. The polysemy ‘to go out’ / ‘to arrive’ is common in the region, and is attested in languages such as Canela–Krahô (*-cato*) or Maxakalí (*-xu-pep* / *-ã-pep*, singular only).

The Proto-Cerrado form is reconstructed as *\*wôj* (finite), *\*-wôc* (nonfinite) in Nikulin 2020: 451 based on reflexes such as Khĩsêtjê *pãji* (finite), *-pôt* (nonfinite) and Xavante *wi* (finite, singular only), *-witsi* (nonfinite, singular only). However, the reconstruction can be amended to *\*wôc* (finite), *\*-wôc* (nonfinite). The Northern Jê languages show a regular lenition of the stem-final stop in the finite form, yielding the reflex *\*bôj* as opposed to the nonfinite form *\*-bôc* (Nikulin & Salanova 2019: 544). In the Akuwẽ languages, the finite form is reconstructed as *\*wi* ‘to arrive.SG’, but the loss of *\*-c* is expected in the finite form, since finite forms only occur clause-finally in Akuwẽ, and the utterance-final allophone of Proto-Akuwẽ *\*/c/* is zero.<sup>20</sup> No cognates in other Macro-Jê languages are known, but if *\*(-)wôc* is shown to be of Macro-Jê origin, the respective PMJ protoform must have been *\*(-)wɣ<sub>1</sub>c<sup>o</sup>*.

The Tupian verbs are class I verbs. In Macro-Jê, the finite stem is reconstructed as absolute (uninflectable), and its nonfinite counterpart is a class I relational stem.

---

that the distinction is a relatively recent innovation restricted to the Guaranian branch (cf. Schleicher 1998, Meira & Drude 2015, Nikulin & Carvalho 2022). The existence of doublets such as *\*-θẽp*/*\*-cẽp* ‘to leave’ suggests that the distinction between *\*θ* and *\*c* cannot continue an ancient Proto-Tupi–Guaranian or Proto-Tupian opposition.

<sup>20</sup> In the utterance-medial position, Proto-Akuwẽ *\*/c/* surfaces as *\*cV*, as in *\*-(<sup>n</sup>)pêcê* ‘good’, *\*-puci* ‘to leave.DU.NF’, *\*-ŷaci* ‘to enter.DU.NF’. These stems surface as *\*-(<sup>n</sup>)pê*, *\*-pu*, *\*-ŷa* in the clause-final position.

## 4.4. Limited distribution in both families

**‘bat’**: Proto-Goyaz *\*nĵêp* : PT *\*juɸ* (Kepkiriwat and Mondé)

Proto-Goyaz *\*nĵêp* ‘bat’ is reconstructed based on reflexes such as Khîsêtjê *ntêp-txi* and Panará *{na}nsêpi* (attested as <incêp> in the early 20<sup>th</sup> century). The root lacks known cognates in other Macro-Jê languages. Nikulin’s (2020: 463) comparison of Proto-Goyaz *\*nĵêp* ‘bat’ with Proto-Akuwê *\*cibi* // *\*ci:bi* ‘tarantula’ must be rejected not only for semantic, but also phonological reasons: the expected cognate of Proto-Goyaz *\*nĵêp* in Proto-Akuwê should have the form *\*\*cipi* // *\*\*ci:bi* (underlying *\*\*/cip/*). Since there are no stronger candidates for the Proto-Macro-Jê term for ‘bat’,<sup>21</sup> it is possible that Proto-Goyaz *\*nĵêp* is a retention from the hypothetical Proto-Macro-Jê form *\*nĵip*<sup>o</sup>.

In Tupian, similar terms for ‘bat’ are found in at least two Rondonian branches, Kepkiriwat (<jêp>) and Mondé (Paiter *líhb*, Zoró *djîp*, among other reflexes; Proto-Mondé *\*nĵi:P*). In João Barbosa de Faria’s notes, <e> or <ê> may stand for Kepkiriwat /i/, a sound transcribed by Cândido M. S. Rondon as <u> (compare Barbosa de Faria’s <queitaá> and Rondon’s <cuitá-á> ‘earth’). Therefore, the Kepkiriwat term for ‘bat’ can be restituted as /jîp/ (/jêp/ is another possibility, but this does not correspond to Proto-Mondé *\*nĵi:P*). A possible cognate in the Rama-Puru branch is Puruborá *ŝip{ê}* (Monserrat 2005: 16), though the morphological segmentation is unclear. A much weaker candidate for the Proto-Tupian term for ‘bat’ is seen in the Tuparikém (Makurap *waxariax*, Wayoró *ngwaria*, Tup *wári’a*, Sakurabiat *kwarisa*, Karitiana *asori*, Arikém /ɸjɸri/) branch and in Awetí (*tati’a*). However, the correspondences are entirely irregular: Core Tuparian languages point to PT *\*wari?a* ~ *\*wari?a*, Makurap to *\*wajari?aC* ~ *\*wajari?aC*, Karitiana to *\*wejari* ~ *\*wejari*, Arikém to *\*aiari* ~ *\*aiari* or the like, and Awetí to *\*wake?a*. This etymology plausibly involves extensive horizontal transmission rather than cognation.

It is unlikely that the similarity between Proto-Goyaz, Kepkiriwat, and Mondé forms is due to contact. Note that the Goyaz languages are mostly spoken more than 1,000 km east from the Kepkiriwat- and Mondé-speaking area. An exception is constituted by the westernmost Goyaz languages, Kajkwakhrattxi (until the 20<sup>th</sup> century) and Khîsêtjê (until the 19<sup>th</sup> century), which used to be spoken in the Tapajós River basin, some 300 km east from the easternmost Mondé territory. However, Kajkwakhrattxi and Khîsêtjê are known to be newcomers in that region; moreover, these languages characteristically reflect Proto-Goyaz *\*nĵ* as *nt* /<sup>n</sup>t/.

Both in Goyaz and Tupian, the term for ‘bat’ is an absolute stem.

**‘to dig’**: PMJ *\*-kut* (Eastern only) : Proto-Mundurukuan *\*-je-kot*

The Proto-Macro-Jê reconstruction, taken from Nikulin 2020: 411, is based on reflexes restricted to the Eastern branch, such as Maxakalí *-kot*, Khîsêtjê *-khwâ* (nonfinite *-khwân*), Laklânô *ka* ‘to dig’. The correspondences are regular, except that in the Northern Jê languages the finite form (PNJ *\*-kwâ*) was analogically remodeled based on the regular nonfinite form *\*-kwâ-ñi*; the expected reflex of the finite form would have been *\*\*-kwa* (the sound change *\*wa* > *\*wâ* normally takes place only in closed syllables).

The Proto-Mundurukuan reconstruction, taken from Picanço (2019: 137), is based on Mundurukú *-je-kot* and Kuruaya *-de-kot* ‘to dig’. This verb includes a middle voice prefix, Mundurukú *je-* / Kuruaya *de-* (Gomes 2007). The root lacks known cognates in other Tupian

<sup>21</sup> Each Macro-Jê branch employs its own root for this meaning: Krenak *kiiñat* ~ *hiññat* ~ *ñiñat*; Maxakalí *xūñim*; PSJ *\*k(r)ŋŋθej*; Proto-Akuwê *\*arobo*; Karajá *tyrèhè*; Ofayé *ɸoktae?* ~ *ɸektaj?* (underlying /ɸəŋtan<sup>o</sup>/ or the like); Rikbaktsa *byrizuk*; Arikapú *arokäi*; Djeoromitxi *beretxe*; Proto-Chiquitano *\*šyβijucy-* (~ *\*ši-* ~ *\*šu-*).

languages, however, there are no stronger candidates for the Proto-Tupian verb for ‘to dig’<sup>22</sup>. A semantically close verb *\*-kəC* probably rather meant ‘to plant’, as evidenced by its reflexes in Mondé, Juruna (also ‘to bury’), or Sateré-Mawé (Nikulin & Carvalho 2022: 30); only the Awetí–Guarani languages show the meaning ‘to dig’, and even then usually in compounds, such as Apyãwa *-’ywy-kaj* (with a historically incorporated root *ywy* ‘earth’). Therefore, it is quite possible that the Proto-Mundurukuan root *\*-kot* is an archaism. The respective Proto-Tupian form could have been *\*-ko(:)T*, *\*-k̃o(:)T*, or *\*-ŋgo(:)T*. A possible semantically shifted cognate, *kohr{a}* or *kor-kor* ‘to paddle’, is seen in Paiter. If the Proto-Tupian reconstruction is shown to be *\*-k̃oT*, Wayoró *-pi-ot* could be claimed to be a partial cognate (but see fn. 22).

**‘to enter’:** PJ *\*ŋgê<sub>2</sub>* (plural only) : PT *\*-ke* ~ *\*-k̃e* (Eastern)

The Proto-Jê reconstruction is from Nikulin 2020: 443. It is reflected as Khîsêtjê *angrê* (nonfinite *-ngrêt*), Xavante *ãdza* (nonfinite *-dzatsi*, dual only), Kaingang *ge*, among other reflexes. In fact, the Cerrado languages show reflexes of three morphologically related verbs: *\*a:ŋgja* (nonfinite *\*-ŋgjac*) ‘to enter.PL’, *\*-ŋgja* (nonfinite *\*-ŋgjañ*) ‘to insert.PL’, and *\*-janŋja* (nonfinite *\*-janŋjañ*) ‘to wear.PL’. The correspondences are regular, except that the finite forms in Parkatêjê (*akjêj*) and Canela–Krahô (*acjêj*) have been remodeled based on the nonfinite form; the expected finite forms in these languages would be *\*akjê/\*acjê* or *\*akia/\*aquij*. Khîsêtjê *ngr* /ŋr/ [ŋgɹ] is possibly the regular reflex of PNJ *\*/ŋj/*, though no supporting examples are known. If PJ *\*ŋgê<sub>2</sub>* is of Macro-Jê origin, the respective protoform can be reconstructed as *\*ŋgi<sub>2</sub>(C)*. No alternative candidate for ‘to enter.PL’ can be reconstructed. Its singular counterpart is reconstructed as Proto-Macro-Jê *\*jâp* (Nikulin 2020: 400), based on reflexes in the Cerrado languages (*\*a:jâ*, nonfinite *\*-jâpr*) and Ofayé *-xâh*. However, it is equally possible that the Ofayé verb is cognate with Karajá *-lò* ‘to enter’, Djeoromitxí *hu/-ru* ‘to enter’, and possibly Arikapú *-txu{rü}* ‘to enter’ (with the unexplained element *-rü*). In this case, one should reconstruct PMJ *\*jy(C)* ‘to enter.SG’. Rikbaktsa *-tsuk* ‘to enter.SG’ shows no regular correspondence to the aforementioned forms, despite being superficially similar.

In Tupian, the reflexes of *\*-ke* ~ *\*-k̃e* ‘to enter’ are only seen in the Eastern branch and include Mundurukú *-je-xé* ‘to come home’ (with a middle voice prefix), Sateré-Mawé *-(w)e-ke* ‘to enter’ (with a reflexive prefix), Apyãwa *-ke* ‘to enter’, among other reflexes. In some TG languages, the verb shows a prefixal alternation between the finite and nonfinite stems. For example, Old Tupí has the finite stem *-ike*, whereas in the nonfinite paradigm the class II stem *teîke* (*-reîke*, 3 *s-eîke*) is found (Barbosa 1956: 307). Similar alternations affect several other *\*i-/\*e-*initial verbs; I assume that the alternation in question originated as an absolute/relational alternation, also found in pairs such as PT *\*ĩrĩ* ‘hammock (absolute)’ and *\*-j-ẽrĩ* ‘hammock (relational)’. If this turns out to be an archaism, one can reconstruct PT *\*-ike* ~ *\*-ik̃e* (finite, absolute) and *\*-j-eke* ~ *\*-j-ek̃e* (non-finite, relational) ‘to enter, to come home’, with the loss of the initial vowel in languages such as Apyãwa. This verb was in any case distinct from PT *\*-wũp* ‘to enter’, with reflexes in the Tuparikém (Makurap *-mum/-mu-ã*, Wayoró *-ngũ-ã*, Karitiana *-mem*) and Eastern (Mundurukú *-õm* ‘to enter’) branches<sup>23</sup>.

<sup>22</sup> Each Tupian branch employs its own root for this meaning: Tuparí *-ay*, Wayoró *-pi-ot*, Akuntsu *-poro-ka*, Makurap *-kix*, Karitiana *-yt*, Sateré-Mawé *-pan*, Awetí *-koy* (from Proto-Tupian *\*-kəC* ‘to plant’), PTG *\*-ʔiβi-koC* (*\*iβi* ‘earth’ is historically an incorporated object, and *\*-koC* goes back to *\*-kəC* ‘to plant’). The element *-ot* in Wayoró *-pi-ot* could be cognate with Karitiana *-yt*, pointing to Proto-Tuparikém *\*-oT*, but Nogueira (2019: 175) analyzes the Wayoró verb as ‘to go inside’, where *-ot* stands for ‘to go’.

<sup>23</sup> The Mundurukú reflex *õ* of PT *\*wũ* is not known to be regular, but a similar sound correspondence is seen in Mundurukú *ó-ʔa* (< PT *\*wi* ‘ax’).

In both language families, the finite verb appears to have been originally absolute (and fossilized voice prefixes are seen in the Cerrado languages and in Tupian), and its nonfinite counterpart is reconstructed as a relational stem (class I in Jê, class II in Tupian).

**‘to pierce’**: PCerr \**-pôk* (SG), \**-japôk* (PL) : PTG \**-puk*

The Proto-Cerrado reconstruction is from Nikulin (2020: 446), who also considers the possibility that the singular stem had a distinct finite form \**-pôr* (preserved in Akuwẽ only), but note that the alternation \**-r* (finite) / \**-k* (nonfinite) is otherwise normally found in intransitive verbs only. Reflexes include Canela–Krahô *-{jõ}pôc* ‘to gut’, *-japôc* ‘to pierce.PL’ and Xavante *puru* (finite), *-pu’u* // *-pu* (nonfinite) ‘to pierce.SG, to spill’, *-dzapu’u* // *-dzapu* ‘to pierce.PL’. No cognates in other Macro-Jê languages are known, but no stronger candidates for the Proto-Macro-Jê term for ‘to pierce’ are known either. If this root does go back to Proto-Macro-Jê, its original PMJ form can be reconstructed as \**-py<sub>1</sub>k<sup>o</sup>* ~ \**(- )py<sub>1</sub>ŋ<sup>o</sup>*.

On the Tupian side of the comparison, one finds Kawaiwete *-fuk* ‘to be pierced’, Apyãwa *-pok* ‘to bleed’, Old Tupí *-puk* ‘to have a hole, to break (intr.)’, pointing to PTG \**-puk* ‘to be pierced’. If this root does go back to Proto-Tupian, its original PMJ form can be reconstructed as \**-po(:)k* or \**-mbo(:)k*. Phonetically similar verbs in other Tupian languages, such as Sateré-Mawé *-puk* ‘to swell’ or Makurap *-pok* ‘to beat, to kill’, are too semantically distant from the TG verb, and are not considered to be cognate.

The Proto-Cerrado verb is reconstructed as transitive (class I), and the Proto-Tupian as intransitive (class I).

**‘son’**: Proto-Chiquitano \**’-tsay*

: Proto-Tuparian \**-ja* ~~*ʔ*~~*P* or Proto-Mawé–Guaranian \**-ca* ~~*ʔ*~~*T*

On the Macro-Jê side, one finds Proto-Chiquitano \**’-tsay* ‘son’, where \**ts* is a thematic consonant: compare 1SG \**’i-tsay*, 1+2 \**’ú-tsay*, with the thematic consonant, and 2SG \**ʔ-áy*, 3SG \**’áŷ-šy* without it. Reflexes are seen in all Chiquitano varieties, such as Bésiro *’-sai*. No cognates in other Macro-Jê languages have been found, but it could in principle go back to PMJ \**-jay(C)* or a similar protoform.

There are two similar forms in the Tupian languages. Proto-Tuparian \**-ja* ~~*ʔ*~~*P* ‘son, fraternal nephew (male ego)’ has reflexes in all Tuparian languages: Makurap *-xaup* (also ‘sperm’), Wayoró *-ndaup*, Tuparí *-ha’úp*, Sakurabiat *-taup*, Akuntsú *-taiP* (Nogueira et al. 2019: 43; Nikulin & Andrade 2020: 295). It lacks known cognates in other Tupian languages, but could in principle go back to PT \**-ja* ~~*ʔ*~~*uP* or \**-ja* ~~*ʔ*~~*kuP*; the former could be related to Proto-Chiquitano \**’-tsay*. In the Mawé–Guaranian languages, one finds reflexes of Proto-Mawé–Guaranian \**-ca* ~~*ʔ*~~*T* ‘son / fraternal nephew (male ego)’, reflected as Sateré-Mawé *-sa’yr{u}*, Awetí *ta’yt/-a’yt*, PTG \**ta* ~~*ʔ*~~*T* / \**-ra* ~~*ʔ*~~*T* (Carvalho & Birchall 2022: 27). Unless it is related to Proto-Tupian \**-kuT* ‘child’, with reflexes in Tuparikém and Mundurukuran, it has no known cognates elsewhere in Tupian. Its possible original PT form could be \**-ja* ~~*ʔ*~~*T*, \**-ja* ~~*ʔ*~~*uT*, \**-ja* ~~*ʔ*~~*kuT*, or \**-ja* ~~*ʔ*~~*kuT*; the former two could be technically related to Proto-Chiquitano \**’-tsay*.

All the aforementioned forms are class II relational stems. In Tupi–Guaranian, \**ta* ~~*ʔ*~~*P* / \**-ra* ~~*ʔ*~~*P* belongs to the so-called subclass IIb, which includes a handful of kinship terms; it differs from other class II subtypes in having a third-person form identical to the absolute one (\**ta* ~~*ʔ*~~*P* ‘his son/fraternal nephew’).

**‘sour’**: PJ \**-jô<sup>k</sup>* ‘sour, salty’ : Karitiana *-syk*

The Proto-Jê reconstruction is from Nikulin 2020: 437. Reflexes include Khîsêtjê *-twa* ‘sour’, Canela–Krahô *-xwa* ‘sour, salty’, Pykobjê–Křikatí *-xwa* ‘salty’, Kaingang *-{ka}jã* ‘salty, sour’. PNJ \**ka:jwa* ‘salt’ is likely related. No cognates in other Macro-Jê languages are known, but no



stronger candidates for the Proto-Macro-Jê term for ‘sour’ are known either (see the discussion under ‘bitter’ in 4.3). If this root does go back to Proto-Macro-Jê, its original PMJ form can be reconstructed as *\*-juk* (the stem-final velar stop can be recovered based on the Kaingang verbal derivative *-kajã-g* ‘to become sour’).

On the Tupian side of the comparison, one finds Karitiana *-syk* ‘sour, to become sour’, attested in Rocha 2011: 218. A possible cognate is Karo *-{xa}’yōk* ‘sour; to be drunk’, prompting the reconstruction PT *\*-joK*. The element *xa’-* in *-xa’yōk* is plausibly a fossilized prefix with an unclear meaning, also found in *xa’kĩn* ‘monkey (sp.)’, *xa’wūt* ‘thorn’, *xa’wap* ‘sun’, *a’-xa’pe* ‘bark’ (compare PT *\*-pe* ‘bark, skin’); the nasal vowel *ō* is unexpected, but parallels do exist (Karo *-yakōp* ‘hot’ < PT *\*-jakop*). Alternatively, the Karo form can also be compared to the final syllable of Sateré-Mawé *-jējuḡ* ‘sour’ (only the third-person form *h-ējuḡ* is in fact attested in Ribeiro 2010: 58); this would account for the nasal vowel in Karo, but not for the stem-final *k* (Karo *\*...yōḡ* would be expected). Karitiana *-syk* is also similar to Proto-Mundurukuan *\*-sak* ‘to be sour’ (Picanço 2019: 138), but there are no regular correspondences between these forms, and the similarity must be accidental. The hypothetical Mundurukuan cognate of Karitiana *-syk* would be *\*\*-ðak*; note that Proto-Mundurukuan *\*s* normally results from contraction, as in *\*māsik* < PT *\*mānĩ-ʔak* (possibly through the stages *\*māndjik* < *\*māndiʔik*).

Be it as it may, Karitiana *-syk* and possibly Karo *-xa’yōk* can technically go back to PT *\*-joK*. The meaning ‘sour’ could probably be alternatively expressed by PT *\*-ati* or *\*-jati* ‘pain, to hurt’ (Nikulin & Carvalho 2022: 29), as is still the case in Tuparian (Makurap *-xati*, Wayoró *-ati*, Tuparí *-asi*); compare Yudja *-xadi* and Xipaya *-xadi* ‘to become sour’, and possibly Aruá *⟨tatit⟩*. If the root *\*-joK* coexisted with it, its semantics must have been more restricted.

#### **‘sweet’**: PMJ *\*-jāñ* (Eastern) : Tuparí *-hoy*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 400. Reflexes are found in the Eastern branch only and include Maxakalí *-xux-pek*, Khîsêtjê *-tán // -táni*, Xavante *-dzei // -dze*. The correspondences are regular, except that the Canela–Krahô and Pykobjê–Krîkatí reflex *-xen* shows an irregular fronted reflex of the vowel *\*ə* (the third-person forms Canela–Krahô *h-àn* and Pykobjê–Krîkatí *h-ỳn* are, however, regular).

Tuparí *-hoy* ‘sweet’ is attested in Alves 2004: 179. No cognates are known elsewhere in Tupian, but if this form is shown to go back all the way to Proto-Tupian, the respective proto-form should be reconstructed as *\*-joC*. Each Tuparikém language shows a different root for ‘sweet’ (Wayoró *-tui*, Akuntsú *-kon*, Makurap *⟨čan⟩*, Karitiana *-kowot /-kowitz/*), and other branches of Tupian also show noncognate forms: Karo *-pewit*, Aruá *⟨čiim⟩*, Proto-Juruna *\*-etšākū* (> Yudja *-etxākū*, Xipaya *-etākū*), Mundurukú *-kurúkurú*, Proto-Mawé–Guaranian *\*-cēʔē* (> Sateré-Mawé *-jēʔē*, Awetí *-eʔē*, Apyãwa *’eʔē*, Old Tupí 3 *s-eʔē* ‘sweet, salty’). Therefore, even though the chances of Tuparí *-hoy* to be an archaism are rather slim, there are no stronger candidates for the Proto-Tupian term for ‘sweet’ anyway.

The Proto-Macro-Jê stem is a class II relational stem, as seen in Khîsêtjê 3 *s-án // s-áni*, Canela–Krahô *h-àn*, Pykobjê–Krîkatí *h-ỳn*. Tuparí has lost the class I/class II distinction, but *h*-initial relational stems in that language typically go back to Proto-Tuparian (and Proto-Tupian) class II stems.

#### 4.5. Noncognate lookalikes or loans

##### **‘flat’**: Proto-Mawé–Guaranian *\*-peP* and Ofayé *-phiʔ*

The Proto-Mawé–Guaranian reconstruction is from Meira & Drude 2015: 293. Reflexes include Sateré-Mawé *-pēp*, Awetí *-pep* (in *mōj-pep* ‘flat snake’, *tatu-pep* ‘armadillo’, *ywy-pep*

‘ground’), Apyãwa *-pew-a*, all meaning ‘flat’. No cognates elsewhere in Tupian are known.<sup>24</sup>

The Ofayé term for ‘flat’ is very scarcely attested. Gudschinsky documents it only in compounds (*krej-phi?* ‘blind = eye-flat’; *kãñõ:r-phi?* ‘cockroach = ?-flat’) and states that the bare root did not occur in her corpus, though her consultant did recognize the element *-phi?* as a term for ‘flat’. Although Ofayé *-phi?* is technically comparable to Proto-Mawé-Guaranian *\*-pe:P*, I propose that it is more likely to be related to Maxakalí *-pex /-pek/* ‘flat’, attested in the compound *kot-pex* ‘beiju = manioc-flat’ (Silva 2020: 260). The Proto-Macro-Jê etymon can then be reconstructed as *\*-pek(°)*, *\*-pêk(°)*, *\*-mbek(°)*, or *\*-mbêk(°)*, with a velar coda that is not compatible with the labial coda seen in Tupian.

**‘to kill’:** Ofayé *-kãj?*, Proto-Chiquitano *\*kõ õj-* ‘to kill, to die’, and Awetí *-kỹj*

If the Ofayé and Chiquitano forms are indeed cognate, the Proto-Macro-Jê form can be tentatively reconstructed as *\*-ŋãj*. Since PMJ *\*k* yields Ofayé *h* (Nikulin 2020: 108), one is forced to reconstruct the initial consonant as *\*ŋ*, whose development in Ofayé has hitherto remained unknown. However, this proposal is not compatible with the idea that Ofayé *hãũjẽ?* ‘earth’ is cognate with Kaingang *ga* ‘earth’, as hypothesized in section 4.1. The vowel can be reconstructed as *\*ã* (the only PMJ nasal vowel whose Ofayé reflex is unknown), and the coda can be reconstructed as palatal based on Ofayé *j* and the Chiquitano third-person finite forms such as Bésiro *kóĩñ-o* or Migueleño *kóoñ-o*. Note that Chiquitano shows pervasive patientive lability, and the verb is used both for describing spontaneous death (with an absolutive subject) and unnatural death (with an absolutive patient and ergative agent). A possible cognate in the Eastern branch Maxakalí is *-kux* ‘to finish’ (Silva 2020: 275).

Awetí *-kỹj* ‘to kill’ (Sabino 2016: 56) is superficially similar to the aforementioned data. This root is isolated within Tupian: the meaning ‘to kill’ is rather expressed by reflexes of PT *\*-ʔaoka ~ \*-ʔaokã* ‘to kill, to beat’ in other languages of the Eastern branch. It cannot be a loan from Ofayé or Chiquitano, because the Awetí live far away from the Ofayé (1,000 km to the north) and the Chiquitano (800 km to the northeast), and there is no reason to suspect these peoples have ever been in contact.

**‘liquid’:** PT *\*ʔuu / \*-j-uu* and Proto-Jabutian *\*-y*

The Proto-Tupian reconstructions are from Nikulin & Carvalho (2022: 30, 37). PT *\*ʔuu* was an absolute noun (the basic term for ‘water’), whereas *\*-j-uu* was a relational class II noun, used in compounds that denoted liquids. Reflexes are found in all branches, including Kepkiriwat

<sup>24</sup> Gerardi et al. (2022, concept FLAT) propose a number of competing cognate sets involving terms for ‘flat’, but none withstands scrutiny. Their cognate set 6281 includes Mundurukú *-sẽm* ‘smooth’ (Crofts 1973) and reflexes of an unrelated Proto-Mawé-Guaranian etymon *\*-tĩP* ‘smooth’ (whence Sateré-Mawé *-tym* ‘smooth’; Ribeiro 2010: 63). Gerardi et al.’s (2022) cognate set 6282 includes Karo *xẽrat* ‘smooth’ (Gabas Jr 1999: 22) and Kuruaya *-korop*, a term obviously noncognate with Karo *xẽrat* and cognate with Mundurukú *-kóróp* ‘smooth’ (Crofts 1973). Their cognate set 6283 lists three cognate terms restricted to the Mondé branch (Gavião *finínĩp*, Mondé *sinĩp*, Suruí-Paiter *firip*) as well as an obviously noncognate Karitiana term *-kỹkyn* ‘smooth’ (Landin 2005: 15). Finally, Gerardi et al.’s (2022) cognate set 6284 lists a form cited as Kamayurá *ojim* — which is obviously a mistranscription of a root whose third-person form is attested in Seki (2000: 413) as *i-jym* ‘it is smooth’ (ultimately a reflex of Proto-Mawé-Guaranian *\*-tĩP* ‘smooth’) — and Paraguayan Guaraní *-joja*, which is ultimately derived from the verb *-ja* ‘to stick’ (< Proto-Tupí-Guaranian *\*-jat*) by means of a reciprocal prefix. Needless to say, the Kamayurá and Paraguayan Guaraní forms given in Gerardi et al. (2022) cannot be cognate. This example is illustrative of the general careless approach to etymology, morphological segmentation, and semantics in Gerardi et al.’s (2022) database, which unfortunately cannot be used as a reference source for comparative Tupian studies.

(<i-ü> ‘water’), Tuparikém (Makurap *u* ‘water’, Wayoró *u-gu* ‘water’, Karitiana *e* ‘rain’, *e-se* ‘water’, *-se* ‘liquid’), Rama-Puru (Karo *i-xû* ‘water’, Puruborá *fi* ‘liquid, chicha’), Mondé (Paitei *ih* ‘water’, Aruá <endjatji> ‘tear’, <namdji> ‘milk’), and Eastern (Mundurukú *i-dî-bî* ‘water’, *-di* ‘liquid’, Apyãwa *’y-θ* ‘water’); see Galucio et al. 2015: 258 for a selection of reflexes.

Proto-Jabutian *\*-y* ‘liquid’ (Voort 2007: 159) is reflected as Arikapú *-ü*, Djeoromitxí *-i*. Chiquitano, Rikbaktsa, and Maxakalí have remotely similar forms: Bésiro *n-θ-iyí-xí* ‘its juice, broth’, Rikbaktsa *tsik* ‘chicha’, *tsik-/tsik* ‘liquid’ (cf. also *-hik* in *pi-hik* ‘water’), Maxakalí *-hep* ‘liquid, blood’, but these are hardly cognate with Proto-Jabutian *\*-y*. The Proto-Jabutian vowel *\*y* has no known Macro-Jê origin, suggesting that *\*-y* is a likely loan from Tupian. It is however unclear why and how the absolute stem *\*?u* (or its reflex in some specific branch of Tupian) could have been borrowed as a relational stem in Jabutian.

**‘louse’**: PMJ *\*-ngy<sub>1</sub>n°* (Eastern only) and Proto-Core Mondé *\*git*

The Proto-Macro-Jê reconstruction is based on reflexes found in the Eastern branch only: Maxakalí *-kut*, Khîsêjtjê *-ngô*, Xavante *-’u*, Kaingang *-ga* ‘louse, maggot’. It is erroneously given as *\*ngy<sub>1</sub>t* in Nikulin 2020: 419, but the Pykobjê–Křikatí cognate *-ncuu* with its long vowel shows that the protoform must be reconstructed with a nasal coda followed by an echo vowel. The expected reflex of *\*-ngyn°* in Xavante would be *\*-’õno* in the utterance-medial position and *-’u* in the utterance-final position, but the former is not attested in my sources on Xavante; I assume it was ousted by the utterance-final allomorph.

On the Tupian side of the comparison, one finds Gavião, Aruá, and Zoró *git* (the Cinta Larga term for ‘louse’ is not attested in the sources I am aware of). The stem-final consonant matches the Macro-Jê forms, in stark contrast with all other Mondé and, more broadly, Tupian languages, which uniformly show reflexes of Proto-Tupian *\*(-)ngup* (Nikulin & Carvalho 2022: 33): compare Paitei *gib*, Salamã *gip*, Makurap *gup*, Wayoró *-a-ngup*, Karitiana *ngep*, Puruborá *típ*, Yudja *kípá*, Mundurukú *kíp*, Sateré-Mawé *gyp*, Awetí *-’a-kyp*, Apyãwa *-kyw-a*; see Galucio et al. 2015: 252 for a selection of reflexes. I have no explanation regarding the outstanding similarity between Proto-Core Mondé *\*git* and the Macro-Jê forms. It is of course possible, but also undemonstrable, that an extinct branch of Macro-Jê that preserved the place of articulation of the PMJ codas was present in the Mondé-speaking area during the period of the independent evolution of the Core Mondé languages (i.e., after the split-off of Salamã, but before the differentiation of Proto-Core Mondé into dialects), and Proto-Core Mondé could have borrowed the noun *\*git* from the speculative Macro-Jê language. In any case, it cannot be cognate with PMJ *\*-ngy<sub>1</sub>n°*, since the basic term for ‘louse’ both in Proto-Tupian and Proto-Mondé clearly had a labial coda.

The Macro-Jê noun is a relational class I stem. In Tupian, both relational (Wayoró, Awetí, Apyãwa) and absolute (Makurap, Karitiana, Yudja, Mundurukú, Sateré-Mawé) reflexes are attested, suggesting that the Proto-Tupian root was relationally labile.

**‘neck’**: PT *\*-wot* and Proto-Cerrado *\*-mbut*

The Proto-Tupian reconstruction is based on reflexes in the Tuparikém (Makurap *-wot-kup*, Wayoró *-ngot-kup*, Karitiana *-hyt*), Rama-Puru (Karo *-ot ká’*), and Eastern (Sateré-Mawé *-hut-’yp*, Awetí *-tur-’yp*, Apyãwa *-xor-a*) branches; see Galucio et al. 2015: 255 for a selection of reflexes. The correspondences are regular, except for the reflexes in the Siokweriat dialect of Sakurabiat (*-kut-kup* instead of the expected *\*-kot-kup*) and Akuntsu (*-pít-kíp* instead of the expected *\*-kot-kíp*). The reflexes in the Tupian languages, in Arikém (but not in Karitiana), Sateré-Mawé, and Awetí point to the compound *\*-wot-kup*.

The Proto-Cerrado reconstruction is from Nikulin 2020: 449. Reflexes include Khîsêjtjê *-mbut* // *-mburu*, Panará *imputi* ‘nape’, and Xavante *-butu* // *-budu*, and the correspondences are

fully regular. It could technically go back to PJ *\*-mbut*<sup>o</sup> < PMJ *\*-mbot*<sup>o</sup>, but it is unlikely that the root in question is old, given that there are two stronger candidates for the PMJ terms for ‘neck’. PMJ *\*-ndo<sub>1</sub>nĩ* ‘neck’ (Nikulin 2020: 388) is preserved in Chiquitano (Bésiro *-ti*, Migueleño *-tii*), Ofayé (*-tôžž*, underlying /-tôn<sup>o</sup>/ ‘nape’), and Eastern (Kaingang *-nunh*). PMJ *\*-jô(C)-cet* ~ *\*-jô(C)-cek* ~ *\*-jy(C)-cet* ~ *\*-jy(C)-cek* (Nikulin 2020: 401) is reflected as Karajá *-lòti* and Rikbaktsa *-soik*. Therefore, it is quite improbable that *\*-mbot*<sup>o</sup> was the basic term for ‘neck’ in Proto-Macro-Jê.

All the aforementioned forms (except Karajá *-lòti*) are class II relational stems.

**‘powder, paste’:** PT *\*-jōžžP* and Proto-Jabutian *\*-nũ*

The Proto-Tupian reconstruction is based on reflexes in the Tuparikém and Eastern branches, including Wayoró *-yōom* ‘powder’ < Proto-Tuparian *\*-jōžžP* (Nikulin & Andrade 2020: 296), Yudja *-umá* < Proto-Juruna *\*-um-á*, Mundurukú *-nôm* < Proto-Mundurukuan *\*-đōm* (Picanço 2019: 140), Awetí *ywy-lu’um* ‘dirt’, *-enta-lu’um* ‘rheum’, and Apyãwa *to’om-a/-ro’om-a* ‘paste’. The correspondences are regular, except that Awetí *l* is not a regular reflex of PT *\*j* > Proto-Mawé-Guaranian *\*c*.

Proto-Jabutian *\*-nũ* ‘pamonha, porridge, food’ is reconstructed in Voort 2007: 156, who notes the similarity of this term with classifiers for ‘pamonha, flour’ in different unrelated languages of the Guaporé area. However, this noun is a reflex of PMJ *\*-ñũ<sub>2</sub>(C)* ‘food’ (Nikulin 2020: 403), whence Eastern Chiquitano *-ō’ō*, Karajá *dòd* ‘solid food, such as fish, turtle or meat’, Khĩsètjê *-nho* ‘food’, Xavante *-nho* ‘food’.

Despite the similarity in form and the fact that both Tupian and Macro-Jê comparanda are relational class II stems, the semantic difference between the Proto-Tupian and Proto-Macro-Jê forms renders the comparison unattractive.

**‘thorn’:** PMJ *\*-ñĩn<sup>o</sup>* ~ *\*-ñĩñ<sup>o</sup>* and Tuparí *-ĩ*

The Proto-Macro-Jê reconstruction is given as *\*-ñĩ(C)* in Nikulin 2020: 406. Reflexes are found in the Western (Djeoromitxí *-nĩ* ‘leaf’, Rikbaktsa *-ni*), Karajá (*dê~dè*), and Eastern (Khĩsètjê *-khrã-nhi*, Měbêngôkre *mrÿ-nhĩ*, Apinajé *-nhĩ*, Pykobjê–Krĩkatí *hum-jêeh*, Akwẽ-Xerente *-krã-nĩ*) branches. The Pykobjê–Krĩkatí reflex with a long vowel suggests that the PMJ form ended in a nasal consonant followed by an echo vowel.

Tuparí *-ĩ* ‘thorn, grain’ (Alves 2004: 185) lacks known cognates in other Tupian languages. It could technically go back to Proto-Tuparian *\*-jĩ* and PT *\*-jĩ(C)* ~ *\*-jĩ(C)* (the loss of *\*j* before *\*ĩ* is regular in that language; see Nikulin & Andrade 2020: 296), but it is unlikely that the root in question is old, given that a different root *\*wo*: ‘thorn’ (whence Wayoró *ngoo*, Karitiana *hy*, Sateré-Mawé *hu*, Apyãwa *xo-∅*, etc.) can be reconstructed. Instead, Tuparí *-ĩ* could be an Arikapú borrowing. Mundurukú *-ĩ* ‘CL:nuts’ (Crofts 1985: 313) is probably unrelated.

## 5. Regular sound correspondences

Now that 38 Macro-Jê–Tupian possible cognate sets have been identified (4.1–4.4; the data from 4.5 are discarded), I proceed to examine the sound correspondences that recur in my comparative corpus (5.1). Non-recurrent correspondences may signal that a given comparison is spurious, and should be discarded over the next iteration. In section 5.2, I address the possibility of identifying additional sound correspondences, which violate the constraints set out in the preamble of 4 — notably the full match between the places of articulation of the onsets and codas — but could nevertheless be regular.

### 5.1. Main sound correspondences

In this section, I make an attempt at determining the sound correspondences between PMJ and PT. In reproducing the data from the preceding section, I adhere to the following principles.

Whenever the data allow for multiple diachronic interpretations, I choose the option that best matches the correspondence sets whose existence is independently established. For example, the data of the Macro-Jê languages are insufficient to determine whether PMJ *\*-we(C)* ‘to go up’ contained a coda or not. In this section, this form is rewritten as *\*-wep ~ \*-wem<sup>o</sup>*, since these are the only possibilities that can match PT *\*-we(:)P*.

For PMJ and PT hypothetical reconstructions based on evidence from a single branch (4.2–4.4), I employ the symbol # instead of the asterisk. For example, the hypothetical Proto-Tupian ancestor of Proto-Mundurukuan *\*-kāj* ‘hole’ can, in theory, be reconstructed as *\*-kājāc*, *\*-kājāc*, *\*-kājāc*, *\*-kājāc*, *\*-kājāc*, or *\*-kājāc*. Of these, *\*-kājāc* is the option that best matches PMJ *\*-kuñ<sup>o</sup>*, and it is reproduced in this section as PT #*-kājāc*.

Table 4 shows the sound correspondences between PMJ and PT onsets. PMJ *\*/c/* and *\*/ñ/* each occur only once in the corpus, hence it is unsurprising that the respective correspondences are not recurrent. In the cognate sets for ‘arm’, ‘foot’, and ‘liver’, Tupian shows an alternation between *\*/p/* in relational stems and *\*/m/* in absolute ones. Macro-Jê would appear to have generalized the relational stems for ‘arm’ and ‘foot’, and the absolute one for ‘liver’. As for the cognate set PMJ *\*-ja-m* ‘to stand (nonfinite)’: PT *\*-jāP* ‘to stand’, it may be significant that PMJ lacks relational vowel-initial stems, and makes use of the relationalizing prefix *\*j-* when a vowel-initial root enters a relational stem (see section 3). See 4.3 for a discussion on the root-medial correspondence in the cognate set for ‘smoke’.

Two non-recurrent correspondences are PMJ *\*/ŋ/* : PT *\*/k/* (‘earth’) and PMJ *\*/ŋ/* : PT *\*/k/* or *\*/k̃/* (‘to enter’). Of these, the former could be due to an erroneous inclusion of PSJ *\*ŋgə* ‘earth’ into the comparison; if it turns out to be noncognate, the PMJ term for ‘earth’ can be reconstructed as *\*kyñ<sup>o</sup>* instead (with reflexes in Chiquitano and Ofayé), thus instantiating the recurrent correspondence PMJ *\*k* : PT *\*k̃*. Alternatively, one could surmise that historically PMJ had an alternation between relational *\*/k/-*initial stems and absolute *\*/ŋ/-*initial stems (a similar alternation is reconstructible for Proto-Tupian based on evidence from Sateré-Mawé and Mondé). Note that PMJ *\*ŋgyñ<sup>o</sup>* ‘earth’ and #*ŋgi<sub>2</sub>* ‘to enter.PL’ (finite) are absolute and do not take prefixes, whereas PMJ *\*-ko<sub>2</sub>* ‘to ingest’, *\*(-)ky<sub>1</sub>m<sup>o</sup>* ‘tree(-like)’, #*-ki* ‘to do, to say’, *\*-kut* ‘to dig’, *\*-kuñ<sup>o</sup>* ‘hole’, #*-kân<sup>o</sup>* ‘white’ all take absolutive or accusative indices (*\*(-)ky<sub>1</sub>m<sup>o</sup>* is relationally labile).

Table 5 shows the sound correspondences between PMJ and PT vowels (vowel nasality is ignored at this stage). PMJ *\*/â/* and *\*/ê/* occurred each only once in the corpus, hence it is unsurprising that the respective correspondences are not recurrent. The cognate set for ‘feces’ presents insurmountable difficulties regarding the reconstruction of its nasal vowel in both protolanguages: the correspondences are unique in both Macro-Jê and Tupian. In the cognate set for ‘smoke’, Tupian could have contracted a disyllabic sequence into a long vowel, as suggested in 4.3. The vowel correspondence in the term for ‘bat’ appears to be truly irregular; recall, however, that the reconstruction of PT *\*u* in *\*juP* hinges on one’s interpretation of Barbosa de Faria’s attestation of ⟨ê⟩ in Kepkiriwat as an instance of /i/. If ⟨jêp⟩ is a representation of /jiP/ rather than /jîP/, the cognate set for ‘bat’ instantiates the recurrent correspondence PMJ *\*i* : PT *\*i*. The vowel correspondence in the term for ‘to pierce’ is unique; combined with the discrepancy in the transitivity of the PMJ and PT verbs (transitive and intransitive, respectively), this is a sufficient reason to discard the etymology.

Although oral and nasal vowels are not distinguished in the correspondences in Table 5, there is a systematic tendency for PMJ oral vowels to correspond to PT oral vowels (28 examples), whereas PMJ nasal vowels correspond to PT nasal vowels (‘to go/come’, ‘to kill’, ‘smoke’,

| PMJ                                | PT            | examples   |
|------------------------------------|---------------|--|
| */p/                               | */p/          | <p><b>‘arm’</b>: PMJ <i>*-pa</i> ‘arm’ : PT <i>*-pə</i> / <i>*mbə</i> ‘hand’</p> <p><b>‘to burn’</b>: PMJ <i>*(-)py<sub>1</sub>k<sup>o</sup></i> ~ <i>*(-)py<sub>1</sub>ŋ<sup>o</sup></i> : PT <i>*-puuk</i></p> <p><b>‘foot’</b>: PMJ <i>*-pâr<sup>o</sup></i> : PT <i>*-pi</i> / <i>*mbi</i></p> <p><b>‘heavy’</b>: PMJ <i>#-pVtVJ(°)</i> : PT <i>*-pətic</i></p> <p><b>‘to pierce’</b>: PMJ <i>#-py<sub>1</sub>k<sup>o</sup></i> ~ <i>#-py<sub>1</sub>ŋ<sup>o</sup></i> : PT <i>#-poK</i></p> <p><b>‘to wake up’</b>: PMJ <i>#-paK(°)</i> : PT <i>*-paK</i></p>   |
| */m/ (*[mb], *[m])                 | */m/          | <p><b>‘husband’</b>: PMJ <i>*-mbi<sub>2</sub>n</i> : PT <i>*-mēT</i></p> <p><b>‘liver’</b>: PMJ <i>*-mbâ</i> : PT <i>*-pi(-)ʔa</i> / <i>*mbi(-)ʔa</i></p>  |
| */w/                               | */w/          | <p><b>‘to arrive’</b>: PMJ <i>#(-)wy<sub>1</sub>c<sup>o</sup></i> : PT <i>*-wuuc</i></p> <p><b>‘to go up’</b>: PMJ <i>*-wep</i> ~ <i>*-wem<sup>o</sup></i> : PT <i>*-we(:)P</i></p> <p><b>‘to kill’</b>: PMJ <i>*-wī</i> : PT <i>#-wī</i></p>  |
| */t/                               | */t/          | <p><b>‘heavy’</b>: PMJ <i>#-pVtVJ(°)</i> : PT <i>*-pətic</i></p> <p><b>3NCRF prefix</b>: PMJ <i>*ta-</i> : PT <i>*tə-</i></p> <p><b>‘to go/come’</b>: PMJ <i>*tē</i> / <i>*-tē-m</i> : PT <i>*-tēP</i></p>   |
| */n/ (*[nd], no examples for *[n]) | */ð/          | <p><b>‘bitter’</b>: PMJ <i>#-ndap<sup>o</sup></i> : PT <i>*-ðəP</i></p> <p><b>‘ripe’</b>: PMJ <i>*-ndêp<sup>o</sup></i> : PT <i>#-ðeP</i></p>  |
| */c/                               | */c/          | <b>3NCRF</b> : PMJ <i>*c-</i> : PT <i>*c-</i>  |
| */ñ/                               |               | <b>‘bat’</b> : PMJ <i>#ñjip<sup>o</sup></i> : PT <i>*juP</i>   |
| */j/ (*[j], *[ɲ])                  | */j/          | <p><b>‘father’</b>: PMJ <i>*-jo<sub>2</sub>m</i> : PT <i>*-joP</i></p> <p><b>‘feces’</b>: PMJ <i>*-ñVt<sup>o</sup></i> : PT <i>*-jVt</i></p> <p><b>‘meat (rel.)’</b>: PMJ <i>*-ñit</i> : PT <i>*-jēT</i></p> <p><b>‘name’</b>: PMJ <i>*-jet</i> : PT <i>*-jeT</i></p> <p><b>‘pus’</b>: PMJ <i>*-jo<sub>2</sub>w<sup>o</sup></i> : PT <i>*-joP</i></p> <p><b>‘sweet’</b>: PMJ <i>*-jāñ</i> : PT <i>#-joC</i></p> <p><b>‘smoke’</b>: PMJ <i>*-ñijək</i> : PT <i>*-jī:K</i></p> <p><b>‘son’</b>: PMJ <i>#-jayC</i> : PT <i>#-jaʔuP</i> or <i>#-jaʔuT</i></p> <p><b>‘sour’</b>: PMJ <i>#-juk</i> : PT <i>#-joK</i></p> <p><b>‘to stand’</b>: PMJ <i>*ja</i> : PT <i>*-ja</i></p> <p><b>‘tooth’</b>: PMJ <i>*-juñ<sup>o</sup></i> : PT <i>*-jāC</i></p> |
| */k/                               | */k/          | <p><b>‘to ingest’</b>: PMJ <i>*-ko<sub>2</sub></i> : PT <i>*-ko</i></p> <p><b>‘tree(-like)’</b>: PMJ <i>*(-)ky<sub>1</sub>m<sup>o</sup></i> : PT <i>*(-)kuP</i></p> <p><b>‘to do, to say’</b>: PMJ <i>#-ki</i> : PT <i>*-ke</i></p>  |
|                                    | */k̃/         | <p><b>‘to dig’</b>: PMJ <i>*-kut</i> : PT <i>#-k̃oT</i></p> <p><b>‘hole’</b>: PMJ <i>*-kuñ<sup>o</sup></i> : PT <i>#-k̃āʔāC</i></p> <p><b>‘white’</b>: PMJ <i>#-kân<sup>o</sup></i> : PT <i>*-k̃iT</i></p>   |
| */ŋ/ (*[ŋg], no examples for *[ŋ]) | */k/          | <b>‘earth’</b> : PMJ <i>*ŋgyñ<sup>o</sup></i> : PT <i>*kuuc</i>  |
|                                    | */k/ or */k̃/ | <b>‘to enter’</b> : PMJ <i>#ŋgi<sub>2</sub></i> : PT <i>*-ke</i> ~ <i>*-k̃e</i>  |
| *∅                                 | *∅            | <p><b>‘to give’</b>: PMJ <i>*-ūp</i> : PT <i>*-ōP</i></p> <p><b>‘I’</b>: PMJ <i>#u</i> : PT <i>*o-</i></p> <p><b>‘meat (abs.)’</b>: PMJ <i>*it</i> : PT <i>*ēT</i></p>   |
|                                    | */ʔ/          | <p><b>‘hole’</b>: PMJ <i>*-kuñ<sup>o</sup></i> : PT <i>#-k̃āʔāC</i></p> <p><b>‘son’</b>: PMJ <i>#-jayC</i> : PT <i>#-jaʔuP</i> or <i>#-jaʔuT</i></p>   |
| */j/                               |               | <b>‘to stand’</b> : PMJ <i>*-ja-m</i> : PT <i>*-ʔāP</i>  |
|                                    | *∅            | <b>‘smoke’</b> : PMJ <i>*-ñijək</i> : PT <i>*-jī:K</i>   |

Table 4. Sound correspondences between Macro-Jê and Tupian onsets

| PMJ           | PT | examples   |
|---------------|----|--|
| *a            | *a | <u>‘to stand’</u> : PMJ * <i>ja</i> / *- <i>ja-m</i> : PT *- <i>ja</i> / *- <i>ʔãP</i><br><u>‘son’</u> : PMJ #- <i>jayC</i> : PT #- <i>jaʔuP</i> or #- <i>jaʔuT</i><br><u>‘to wake up’</u> : PMJ #- <i>paK</i> (°) : PT *- <i>paK</i>  |
|               | *ə | <u>3CRF prefix</u> : PMJ * <i>ta-</i> : PT * <i>tə-</i><br><u>‘arm’</u> : PMJ *- <i>pa</i> ‘arm’ : PT *- <i>pə</i> / * <i>mbə</i> ‘hand’<br><u>‘bitter’</u> : PMJ #- <i>ndap</i> ° : PT *- <i>ðəP</i>  |
| *â            | *i | <u>‘foot’</u> : PMJ *- <i>pâr</i> ° : PT *- <i>pi</i> / * <i>mbi</i><br><u>‘liver’</u> : PMJ *- <i>mbâ</i> : PT *- <i>pi(-)ʔa</i> / * <i>mbi(-)ʔa</i><br><u>‘white’</u> : PMJ #- <i>kân</i> ° : PT *- <i>ķiT</i>   |
| *ê            | *o | <u>‘sweet’</u> : PMJ *- <i>jəñ</i> : PT #- <i>joc</i>  |
| *y            | *u | <u>‘to arrive’</u> : PMJ #(-) <i>wy<sub>1</sub>c</i> ° : PT *- <i>wu<sub>1</sub>c</i><br><u>‘to burn’</u> : PMJ *(-) <i>py<sub>1</sub>k</i> ° ~ *(-) <i>py<sub>1</sub>ŋ</i> ° : PT *- <i>pu<sub>1</sub>k</i><br><u>‘tree(-like)’</u> : PMJ *(-) <i>ky<sub>1</sub>m</i> ° : PT *(-) <i>ku<sub>1</sub>P</i><br><u>‘earth’</u> : PMJ * <i>ŋgyñ</i> ° : PT * <i>ku<sub>1</sub>c</i><br><u>‘son’</u> : PMJ #- <i>jayC</i> : PT #- <i>jaʔuP</i> or #- <i>jaʔuT</i> |
| *e            | *e | <u>‘to go up’</u> : PMJ *- <i>wep</i> ~ *- <i>wem</i> ° : PT *- <i>we(:)P</i><br><u>‘name’</u> : PMJ *- <i>jet</i> : PT *- <i>jet</i><br><u>‘to go/come’</u> : PMJ * <i>tē</i> / *- <i>tē-m</i> : PT *- <i>tēP</i>   |
| *ê            |    | <u>‘ripe’</u> : PMJ *- <i>ndêp</i> ° : PT #- <i>ðeP</i>  |
| *i            | *i | <u>3NCRF prefix</u> : PMJ * <i>i-</i> : PT * <i>i-</i><br><u>‘to kill’</u> : PMJ *- <i>wī</i> : PT #- <i>wī</i><br><u>‘smoke’</u> : PMJ *- <i>ñijǎk</i> : PT *- <i>jī:k</i>  |
|               | *e | <u>‘to do, to say’</u> : PMJ #- <i>ki</i> : PT *- <i>ke</i><br><u>‘to enter’</u> : PMJ # <i>ŋgi<sub>2</sub></i> : PT *- <i>ķe</i><br><u>‘husband’</u> : PMJ *- <i>mbi<sub>2</sub>n</i> : PT *- <i>mēT</i><br><u>‘meat’</u> : PMJ * <i>īt</i> / *- <i>ñīt</i> : PT * <i>ēT</i> / *- <i>jēT</i>  |
| *o            | *o | <u>‘father’</u> : PMJ *- <i>jo<sub>2</sub>m</i> : PT *- <i>joP</i><br><u>‘to ingest’</u> : PMJ *- <i>ko<sub>2</sub></i> : PT *- <i>ko</i><br><u>‘pus’</u> : PMJ *- <i>jo<sub>2</sub>w</i> ° : PT *- <i>joP</i>   |
| *u            | *a | <u>‘tooth’</u> : PMJ *- <i>juñ</i> ° : PT *- <i>jāC</i><br><u>‘hole’</u> : PMJ *- <i>kuñ</i> ° : PT #- <i>ķãʔãC</i>  |
|               | *o | <u>‘to dig’</u> : PMJ *- <i>kut</i> : PT #- <i>ķoT</i><br><u>‘to give’</u> : PMJ *- <i>ūp</i> : PT *- <i>ōP</i><br><u>‘I’</u> : PMJ # <i>u</i> : PT * <i>o-</i><br><u>‘sour’</u> : PMJ #- <i>juk</i> : PT #- <i>joK</i>  |
| non-recurrent |    | <u>‘bat’</u> : PMJ # <i>njip</i> ° : PT * <i>ju<sub>1</sub>P</i><br><u>‘feces’</u> : PMJ *- <i>ñVt</i> ° : PT *- <i>jVT</i><br><u>‘to pierce’</u> : PMJ #- <i>py<sub>1</sub>k</i> ° ~ #- <i>py<sub>1</sub>ŋ</i> ° : PT #- <i>poK</i><br><u>‘smoke’</u> : PMJ *- <i>ñijǎk</i> : PT *- <i>jī:k</i>   |

Table 5. Sound correspondences between Macro-Jê and Tupian vowels

‘meat’, ‘to give’, ‘feces’). An exception is constituted by four cognate sets which show a PMJ oral vowel corresponding to a PT nasal vowel (5).

- (5) PMJ PT
- |               |                      |         |
|---------------|----------------------|---------|
| a. ‘to stand’ | *-ja-m               | *-ʔãP   |
| b. ‘husband’  | *-mbi <sub>2</sub> n | *-mẽT   |
| c. ‘tooth’    | *-juñ <sup>o</sup>   | *-jãC   |
| d. ‘hole’     | *-kuñ <sup>o</sup>   | #-kãʔãC |

In all these cognate sets, the oral vowel in PMJ is followed by a nasal coda. It is tempting to assume that the respective Proto-Macro-Jê–Tupian etyma likewise contained an oral vowel followed by a nasal coda, and that the vowel became nasal in Tupian by assimilating the nasality of the erstwhile coda. However, there are also several cognate sets which feature an oral vowel followed by a nasal coda in PMJ, yet the PT cognate has an oral vowel (6).

- (6) PMJ PT
- |                  |                                    |         |
|------------------|------------------------------------|---------|
| a. ‘white’       | #-kân <sup>o</sup>                 | *-kĩT   |
| b. ‘sweet’       | *-jâñ                              | #-joC   |
| c. ‘tree(-like)’ | *(-)ky <sub>1</sub> m <sup>o</sup> | *(-)kũP |
| d. ‘earth’       | *ηγyñ <sup>o</sup>                 | *kũC    |
| e. ‘father’      | *-jo <sub>2</sub> m                | *-joP   |

Even though the data are too scarce to warrant a firm conclusion, it is noteworthy that the examples in 5 and 6 involve different vowel qualities: pre-PT *\*a* and *\*e* did undergo nasalization to PT *\*ã* and *\*ẽ* before an erstwhile nasal coda, whereas PT *\*ĩ*, *\*ũ*, and *\*o* show no signs of such a process.

In sum, 37 out of 38 candidates for cognate sets (with the exception of ‘to pierce’) show recurrent sound correspondences, or a reasonable explanation is available as for why the sound correspondences are not demonstrably recurrent.

## 5.2. Additional sound correspondences

In the preamble of section 4, I defined the criteria for the cognate search as follows: (i) all PMJ and PT consonants are required to fully match in their place of articulation, (ii) correspondences involving a back vowel in one protolanguage and a front vowel in another are disallowed. Of course, it is perfectly possible that at least some sound correspondences between PMJ and PT violate these constraints: cross-linguistically, it is very common for consonants to diachronically change their place of articulation (or to be lost altogether), and for vowels to diachronically change their backness value. Therefore, any cognates displaying such sound correspondences remained undetected in my initial cognate search. Moreover, my criteria rendered it impossible to detect any Tupian cognates for PMJ stems with complex onsets (*\*/pr/*, *\*/kr/*, *\*/mr/*, *\*/ŋr/*), because they could not be matched to anything in PT, which lacks complex onsets. This section explores the possibility of identifying cognates and sound correspondences that were overlooked in 4 and 5.1 due to the stringency of my initial criteria.

### 5.2.1. PMJ complex onsets corresponding to PT simple onsets

I start by discussing a group of possible cognates that involve a complex onset in PMJ. At least PMJ *\*/mr/* corresponds to a simple onset *\*/m/* without a rhotic in PT, as shown by the following two examples.



**‘ashes’**: PMJ \*(-)mbrôŋ : PT \*-mboK

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 381. Reflexes are found in the Western (Arikapú *pikā-brä*, Rikbaktsa *poro* ‘bamboo, salt’, Ofayé *kõtah* /ktah/) and Eastern (Maxakalí *putohok*, Krenak *proŋ* ‘charcoal’, Khîsêtjê *-mbro*, Xavante *-pro* ‘foam’ in *ö-dzai-pro* ‘foam’, *wede-pro* ‘coffee’, *-dzadai-pro* ‘saliva’, Laklãnõ *mlã*) branches. The correspondences are regular, except that Kaingang *mrêj* shows an unexpected final consonant, and Canela (and possibly other Timbira varieties) has the relational allomorph *-hpro* instead of the expected *\*-mpro*, which must be a back-formation from the regular absolute allomorph *pro*.

The Proto-Tupian term for ‘ashes’ is not readily reconstructible (Nikulin 2020:60, fn. 47). One promising candidate is PT *\*-mboK*. It has semantically shifted reflexes in Tuparikém (Tuparí *-{a}pok* ‘foam’, *-épa-pok* ‘rheum’) and a variety of morphologically complex reflexes in the Eastern branch: Xipaya *-{pu}búk-a* ‘ashes’ (<-puβúka> in Nimuendajú 2013: 205), Kawaiwete *-{’}muk* ‘powder’, Parintintin *yvy-mu-{'}* *mbug* ‘powder-like dust’. The preglottalization in Kawaiwete and Parintintin points to PTG *\*-ʔmbuK*, a form that probably results from vowel syncope and goes back to earlier *\*-ʔimbuK*. The latter form is most clearly seen in the Proto-Awetí–Guaranian compound *\*tatʰa-ʔipuK* ‘ashes’ (literally ‘fire-powder’), as reflected in Awetí *taza-’ipuk* ‘ashes’ and Ka’apor *tat-imbuk*. Quite surprisingly, TG languages other than Ka’apor do not reflect PTG *\*tāt-ʔimbuK*, but rather *\*tānīmbuK*: Kawaiwete *tanimuk*, Parintintin *tanimbug* ‘ember’, Apyãwa *tanimok-a*, Old Tupí *tanimbuk-a*, Guarasugwe *tanimi* (with an irregular final vowel), etc. Be it as it may, Proto-Awetí–Guaranian *\*-ʔipuK* appears to go back to a morphologically complex form, where the element *\*-puK* goes back to PT *\*-mboK*, and the origin of the element *\*ʔi-* is unclear. PT *\*mb* (underlying *\*/m/*) is reconstructed based on the Wayoró reflex *-{a}mbo* ‘foam’, though the velar coda is unexpectedly lacking in that form.

**‘snake’**: PT *\*mbəC* and Proto-Jabutian *\*mrāj*

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 31. Reflexes are found in all branches except Tuparikém, including Kepkiriwat (<bôi>, <boi->), Rama-Puru (Karo *māy{gāra}*, Puruborá *māŋ{ū}p*), Mondé (Zoró *baŋ*), and Eastern (Yudja *hutá*, Mundurukú *pūy-bu*, Sateré-Mawé *moi*, Apyãwa *maj-a*). The correspondences are regular, except that the Awetí reflex *mōi* has an unexpected nasal vowel.

The Proto-Jabutian term for ‘snake’ is reconstructed as *\*mrāj* in Voort 2007: 161 based on Arikapú *mrāy* and Djeoromitxí *mē*. It lacks known cognates in other Macro-Jê languages; Rikbaktsa *pyryhyk* displays some superficial similarity and could be partially cognate, though details of this etymology have not been worked out. However, no stronger candidates for the Proto-Macro-Jê root for ‘snake’ are known either. It has been proposed that Proto-Core Maxakalian *\*kājā*, Proto-Goyaz *\*kaŋā*, and Ofayé *koni* are cognate (Gudschinsky 1971: 12; Nikulin 2015: 287, 297), but these forms show no regular sound correspondences whatsoever.<sup>25</sup> If the Jabutian root is an archaism, the PMJ form can be hypothesized to have been similar to *\*mrāj*.

I have not succeeded at identifying other plausible cognate sets involving PMJ onsets of the type *\*Cr*. There is some similarity between PMJ *\*ŋrVn°* ‘toucan’ (Nikulin 2020: 420;<sup>26</sup> Eastern only) and the second syllable of PT *\*jōkāt* ‘toucan’ (the reflexes in the Arikém languages

<sup>25</sup> Other Macro-Jê branches employ clearly noncognate root(s) for this meaning: Krenak *ŋgraj*; Malalí <checheem>, <háhim>; Proto-Akuwẽ *\*wa:hi* (venomous), *\*amke* (non-venomous); PSJ *\*pān*; Proto-Karajá *\*hemālālā*; Rikbaktsa *pyryhyk*; Proto-Chiquitano *\*išoβo- ~ \*išoβu-*.

<sup>26</sup> Nikulin (2020: 420) actually reconstructs *\*ŋrṼt ~ \*ŋrṼn°*, but the former variant can be excluded based on the Timbira reflex, which has a long vowel.

point to *\*juũʔākāt*), but this comparison involves too many irregularities to be accepted. Another match that should be viewed as spurious is the comparison between Proto-Cerrado *\*prām* ‘blackfly’ (whence Canela–Krahô *prām-re*, *prām-ti*, Akwẽ-Xerente *mrām-rê*; Nikulin 2020:447) and Proto-Tuparikém *\*mērēP* ‘fly’ (whence Makurap *mērō[ā]*, Wayoró *mĩrĩm[a]*, Tuparí *mērém[ʔa]*, Karitiana *mārām*, etc.), which shows poor distribution in both language families and non-recurrent sound correspondences.

### 5.2.2. PMJ palatal coda corresponding to PT zero

There are four pairs of Macro-Jê and Tupian cognate sets with identical or similar meanings where the PMJ (PJ, PCerr) form has a palatal coda, and the Tupian comparanda lack a coda altogether. It is thus possible that some kinds of palatal codas in the hypothetical Proto-Macro-Jê–Tupian language were deleted in the phonological history of Proto-Tupian.

#### **‘urine’**: PMJ *\*-jâc* : PT *\*-jĩ(:)*

The Proto-Macro-Jê reconstruction is from Nikulin 2020: 400. Reflexes are found in Chiquitano (Bésiro *-íʔi*, not listed in Nikulin 2020) and the Eastern branch (Maxakalí *-xux* ‘to urinate’, Khĩsêtjê *-tá*, Xavante *-dzé*, Kaingang *-já-nh* ‘to urinate’). Nikulin (2020: 400) also lists a putative Karajá reflex, given as *\*-lâ*, but the correct Proto-Karajá reconstruction must be *\*-ly* ‘to urinate’, whose reflex is attested e.g. in the form *a-r-i-ly=kre* ‘I will urinate (male speech)’ (Karajá et al. 2013: 5, for some reason with a transitive prefix *i-*). Proto-Karajá *\*y* does not correspond to Maxakalí *u* /*i*/ or Proto-Cerrado *\*a*, and *\*-ly* is thus noncognate with the remaining forms. The Chiquitano reflex shows complexities as well. In addition to the well-attested relational stem *\*-ĩʔi*, there is also a similar absolute stem *\*jĩʔi-ʂi* ‘urine’, whose reflex is attested as <yĩs> in the 18<sup>th</sup> century and as *iũ-rch* in the Brazilian variety of Eastern Chiquitano (Santana 2012: 258). In my field data, *0-yĩʔi-j* (*y-iʔi-j?*) is documented as an irregular third-person singular form of *-íʔi* (thus ‘his/her/its urine’), but I concede that this may be a misanalysis on my part, and it is possible that *yĩʔi-j* could in fact be an absolute (unpossessed) term for ‘urine’ in Migueleño as well. Moreover, the 18<sup>th</sup>-century materials suggest that the relational stem for ‘urine’ takes a thematic consonant and thus has the shape *\*-cĩʔi* (whence <zĩ> *0-zĩʔi* ‘my urine’, <ozĩ> *o-zĩʔi* ‘our (INCL) urine’), which matches the data of other Macro-Jê languages but not of the contemporary Chiquitano varieties (Migueleño *ixh-ĩʔi* / *iy-ĩʔi* ‘my urine (female / male speech)’, Bésiro *n-ixh-ĩʔi* ‘my urine’).<sup>27</sup>

The Proto-Tupian reconstruction is from Nikulin & Carvalho 2022: 29. Reflexes are found at least in the Tuparikém (Wayoró *-ndi-gu*, Sakurabiat <-fũ>, Karitiana *-si*, Arikém <-sĩ>) and Eastern (Sateré-Mawé *-sy*, Apyãwa *ty-0*) branches, with possible cognates in Rama-Puru and Mondé. The sound correspondences are regular, except that the Sakurabiat and Arikém forms (both attested in pre-modern sources only) show a long vowel, which does not match the evidence from Wayoró, Karitiana, and Sateré-Mawé.

In both language families, the root in question is reconstructed as a class II relational stem, except for the aforementioned complexities in Chiquitano. The palatal coda, reconstructed for PMJ based on the Maxakalí reflex *-xux* and the Southern Jê verbalized reflex (Kaingang and Laklãnõ *já-nh* ‘to urinate’), has prevented this comparison from being cited in 4.2. The correspondence between PMJ *\*â* and PT *\*ĩ* is, however, a non-recurrent one (though the only other comparison that instantiates PMJ *\*â* is quite weak, being represented by just one language on the Tupian side). If the Proto-Macro-Jê etymon of Maxakalí *-ptux* ‘heavy’ — a possible cognate

<sup>27</sup> I thank Luca Ciucci for bringing my attention to the 18<sup>th</sup>-century forms.

of PT *\*-pətiC* — is to be reconstructed as *#-pVtəJ(°)* (as opposed to *#-pVtyJ(°)* or *#-pVtj̃J(°)*), one could argue that the correspondence between PMJ *\*ə̃* and PT *\*i* is in fact the regular one, whereas the similarity between PMJ *\*-jə̃ñ* and Tuparí *-hoy* could be spurious.

**‘another, a’:** PMJ *\*-nũc ~ \*-nũ<sub>1j</sub> ~ \*-nũ<sub>1j</sub>°* : PT *\*-nõ*

The Proto-Macro-Jê form is reconstructed as *\*-nũj* in Nikulin 2020: 390 based on reflexes in Karajá (*-nõ*), Maxakalí (*-nõy*), and Timbira (Canela–Krahô *-hnõ*, Pykobjê–Krikatí *-’no*, Parkatêjê *-nõ*). Ramirez et al. (2015: 256) identify additional cognates in other Jê languages (Apinajé *-hõ*, Mëbêngôkre *-’õ*, and Kaingang/Laklãnõ *ũ*), a suggestion rejected in Nikulin (2020) due to apparent phonological irregularities. However, it has since been established (Nikulin & Salanova 2022: 138) that the correspondence between Proto-Timbira *\*/n/*, Apinajé */ʔ/*, Mëbêngôkre */ʔ/*, and Khîsêjtê and Kajkwakhrattxi */tʰ/* is a regular one, and that it goes back to a distinct segment of the protolanguage, despite being exceedingly rare in the lexicon. Therefore, Nikulin’s (2020) criticism of Ramirez et al.’s (2015) proposal is invalid. Additional cognates are Khîsêjtê and Kajkwakhrattxi *-thõ*. I amend the PMJ reconstruction to *\*-nũ<sub>1c</sub> ~ \*-nũ<sub>1j</sub> ~ \*-nũ<sub>1j</sub>°*. The voiceless nasal *\*/ŋ/* has not been posited in earlier works on PMJ phonology. I contend that positing PMJ *\*ŋ* as an independent phoneme helps accounting for the otherwise inexplicable reflexes in Southern Jê (*\*ũ*, with no onset), in the Karajá male genderlect, and in the Javaé dialect of Karajá (*-õ*; Ribeiro 2012a: 139–141). Therefore, PMJ *\*ŋ* must have been preserved in PJ (as well as in PCerr and Proto-Goyaz) and Proto-Karajá. In PNJ, it evolved into *\*ŋ* > Khîsêjtê and Kajkwakhrattxi */tʰ/*, Apinajé and Mëbêngôkre */ʔ/*, Parkatêjê */n/*, and Canela–Krahô and Pykobjê–Krikatí */n/*. It was independently lost in PSJ and in the male genderlect of Karajá (and in the Javaé dialect). In Maxakalí and in the female genderlect of Karajá, it yielded *n* (underlying */d/*).

The Proto-Tupian form has a limited distribution: it is preserved in two Rondonian branches only, Tuparikém (Wayoró *-nõ* ‘another’ < Proto-Tupian *\*-nõ*; Nikulin & Andrade 2020: 306) and Rama-Puru (Karo *-nõ* ‘one of’; Gabas Jr 1999: 30, 2013).

PT *\*-nõ* is functionally identical to PMJ *\*-nũc ~ \*-nũ<sub>1j</sub> ~ \*-nũ<sub>1j</sub>°*, and the phonological similarity is striking. Only the palatal coda, reconstructed for PMJ based on the Maxakalí reflex *-nõy*, has prevented this comparison from being cited in 4.2.

**‘wet’:** Proto-Cerrado *\*-ŋgoñ* : Akuntsú *-ko*

The Proto-Cerrado form is given as *\*-ŋgoj’* in Nikulin 2020: 473, reflected as Khîsêjtê *-ngo* and Akwê-Xerente *-koi // -ko*, among others. This reconstruction must be updated to *\*-ŋgoñ* based on the long vowel in the Pykobjê–Krikatí reflex *-ncoo* (see fn. 2). No cognates in other Macro-Jê languages are known, but no stronger candidates for the Proto-Macro-Jê root for ‘wet’ are known either.<sup>28</sup> If Proto-Cerrado *\*-ŋgoñ* is a retention from Proto-Tupian, the original form must have been *\*-ŋgôñ°*.

On the Tupian side of the comparison, one finds Akuntsú *-ko* ‘wet’ (Aragon 2014: 138), with no known cognates elsewhere in Tupian; even the closely-related Tuparikém languages show noncognate terms for ‘wet’ (Wayoró *-txuup*, Tuparí *-súm-’e*, Makurap *-wuyyo*, Karitiana *-sebok*). No stronger candidate for the Proto-Tupian term for ‘wet’ is known.<sup>29</sup> If Akuntsú *-ko* is an archaism, the original form could have been *\*-ko*, *\*-k̃o*, or *\*-ŋgo*.

<sup>28</sup> Each Macro-Jê branch employs its own root(s) for this meaning: Krenak *hĩñot*; Maxakalí *-pato*; PSJ *\*-paŋpe* (the root is likely just *\*-pe*; compare also Kaingang *mrér* ‘wet’); Proto-Karajá *\*-tuku*; Ofayé <penó>; Rikbaktsa *-bibí* and *-hõrõ ~ -hõ*; Arikapú *-ũ*; Djeoromitxí *-boi* and *-bu*; Proto-Chiquitano *\*pãã-*.

<sup>29</sup> Each Tupian branch employs its own root(s) for this meaning: Puruborá *i(-)pəC* (Monserrat 2005: 19), Paiter *siab*, Mundurukú *-dîrem*, Yudjá *-’urú* and Xipaya *-súru*, Sateré-Mawé *-’apuk*, PTG *\*-ākîp*.

If the forms PMJ  $*\text{-}\eta\text{g}\hat{o}\tilde{n}^\circ$  ‘wet’ and PT  $*\text{-}ko \sim *k\tilde{o} \sim *-\eta\text{go}$  ‘wet’ actually existed, they may have been cognate. No parallels are known that would reveal the PT correspondence for PMJ  $*\hat{o}$ , the initial consonants correspond well, and the coda shows a mismatch that is precisely the object of discussion in this subsection.

**‘water’ / ‘liquid’:** Proto-Cerrado  $*\eta\text{g}\hat{o}j$ ’ and and Proto-Tuparikém  $*-\eta\text{g}\hat{i}$  ‘liquid’

The Proto-Cerrado term for ‘water’ is from Nikulin 2020: 473, where the apostrophe stands for the absence of an echo vowel. It is reflected, for example, as Khĩsêjtjê *ngô*, Panará *inkô*, and Xavante *ui // u* ‘still water’. Despite the striking similarity, it does not regularly correspond to PSJ  $*\eta\text{g}\hat{o}j$  ‘water’ (> Kaingang *goj* and Laklãnô *goj*): the former form points to PJ  $*\eta\text{g}\hat{o}j$ ’, and the latter to  $*\eta\text{gu}_1j$ , with a different nucleus and a different coda. Similar, but unrelated, are the Proto-Akuwê noun  $*kaj // *k\tilde{a}$  ‘flowing water’ (< PCerr  $*wyj$  or  $*wy\tilde{n}$  ‘river, whence Měbêngôkre *by-ti-re* ‘Xingu River’), the PCerr form  $*-\eta\text{go}\tilde{n}$  ‘wet’ (see above), and the Maxakalí noun *kônãg-kox* ‘river’ (possibly from *kônããg* ‘water’ and *-kox* ‘hole’). Maxakalí *-kux* ‘riverbank’ is phonologically comparable with PCerr  $*\eta\text{g}\hat{o}j$ ’ ‘water’, but a semantically closer cognate is available, PNJ  $*\{ca\}k\tilde{a}c$  ‘riverbank’ (> Khĩsêjtjê *sakhát // sakhárá*). I hesitate at deciding whether PCerr  $*\eta\text{g}\hat{o}j$ ’ ‘water’ and PSJ  $*\eta\text{g}\hat{o}j$  ‘water’ should be considered cognate; in any case, these roots are an innovation, since the PMJ term for ‘water’ is clearly reconstructible as  $*mbi_1n^\circ$ . If PCerr  $*\eta\text{g}\hat{o}j$ ’ is a semantically shifted reflex of a PMJ noun, its original form should be reconstructed as  $*\eta\text{gy}_1j$  or  $*\eta\text{gy}_1j^\circ$ .

On the Tupian side of the comparison, one finds Proto-Tuparikém  $*-\eta\text{g}\hat{i}$  ‘liquid’, reflected, among others, as Wayoró *-ngu* ‘liquid’ and Karitiana *-nge* ‘blood’ (Nikulin & Carvalho 2022: 33). The Karitiana meaning is clearly innovative, since the closely related Arikém retains a pan-Tupian root for ‘blood’ (< *nhaé*, < *nyaé* < PT  $*-j\text{uu}$ ; Nikulin & Carvalho 2022: 29). No cognates in other Tupian languages are known, and the Proto-Tupian term for ‘liquid’ is reconstructed as  $*-juu$  (Nikulin & Carvalho 2022: 30). If Proto-Tuparikém  $*-\eta\text{g}\hat{i}$  is nevertheless a semantically shifted reflex of a PT noun, its original form should be reconstructed as  $*-\eta\text{g}\text{uu}$ .

Proto-Cerrado  $*\eta\text{g}\hat{o}j$ ’ is reconstructed as an absolute stem, whose relationalized equivalent is  $*\text{-ga-}\eta\text{g}\hat{o}j$  (> Khĩsêjtjê *-kangô*, Panará *nankô / -rankô*, Xavante *-wa’u*). Proto-Tuparikém  $*-\eta\text{g}\hat{i}$  is a relational class I stem. The palatal coda in Proto-Cerrado is reconstructed based on the Xavante and Akwê-Xerente utterance-medial allomorphs, and it lacks a correspondence in Tuparikém. The correspondence PMJ  $*\eta\text{g} */\eta/$  : PT  $*\eta\text{g} */\eta/$  is unparalleled, but still imaginable, and PMJ  $*y_1$  does correspond to PT  $*u$ .

### 5.2.3. PMJ back vowel corresponding to PT $*e$

Nikulin (2020: 188–189) reconstructs the PMJ second-person pronoun (internal case) as  $*a$ , with reflexes such as Ofayé *e*, *e-*, Krenak *ho-ti*, Panará *ka*, and Kaingang *ã*. In addition, there is a similar second-person index, reconstructed as having two allomorphs,  $*a-$  with class I stems and  $*\emptyset-$  with class II stems (Nikulin 2020: 208–219). Its proposed reflexes are found in all major branches, including Chiquitano (Bésiro *a-/∅-*), Western (Djeoromitxí *a-*, Rikbaktsa *a-*, Ofayé *a-/∅-*), Karajá (*a-/∅-*), and Eastern (Maxakalí *ã-/∅-*, Krenak *a-/h-*, Khĩsêjtjê *a-/ng-/k-*,<sup>30</sup> Panará *a-/k-*,

<sup>30</sup> Khĩsêjtjê *k-* (the expected reflex of PMJ  $*\emptyset-$  before oral vowels, via  $*\emptyset > *fi > *iy > *g > *k$ ) no longer functions as an inflectional marker: it was ousted by the allomorph *ng-* (originally found before nasal vowels only, as in *ng-ĩnti* ‘your name’ or *ng-ũmndât // ng-ũmndârâ* ‘your wrist’), and forms such as *ng-ahrâ* ‘to play with you’ and *ng-ajkhêrê* ‘you yawn’—instead of the expected  $*k\text{-ahrâ}$ ,  $*k\text{-ajkhêrê}$ —are found in Khĩsêjtjê. The allomorph *k-* is preserved in the triadic kinship terms *k-áthêng // k-áthêngê* ‘your son, who happens to be my relative’, *k-átôt(-jê) // k-átôrô* ‘your mother, who happens to be my in-law (avoidance woman)’.

Xavante *a-* /a:-/ ‘second-person honorific’). Both cognate sets show somewhat irregular reflexes in some languages. The reconstruction *\*a* ‘you’ does not account for the Southern Jê reflexes (Kaingang *ã*, Laklãnõ *a* ‘you’), which rather point to PMJ *\*u*; normally PMJ *\*a* yields PJ *\*a* > PSJ *\*ã* > Kaingang *ẽ*, Laklãnõ *ã*. The reconstruction *\*a-* ‘second-person index’ fails to account for the reflexes in Ofayé (*a-* instead of the expected *\*e-*; points to PMJ *\*â-* or *\*y-*) and Krenak (*a-* instead of the expected *\*o-*; points to PMJ *\*a-*). It is important to note that person indices are unstressed in almost all Macro-Jê languages (Chiquitano is an exception), and the development of unstressed PMJ vowels is understudied. In any case, all reflexes of the pronoun and person index point to a PMJ back vowel (be it *\*a*, *\*u*, *\*ə*, *\*â*, or *\*y*).

In Proto-Tupian, *\*e-* is reconstructed as an absolutive/genitive second-person prefix, from which the pronoun *\*ẽT* is derived, just like the pronoun *\*õT* is derived from the first-person prefix *\*o-*. It has reflexes in almost all Tupian languages. Before consonant-initial stems, it is reflected as Makurap *e-*, Wayoró *e-*, Karitiana *a-*, Karo *e-*, Puruborá *ε-*, Paiter *e-*, Yudja *e-*, Mundurukú *e-*, Sateré-Mawé *e-*, Apyãwa *e-* ‘2CRF’, among many other reflexes. Before vowel-initial stems, it shows asyllabic allomorphs in some languages, such as Yudja *l-*. The Tupi–Guaranian reflex is only used anaphorically, particularly when a second-person possessor on a noun or a second-person argument of a gerund of an intransitive verb is coreferential with some other participant. As for noncoreferential uses, it has been ousted by the clitic *\*(e)nde=* in the Tupi–Guaranian languages.

The possibility of linking the Proto-Macro-Jê second-person markers and the Proto-Tupian index *\*e-* was not considered in 4 due to the mismatch in vowel backness. Although the sound correspondence is not recurrent, the cognation hypothesis is still plausible, since irregular vowel changes are otherwise known to be common in grammatical morphemes (as seen in the Macro-Jê cognate set discussed in this subsection).

## 6. Conclusion

In this article, I have assembled the lexical evidence supporting the proposed common origin of the Macro-Jê and Tupian families. Despite their limited number, the matches analyzed in 4–5 show recurrent sound correspondences that are hardly attributable to chance or to language contact: they mostly involve basic vocabulary (including 19 items on the 110-item Swadesh list: ‘meat’, ‘name’, ‘smoke’, ‘tooth’, ‘ingest’ = ‘to eat/drink’, ‘tree’, ‘liver’, ‘foot’, ‘burn’, ‘to give’, ‘to stand’, ‘earth’, ‘to kill’, ‘white’, ‘I’, ‘heavy’, ‘to go/come’, ‘ashes’, ‘snake’) and grammatical morphemes, and multiple proposed cognate sets involve data from Macro-Jê and Tupian branches spoken very far from each other (e.g. Jê and Karitiana). Moreover, the Macro-Jê–Tupian comparanda often involve matching codas, but the Eastern branch of Macro-Jê—the one that most faithfully preserves PMJ codas—is geographically removed from Rondônia, where most (non-Tupi–Guaranian) Tupian languages are spoken. Therefore, the similarities noted above can hardly result from language contact, and common genetic origin is the best explanation available.

Further research will need to concentrate on the lexical reconstruction of Proto-Macro-Jê and Proto-Tupian. In this article, I have proposed multiple hypothetical PMJ and PT forms based on reflexes in only one branch or language; I predict that some of these etymologies can be further strengthened by identifying previously unnoticed cognates in the attested languages. As of now, relatively few Macro-Jê and Tupian etymologies are currently known. Nikulin’s (2020) dissertation lists 188 PMJ reconstructions, of which some are quite dubious, and others involve reflexes in one first-level branch only. Although no comprehensive source on

Tupian etymology exists so far, I am currently working on a Tupian etymological dictionary, and my draft has 255 entries that involve reflexes in more than one first-level branch (including compounds). It is certainly possible to reconstruct many more PMJ and PT forms.

Another direction for further research is to include other language families into consideration. In my opinion, language families and isolates such as Cariban, Bororoan, Karirian, and Yaathê are very likely related to Macro-Jê and Tupian, but a search for possible cognates in these languages is complicated by the fact that Proto-Cariban, Proto-Bororoan, Proto-Karirian, and pre-Yaathê do not have codas (except for the marginal coda *\*-j* in pre-Yaathê; Silva forthc.). If some or all of these languages are ultimately related to Macro-Jê and Tupian, it is unclear whether Macro-Jê and Tupian codas should be matched to zero (under the assumption that *\*C(r)VC*-structures yielded *\*C(r)V* in these languages), or whether Macro-Jê *\*C(r)VC*-structures and Tupian *\*CVC*-structures should be matched to polysyllabic roots in other languages (under the assumption that *\*C(r)VCV*-structures yielded *\*C(r)VC* in Macro-Jê and Tupian). This uncertainty leaves too much room for false positives at this stage of investigation. Some other language families — notably Katukina–Harakmbut, Mataguayan, and Guaicuruan — have comparable syllable structures with robust codas. Indeed, there are several promising lookalikes with matching onsets and codas involving these families, as in PT *\*-kaT* : Proto-Mataguayan *\*-káʔt-* : Harakmbut *-kot* ‘to fall’; PMJ *\*(-)mbrôŋ* : PT *\*-mboK* ‘ashes’ : Proto-Mataguayan *\*-máʔk* ‘powder’; PT *\*atʰa* ‘fire’ : Katukina *ita*, Harakmbut *ʔuta* ‘firewood’ (but Proto-Mataguayan *\*ʔitáχ* ‘fire’, with a uvular coda); Proto-Tupian *\*əK* : Katukina *hak*, Harakmbut *jak* /*hak*/ ‘house’; PMJ *\*-mbâ* : Proto-Tupian *\*-pi(-)ʔa/\*mbi-ʔa* : Katukina *ma*, Harakmbut *-me* ‘liver’. If more of such matches are found and if regular sound correspondences are identified linking the aforementioned languages, the Macro-Jê–Tupian hypothesis may turn out to be the tip of an iceberg — quite possibly, the largest macrofamily in the Americas.

#### Acknowledgments

I acknowledge the generosity of all speakers and researchers of Macro-Jê and Tupian languages who have dedicated their time to answering my questions about the meaning, form, spelling, and inflection of so many words. In particular, I am grateful to Andrés Pablo Salanova, Antônia Fernanda Souza Nogueira, Antonia Socoré Masai, Chadawa Juruna, Ignacia Tomichá Yopié, Jamthô Suyá, Khawiri Suyá, Luca Ciucci, Mário André Coelho da Silva, Mawaré Juruna, Oregô Tapayuna, Ricardo Capêrkô Canela, Talles Huryty Tapirapé, Tawayaku Kayabi, Tenywaawi Tapirapé, Victoriano Julián Laverán Ramos, and Yrywaxã Tapirapé. I thank an anonymous JoLR reviewer for their comments on the manuscript. An early version of this paper was presented at II Seminário do GELCIA / II SIPLI-Norte (Federal University of Pará, Belém) on November 18, 2019; I thank the audience for their comments.

#### Provenance of linguistic data

Unreferenced examples come from the following sources.

|              |   |
|--------------|---|
| Akwê-Xerente | Krieger & Krieger 1994  |
| Akuntsú      | Aragon 2008, 2014   |
| Apinajé      | Albuquerque 2012  |
| Apyãwa       | Almeida et al. 1983; Tenywaawi Tapirapé, p. c.; Yrywaxã Tapirapé, p. c. |
| Arikapú      | R. Ribeiro 2008; Arikapú et al. 2010                                    |
| Arikém       | Nimuendajú 1932; Rondon & Faria 1948                                    |

|                      |   |
|----------------------|---|
| Aruá                 | Sekelj 1948   |
| Awetí                | Sabino 2016   |
| Bésiro               | Parapaino Castro 2008                                       |
| Canela–Krahô         | Grupp 2015  |
| Djeoromitxí          | M. Ribeiro 2008   |
| Eastern Chiquitano   | Fuss & Riester 1986   |
| Gavião               | Moore 1984; Felzke & Moore 2019                             |
| Guarasugwe           | Ramirez et al. 2017   |
| Harakmbut            | Tripp 1995  |
| Ka'apor              | Kakumasu & Kakumasu 2007                                    |
| Kaingang             | Wiesemann 2011  |
| Kajkwakhrattxi       | Camargo 2010  |
| Kamayurá             | Seki 2000   |
| Karajá               | Ribeiro 2012a; Karajá et al. 2013                           |
| Karitiana            | Landin 2005, Rocha 2011; Storto 2019                        |
| Karo                 | Gabas Jr 1999   |
| Katukina             | dos Anjos 2011  |
| Kawaiwete            | Weiss 2005  |
| Kepkiriwat           | Rondon & Faria 1948   |
| Khîsêtjê             | Nonato et al. 2012; Jamthô Suyá, p. c., Khawiri Suyá, p. c. |
| Krenak               | Seki n/d  |
| Kuruaya              | Costa 2002; Picanço 2005, 2019                              |
| Laklânô              | Alves Jr 2014   |
| Makurap              | Sekelj 1948; Braga 2005                                     |
| Malalí               | Silva & Nikulin 2021  |
| Maxakalí             | Silva 2020; Silva, p. c.                                    |
| Mbyá                 | Dooley 2006   |
| Mêbêngôkre           | Salanova, p. c.   |
| Miguelêño Chiquitano | own field data  |
| Mundurukú            | Crofts 1985; Picanço 2005                                   |
| Ofayé                | Hanke 1964; Gudschinsky 1974; Ribeiro 2004b; Oliveira 2006  |
| Old Tupí             | Barbosa 1956  |
| Paiter               | Bontkes 1978  |
| Panará               | Bardagil-Mas 2018   |
| Paraguayan Guaraní   | Centurión Servin & Davalos Arce 2009                        |
| Parintintin          | Betts 1981  |
| Parkatêjê            | Araújo 2016   |
| Pataxó-Hãhãhãe       | Silva & Nikulin 2021  |
| Proto-Mataguayan     | Nikulin & Carol forthc.                                     |
| Puruborá             | Galucio 2005  |
| Pykobjê–Krikatí      | Pries 2008  |
| Rikbaktsa            | Tremaine 2007   |
| Sakurabiat           | Galucio 2001; Snethlage 2015                                |
| Salamãý              | Galucio et al. 2015   |
| Sateré-Mawé          | Ribeiro 2010; Silva 2010                                    |
| Siriono              | Gasparini & Dicarere Méndez 2015                            |
| Tapiete              | Gonzalez 2005   |
| Tuparí               | Alves 2004; Singerman 2018                                  |
| Wayoró               | Nogueira 2011, 2019; Nogueira et al. 2021                   |
| Yudja                | Fargetti 2001; Chadawa Juruna, p. c.                        |
| Xavante              | McLeod & Mitchell 1977; Lachnitt 1987                       |
| Xipaya               | Fargetti & Rodrigues 2008                                   |
| Zoró                 | Galucio et al. 2015   |

**Abbreviations**

- X // Y X is the utterance-medial allomorph, Y is the utterance-final allomorph  
 {X} fossilized material (noncognate part for which cognation is not asserted)  
 <X> material cited *verbatim* after premodern sources  
 X : Y X corresponds to Y

**Grammatical abbreviations:** 1/2/3 = first/second/third person; ALZ = alienizer; ANTP = antipassive; AUG = augmentative; AUX = auxiliary; CRF = coreferential; DU = dual; F = finite; IMPF = imperfective; INCL = inclusive; INV = inverse; NCRF = noncoreferential; NF = nonfinite; NMLZ = nominalizer; PL = plural; PSSD = possessed; PRG = progressive; PRS = present; PST = past; REF = referentializer; SG = singular; <sub>A</sub> = agent; <sub>P</sub> = patient; <sub>V</sub> = verb; NP = noun phrase.

**Phonological abbreviations:** C = consonant; J = palatal consonant; K = velar consonant; N = nasal consonant; V = vowel.

**Language names:** Guaj. = Guajajara (Tenetehara); PCerr = Proto-Cerrado; PJ = Proto-Jê; PMJ = Proto-Macro-Jê; PNJ = Proto-Northern Jê; PSJ = Proto-Southern Jê; PT = Proto-Tupian; PTG = Proto-Tupi-Guaranian; TG = Tupi-Guaranian.

**References**

- Adelaar, Willem F. H. 2008. Relações externas do Macro-Jê. O caso do Chiquitano. In: Stella Telles, Aldir Santos de Paula (orgs.). *Topicalizando Macro-Jê*: 9–28. Recife: Nectar.
- Aikhenvald-Angenot, Alexandra Y., Jean-Pierre Angenot. 1992. The South-American Proto-Ge and the Old World. In: Shevoroshkin, Vitaly V. (ed.). *Nostratic, Dene-Caucasian, Austric and Amerind. Materials from the First International Interdisciplinary Symposium on Language and Prehistory, Ann Arbor, 8–12 November, 1988*: 403–18. Bochum: Universitätsverlag Dr. Norbert Brockmeyer.
- Albuquerque, Francisco Edviges (org.). 2012. *Dicionário Escolar Apinayé. Panhĩ kapēr–kupẽ kapēr. Apinayé–português*. Belo Horizonte: Editora da Faculdade de Letras–UFMG.
- Almeida, Antônio, Irmãzinhas de Jesus, Luiz G. de Paula. 1983. *A língua tapirapé*. Rio de Janeiro: Biblioteca Reprodutiva Xerox.
- Alves, Poliana Maria. 2004. *O léxico do Tuparí: proposta de um dicionário bilíngüe*. Ph.D. dissertation, Universidade Estadual Paulista “Júlio de Mesquita”.
- Alves Jr, Ozias. 2014. *Parlons xokleng / laklãñõ. Langue indigène du sud du Brésil*. Paris: L’Harmattan.
- Aragon, Carolina Coelho. 2008. *Fonologia e aspectos morfológicos e sintáticos da língua Akuntsú*. MA thesis, Universidade de Brasília.
- Aragon, Carolina Coelho. 2014. *A grammar of Akuntsú, a Tupian language*. Ph.D. dissertation, University of Hawai’i at Mānoa.
- Araújo, Leopoldina. 2016. *Dicionário Parkatêjê–português*. Belém: s/ed.
- Arikapú, Mamoá, Nazaré Arikapú, Hein van der Voort. 2010. *Vocabulário Arikapú–português*. Baltimore: Cadernos de Etnolingüística. (Monografias, 1.)
- Barbosa, P<sup>e</sup> A. Lemos. 1956. *Curso de Tupi Antigo: gramática, exercícios, textos*. Rio de Janeiro: Livraria São José.
- Bardagil-Mas, Bernat. 2018. *Case and agreement in Panará (Naamval en congruentie in het Panará)*. Ph.D. dissertation, Rijksuniversiteit Groningen.
- Betts, LaVera. 1981. *Dicionário Parintintin–português, português–parintintin*. Cuiabá: Sociedade Internacional de Lingüística.
- Bontkes, Willem. 1978. *Dicionário preliminar Suruí–português, português–Suruí*. Porto Velho: Summer Institute of Linguistics.
- Braga, Alzerinda. 2005. *Aspects morphosyntaxiques de la langue Makurap/Tupi*. Ph.D. dissertation, Université de Toulouse–Le Mirail.
- Camargo, Nayara da Silva. 2010. *Língua Tapayúna: aspectos sociolingüísticos e uma análise fonológica preliminar*. MA thesis, Universidade Estadual de Campinas. doi:10.47749/T/UNICAMP.2010.770701
- Carvalho, Fernando O. de. 2019. Revisitando o Proto-Jurúna: a reconstrução das oclusivas orais. In: Edna dos Santos Oliveira, Eduardo Alves Vasconcelos, Romário Duarte Sanches (eds.). *Estudos linguísticos na Amazônia*: 215–236. Campinas: Pontes Editores.



- Carvalho, Fernando O. de. 2022. On the Guaranian evidence for two Proto-Tupi–Guarani affricates. *Journal of Language Relationship* 20(1): 81–112.
- Carvalho, Fernando O. de. 2023. Proto-Tupi–Guarani did not have a palatalized velar stop. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 18(1): e20220013. doi:10.1590/2178-2547-BGOELDI-2022-0013
- Carvalho, Fernando O. de. Forthcoming. *A comparative reconstruction of Proto-Tupi-Guarani phonology* [provisional title]. Berlin: Language Science Press. (Topics in Phonological Diversity.)
- Carvalho, Fernando O. de., Joshua Birchall. 2022. A comparative reconstruction of Proto-Tupi–Guarani kinship terminology. *LIAMES* 22: e022001. doi:10.20396/liames.v22i00.8666489
- Centurión Servín, Celsa, Juana Davalos Arce. 2009. *Milibri pocket. Castellano–guaraní, guaraní–castellano: grafía tradicional y científica*. Asunción: Milibri Ediciones.
- Ciucci, Luca. 2020. *Los clasificadores del chiquitano (bésiro)*. Talk presented at Kusíribo Oberabakáx ñomé Mumanit-yakatoe aukí na Niki Mamonkóka (November 13, 2020). doi:10.13140/RG.2.2.30849.68963
- Corrêa-da-Silva, Beatriz Carretta. 2010. *Mawé/Awetí/Tupi-Guaraní: relações lingüísticas e implicações históricas*. Ph.D. dissertation, Universidade de Brasília.
- Costa, Raimundo Nonato Vieira. 2002. Fonologia segmental da língua Kuruaya. *Moara* 17: 85–101. doi:10.18542/moara.v1i17.3151
- Crofts, Marjorie. 1973. *Gramática Munduruku*. Brasília: Summer Institute of Linguistics.
- Crofts, Marjorie. 1985. *Aspectos da língua Mundurukú*. Brasília: Summer Institute of Linguistics.
- Davis, Irvine. 1968. Some Macro-Jê relationships. *International Journal of American Linguistics* 34(1): 42–47. doi:10.1086/464994
- Dooley, Robert A. 2006. *Léxico Guaraní, dialeto Mbyá com informações úteis para o ensino médio, a aprendizagem e a pesquisa lingüística*. Cuiabá: Sociedade Internacional de Linguística.
- dos Anjos, Zoraide. 2011. *Fonologia e gramática Katukina-Kanamari*. Ph.D. dissertation, Vrije Universiteit Amsterdam.
- Dourado, Luciana Gonçalves. 2001. *Aspectos morfosintáticos da língua Panará (Jê)*. Ph.D. dissertation, Universidade Estadual de Campinas. doi:10.47749/T/UNICAMP.2001.206356
- Drude, Sebastian. 2009. Nasal harmony in Awetí: a declarative account. *ReVEL — Revista Virtual de Estudos da Linguagem* 7(3).
- Drude, Sebastian. 2011. Awetí in relation with Kamayurá: the two Tupian languages of the Upper Xingu. In: Bruna Franchetto (org). *Alto Xingu. Uma sociedade multilíngüe*: 155–192. Rio de Janeiro: Museu do Índio/FUNAI.
- Estevam, Adriana Machado. 2011. *Morphosyntaxe du xavante. Langue jê du Mato Grosso (Brésil)*. Ph.D. dissertation, Université Paris-Diderot (Paris-VII).
- Fargetti, Cristina Martins. 2001. *Estudo fonológico e morfosintático da língua Juruna*. Ph.D. dissertation, Universidade Estadual de Campinas. doi:10.47749/T/UNICAMP.2001.222157
- Fargetti, Cristina Martins, Carmen L. Reis Rodrigues. 2008. Consoantes do Xipaya e do Juruna — uma comparação em busca do proto-sistema. *Alfa* 52(2): 535–563.
- Fargetti, Cristina Martins, Carmen Lúcia Reis Rodrigues. 2021. Vogais do Juruna e do Xipaya — uma comparação. *Moara* 58: 186–204. 10.18542/moara.v0i58.10906
- Felzke, Lediane Fani, Denny Moore. 2019. Terminologias de parentesco dos grupos da família lingüística Mondé. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 14(1): 15–32. doi:10.1590/1981.81222019000100003
- Fuss, Max, Jürgen Riester. 1986. *Zúbaka. La Chiquitania: visión antropológica de una región en desarrollo. Tomo I: Vocabulario español–chiquito y chiquito–español*. La Paz: Los Amigos del Libro.
- Gabas Jr, Nilson. 1989. *Estudo fonológico da língua Karo (Arara de Rondônia)*. MA thesis, Universidade Estadual de Campinas. doi:10.47749/T/UNICAMP.1989.38012
- Gabas Jr, Nilson. 1999. *A grammar of Karo, Tupi (Brazil)*. Ph.D. dissertation, University of California, Santa Barbara.
- Galucio, Ana Vilacy. 1994. Fonologia segmental da língua Mekens. In: *Anais do IX Encontro Nacional da ANPOLL (Caxambu, junho de 1994). Área de Lingüística, v. 2*: 988–997. João Pessoa: Associação Nacional de Pós-Graduação e Pesquisa em Letras e Lingüística.
- Galucio, Ana Vilacy. 2001. *The morphosyntax of Mekens (Tupi)*. Ph.D. dissertation, University of Chicago.
- Galucio, Ana Vilacy. 2005. Puruborá: notas etnográficas e lingüísticas recentes. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 1(2): 159–192.
- Galucio, Ana Vilacy, Antônia Fernanda de Souza Nogueira. 2012. Comparative study of the Tupari branch of the Tupi family: contributions to understanding its historical development and internal classification. In: *Memorias del V Congreso de Idiomas Indígenas de Latinoamérica, 6–8 de octubre de 2011, Universidad de Texas en*

- Austin. Austin: Center for Indigenous Languages of Latin America (CILLA). Available online at: [ailla.utexas.org/sites/default/files/documents/Galucio\\_Nogueira\\_CILLA\\_V.pdf](http://ailla.utexas.org/sites/default/files/documents/Galucio_Nogueira_CILLA_V.pdf), accessed on 24/02/2023.
- Galucio, Ana Vilacy, Antônia Fernanda de Souza Nogueira. 2018. From object nominalization to object focus: the innovative A-alignment in the Tuparian languages (Tupian family). *Journal of Historical Linguistics* 8(1): 95–127. doi:10.1075/jhl.16025.gal
- Galucio, Ana Vilacy, Sérgio Meira, Joshua Birchall, Denny Moore, Nilson Gabas Júnior, Sebastian Drude, Luciana Storto, Gessiane Picanço, Carmen Reis Rodrigues. 2015. Genealogical relations and lexical distances within the Tupian linguistic family. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 10(2): 229–274. doi:10.1590/1981-81222015000200004
- Gasparini, Noé, Victor Hugo Dicarere Méndez. 2015. *Diccionario Siriono. Siriono–castellano–inglés y castellano–siriono*. Trinidad: Editorial Tiempos del Beni.
- Gerardi, Fabrício Ferraz, Stanislav Reichert, Carolina Aragon, Tim Wientzek, Johann-Mattis List, Robert Forkel. 2022. *TuLeD. Tupian Lexical Database* (v0.12). doi:10.5281/zenodo.6572576
- Gomes, Dionei Moreira. 2007. Voz média em Mundurukú: uma análise do morfema *je-* (*dze-*). In: Ana Suelly Ar-ruda Câmara Cabral, Aryon Dall’Igna Rodrigues (orgs.). *Línguas e culturas Tupi, vol. I*: 317–323. Campinas: Curt Nimuendajú; Brasília: LALLI/UnB.
- Gonzalez, Hebe Alicia. 2005. *A grammar of Tapiete (Tupi–Guarani)*. Ph.D. dissertation, University of Pittsburgh.
- Gudschinsky, Sarah C. 1971. Ofaié-Xavante, a Jê language. In: Sarah C. Gudschinsky (ed.). *Estudos sobre línguas e culturas indígenas: Trabalhos lingüísticos realizados no Brasil*: 1–16. Brasília: SIL–Instituto Lingüístico de Verão.
- Gudschinsky, Sarah C. 1974. Fragmentos de Ofaié: a descrição de uma língua extinta. In: Loraine Irene Bridgeman (ed.). *Série Lingüística, no. 3*: 177–249. Brasília: Summer Institute of Linguistics.
- Guérios, Rosário F. Mansur. 1939. O nexó lingüístico Bororo/Merrime-Caiapó (contribuição para a unidade genética das línguas americanas). *Revista do Círculo de Estudos “Bandeirantes”* 2: 61–74.
- Greenberg, Joseph H. 1987. *Language in the Americas*. Stanford University Press.
- Grupp, Bernhard. 2015. *Dicionário Canela: Canela–português–inglês, português–Canela, inglês–Canela*. 2<sup>nd</sup> ed. Barra do Corda: Missão Cristã Evangélica do Brasil.
- Hanke, Wanda. 1964. *Völkerkundliche Forschungen in Südamerika: Verlöschende Urzeit im Innern Brasiliens (= Kultur-geschichtliche Forschungen)*. Braunschweig: Limbach.
- Hanke, Wanda, Morris Swadesh, Arion Dall’Igna Rodrigues. 1958. Notas de fonologia Mekens. In: Juan Comas (ed.). *Miscellanea Paul Rivet octogenario dicata, vol. II*: 187–217. Mexico City: Universidad Nacional Autónoma de México.
- Harrison, Carl, Carole Harrison. 2013. *Dicionário Guajajara–português*. Anápolis: Associação Internacional de Lingüística–SIL Brasil.
- Kakumasu, James Y., Kiyoko Kakumasu. 2007. *Dicionário por tópicos Kaapor–português*. Cuiabá: Associação Internacional de Lingüística–SIL Brasil.
- Karajá, Hatawaki, Juanahu Karajá, Leandro Karajá, Teribre Karajá, Wadoi Karajá, Woubedu Karajá. 2013. *Dicionário Enciclopédico Inyrybè/Karajá–português brasileiro*. Rio de Janeiro: Museu do Índio/Prodoclin. Available online at: [http://prodoclin.museudoindio.gov.br/images/conteudo/Karaja/produtos\\_pesquisadores/Karaj%C3%A1\\_Iny\\_Lexicon.pdf](http://prodoclin.museudoindio.gov.br/images/conteudo/Karaja/produtos_pesquisadores/Karaj%C3%A1_Iny_Lexicon.pdf), accessed on 24/02/2023.
- Krieger, Wanda Braidotti, Guenther Carlos Krieger (orgs.). 1994. *Dicionário escolar: Xerente–português; português–Xerente*. Rio de Janeiro: Junta das Missões Nacionais da Convenção Batista Brasileira.
- Lachnitt, Georg. 1987. *Romnhitsi’ubumro. A’uwē mreme–waradzu mreme. Dicionário xavante–português*. Campo Grande: Missão Salesiana de Mato Grosso.
- Landin, David. 2005. *Dicionário e léxico Karitiana / português*. 2<sup>nd</sup> ed. Cuiabá: Sociedade Internacional de Lingüística.
- Lima, Suzi. 2014. *The grammar of individuation and counting*. Ph.D. dissertation, University of Massachusetts Amherst. doi:10.7275/msad-1e04
- Martins, Andrébio Márcio Silva. 2009. Um estudo comparativo-lexical das famílias Kamakã e Purí. In: Silvia Lúcia Bigonjal Braggio, Sinval Martins de Sousa Filho (orgs.). *Línguas e culturas Macro-Jê*: 231–238. Goiânia: Vieira.
- Martins, Andrébio Márcio Silva. 2011. *Uma avaliação da hipótese de relações genéticas entre o Guató e o tronco Macro-Jê*. Ph.D. dissertation, Universidade de Brasília.
- McLeod, Ruth, Valerie Mitchell. 1977. *Aspectos da língua Xavante*. Brasília: Summer Institute of Linguistics.
- Meira, Sérgio, Spike Gildea, B. J. Hoff. 2010. On the origin of ablaut in the Cariban family. *International Journal of American Linguistics* 76(4): 477–515. doi:10.1086/658055

- Meira, Sérgio, Sebastian Drude. 2013. Sobre a origem histórica dos “prefixos relacionais” das línguas Tupí–Guaraní. *Cadernos de Etnolingüística* 5(1): 1–30.
- Meira, Sérgio, Sebastian Drude. 2015. A summary reconstruction of Proto-Maweti–Guarani segmental phonology. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 10(2): 275–296. doi:10.1590/1981-81222015000200005
- Monserrat, Ruth Maria Fonini. 2005. Notícia sobre a língua Puruborá. In: Aryon Dall’Igna Rodrigues, Ana Suely Arruda Câmara Cabral (eds.). *Novos estudos sobre línguas indígenas*: 9–22. Brasília: Editora UnB.
- Moore, Denny. 1984. *Syntax of the language of the Gavião Indians of Rondônia, Brazil*. Ph.D. dissertation, City University of New York.
- Moore, Denny, Ana Vilacy Galucio. 1994. Reconstruction of Proto-Tupari consonants and vowels. In: Margaret Langdon, Leanne Hinton (eds.). *Proceedings of the Meeting of the Society for the Study of the Indigenous Languages of the Americas, July 2–4, 1993, and the Hokan-Penutian Workshop, July 3, 1993, both held at the 1993 Linguistic Institute at Ohio State University in Columbus, Ohio*: 119–137. Berkeley: Survey of California and other Indian Languages.
- Nikulin, Andrey. 2015. Apofonia e sistema vocálico do Proto-Jê Meridional: contribuição para estudos comparativos das línguas Jê. *Revista Brasileira de Linguística Antropológica* 7(2): 275–306. doi:10.26512/rbla.v7i2.20601
- Nikulin, Andrey. 2017. A phonological reconstruction of Proto-Cerrado (Jê family). *Journal of Language Relationship* 15(3): 147–180. doi:10.31826/jlr-2018-153-404
- Nikulin, Andrey. 2020. *Proto-Macro-Jê: um estudo reconstrutivo*. Ph.D. dissertation, Universidade de Brasília.
- Nikulin, Andrey. Forthc. Reconstrução fonológica, ou como reconstruir o Proto-Tuparikém. Submitted to: Stella Telles, Cláudio Couto, Emerson Carvalho, and W. Leo Wetzels (eds.). *Tópicos na estrutura sonora das línguas indígenas sul-americanas: uma introdução à análise fonológica*. Available online at: [www.academia.edu/85174094/Reconstru%C3%A7%C3%A3o\\_fonol%C3%B3gica\\_ou\\_como\\_reconstruir\\_o\\_Proto\\_Tuparik%C3%A9m\\_submetido\\_](http://www.academia.edu/85174094/Reconstru%C3%A7%C3%A3o_fonol%C3%B3gica_ou_como_reconstruir_o_Proto_Tuparik%C3%A9m_submetido_), accessed on 24/02/2023.
- Nikulin, Andrey, Rafael Andrade. 2020. The rise and fall of approximants in the Tuparian languages. *Journal of Language Relationship* 18(4): 284–319. doi:10.31826/jlr-2020-183-412
- Nikulin, Andrey, Javier Carol. Forthc. *Historical phonology of Mataguayan*. Berlin: Language Science Press. (Topics in Phonological Diversity.)
- Nikulin, Andrey, Fernando O. de Carvalho. 2018. Prehistoria de las lenguas y familias lingüísticas del Gran Chaco, de la meseta brasileña y cercanías: propuesta de base de datos léxicos y resultados preliminares. In: María Alejandra Regúnaga, Silvia Andrea Spinelli, María Emilia Orden (eds.). *IV Encuentro de Lenguas Indígenas Americanas-ELIA: libro de Actas*: 545–560. Santa Rosa: Universidad Nacional de La Pampa.
- Nikulin, Andrey, Fernando O. de Carvalho. 2019. Estudos diacrônicos de línguas indígenas brasileiras: um panorama. *Macabéa — Revista Eletrônica do Netlli* 8(2): 255–305. doi:10.47295/mren.v8i2.1910
- Nikulin, Andrey, Fernando O. de Carvalho. 2022. A revised reconstruction of the Proto-Tupian vowel system. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 17(2): e20210035. doi:10.1590/2178-2547-BGOELDI-2021-0035
- Nikulin, Andrey, Andrés Pablo Salanova. 2019. Northern Jê verb morphology and the reconstruction of finiteness alternations. *International Journal of American Linguistics* 85(4): 533–567. doi:10.1086/704565
- Nikulin, Andrey, Andrés Pablo Salanova. 2022. O enfraquecimento diacrônico de consoantes em Mëbêngôkre. In: Edna dos Santos Oliveira, Eduardo Alves Vasconcelos, Romário Duarte Sanches (eds.). *Estudos Linguísticos na Amazônia, vol. 2*: 121–143. Campinas: Pontes.
- Nikulin, Andrey, Mário André Coelho da Silva. 2020. As línguas Maxakalí e Krenák dentro do tronco Macro-Jê. *Cadernos de Etnolingüística* 8(1): 1–64.
- Nimuendajú, Curt. 1932. Wortlisten aus Amazonien. *Journal de la Société des américanistes* 24(1): 93–119. doi:10.3406/jsa.1932.1846
- Nogueira, Antônia Fernanda de Souza. 2011. *Wayoro êmêto: fonologia segmental e morfossintaxe verbal*. MA thesis, Universidade de São Paulo. doi:10.11606/D.8.2011.tde-22102012-093617
- Nogueira, Antônia Fernanda de Souza. 2019. *Predicação na língua Wayoro (Tupi): propriedades de finitude*. Ph.D. dissertation, Universidade de São Paulo. doi:10.11606/T.8.2019.tde-14082019-101006
- Nogueira, Antônia Fernanda Souza, Ana Vilacy Galucio, Nicole Soares-Pinto, Adam Roth Singerman. 2019. Termos de parentesco nas línguas Tuparí (família Tupí). *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 14(1): 33–64. doi:10.1590/1981.81222019000100004
- Nogueira, Antônia Fernanda de Souza, comunidade Wajuru, Clenilson Miranda de Sousa. 2021. Apresentação da primeira versão do vocabulário Wayoro–português. In: Patrícia Goulart Tondineli (org.). *(Re)vitalizar línguas minorizadas e/ou ameaçadas: teorias, metodologias, pesquisas e experiências*: 83–108. Porto Velho: Coleção Pós-Graduação da UNIR–EDUFRO.

- Nonato, Rafael, Jamtô Suyá, Kawiri Suyá. 2012. *Dicionário Kĩsêdjê-português*. Rio de Janeiro: Museu do Índio/Prodoclin. Available online at: [http://prodoclin.museudoindio.gov.br/images/conteudo/kisedje/dicionario\\_kisedje1.pdf](http://prodoclin.museudoindio.gov.br/images/conteudo/kisedje/dicionario_kisedje1.pdf), accessed on 24/02/2023.
- Oliveira, Christiane Cunha de. 2005. *The language of the Apinajé people of Central Brazil*. Ph.D. dissertation, University of Oregon.
- Oliveira, Maria das Dores de. 2006. *Ofayé, a língua do povo do mel: fonologia e gramática*. Ph.D. dissertation, Universidade Federal de Alagoas.
- Pache, Matthias. 2023. Evidence for a Chibcha–Jê connection. *International Journal of American Linguistics* 89(2): 219–253. doi:10.1086/723641
- Palha, Luiz. 1942. *Ensaio de gramática e vocabulário da língua Karajá falada pelos índios remeiros do rio 'Araguaia'*. Rio de Janeiro: Olímpica.
- Parapaino Castro, Pablino. 2008. *Isiukiché nikorokó bésiro. Guía de escritura del idioma bésiro*. Santa Cruz de la Sierra: Unión de Artesanos de la Tierra – UNIARTE.
- Picanço, Gessiane Lobato. 2005. *Mundurukú: phonetics, phonology, synchrony, diachrony*. Ph.D. dissertation, University of British Columbia. doi:10.14288/1.0092991
- Picanço, Gessiane Lobato. 2019. *A fonologia diacrônica do Proto-Mundurukú (Tupí)*. Curitiba: Appris.
- Pinheiro, Edilson, Maria de Nazaré Duarte, Gessiane Picanço. 2020. A ortografia Mundurukú em questão: uma proposta de adaptação com foco em não falantes. *Revista Brasileira de Línguas Indígenas* 3: 28–39. doi:10.18468/rbli.2020v3n2.p28-39
- Pries, Stanley T. 2008. *Dicionário Gavião-Krikati*. Mimeo.
- Ramirez, Henri, Valdir Vegini, Maria Cristina Victorino de França. 2015. Koropó, puri, kamakã e outras línguas do Leste brasileiro: revisão e proposta de nova classificação. *LIAMES* 15(2): 223–277. doi:10.20396/liames.v15i2.8642302
- Ramirez, Henri, Valdir Vegini, Maria Cristina Victorino de França. 2017. O warázu do Guaporé (tupí-guarani): primeira descrição linguística. *LIAMES* 17(2): 411–506. doi:10.20396/liames.v17i0.8647468
- Reiter, Sabine. 2011. *Ideophones in Awetí*. Ph.D. dissertation, Christian-Albrechts-Universität zu Kiel.
- Ribeiro, Eduardo Rivail. 2002. O marcador de posse alienável em Karirí: um morfema Macro-Jê revisitado. *LIAMES* 2: 29–46. doi:10.20396/liames.v2i1.1403
- Ribeiro, Eduardo Rivail. 2004a. Prefixos relacionais em Jê e Karajá: um estudo histórico-comparativo. *LIAMES* 4: 91–101. doi:10.20396/liames.v4i1.1427
- Ribeiro, Eduardo Rivail. 2004b. *A língua Ofayé: documentando uma língua ameaçada*. Talk presented at IEL/Universidade Estadual de Campinas on 2/8/2004.
- Ribeiro, Eduardo Rivail. 2009. Old data, new cognates: the case of the “marker of alienable possession” in the Kamakã, Purí, and Krenák families. *Cadernos de Etnolingüística. Série Notas* 2: 1–5.
- Ribeiro, Eduardo Rivail. 2011. Prefixos relacionais como evidência histórico-comparativa: os casos Chiquitano e Jabutí. In: Ana Suelly Arruda Câmara Cabral, Aryon Dall’Igna Rodrigues, Jorge Domingues Lopes, Maria Risolêta Silva Julião (orgs.). *Línguas e Culturas Macro-Jê, vol. 2*: 105–120. Campinas: Curt Nimuendajú; Brasília: Laboratório de Línguas Indígenas.
- Ribeiro, Eduardo Rivail. 2012a. *A grammar of Karajá*. Ph.D. dissertation, University of Chicago.
- Ribeiro, Eduardo Rivail. 2012b. Final consonants in Maxakalí and their comparative status. *LIAMES* 12: 189–93. doi:10.20396/liames.v0i12.1489
- Ribeiro, Eduardo Rivail, Hein van der Voort. 2010. Nimuendajú was right: The inclusion of the Jabutí language family in the Macro-Jê stock. *International Journal of American Linguistics* 76(4): 517–570. doi:10.1086/658056
- Ribeiro, Maria de Jesus Pacheco. 2010. *Dicionário Sateré-Mawé/português*. MA thesis, Universidade Federal de Rondônia (Guajará-Mirim campus).
- Ribeiro, Michela Araújo. 2008. *Dicionário Djeoromitxi-português: registro da diversidade lingüística do povo Jabuti*. MA thesis, Universidade Federal de Rondônia (Guajará-Mirim campus).
- Ribeiro, Rosa Maria de Lima. 2008. *Dicionário Arikapu-português: registro de uma língua indígena amazônica*. MA thesis, Universidade Federal de Rondônia (Guajará-Mirim campus).
- Rocha, Ivan. 2011. *A estrutura argumental da língua Karitiana: desafios descritivos e teóricos*. MA thesis, Universidade de São Paulo. doi:10.11606/D.8.2011.tde-12092012-120027
- Rocha, Ivan. 2022. Interpretação temporal em orações não finitas em Karitiana: a contribuição do aspecto. In: Dionei Moreira Gomes, María Alejandra Regúnaga, Arthur Britta Scandelari (orgs.). *Diversidade lingüística na América: línguas ameríndias (v. 1)*: 229–263. Brasília: Editora UnB.

- Rodrigues, Aryon Dall’Igna. 1985. Evidence for Tupi–Carib relationships. In: Harriet E. Manelis Klein, Louisa R. Stark (eds.). *South American Indian languages: retrospect and prospect*: 371–404. Austin: University of Texas Press.
- Rodrigues, Aryon Dall’Igna. 1993. Uma hipótese sobre flexão de pessoa em Bororo. In: *Anais da 45ª Reunião Anual da Sociedade Brasileira para o Progresso da Ciência*: 505. Recife: Sociedade Brasileira para o Progresso da Ciências.
- Rodrigues, Aryon Dall’Igna. 1999. Macro-Jê. In: R. M. W. Dixon, Alexandra Y. Aikhenvald (eds.). *The Amazonian Languages*: 165–206. Cambridge University Press.
- Rodrigues, Aryon Dall’Igna. 2000. Ge–Pano–Carib x Jê–Tupí–Karíb: sobre relaciones prehistóricas en Sudamérica. In: Luis Miranda (org.). *Actas del I Congreso de Lenguas Indígenas de Sudamérica, tomo I*: 95–105. Lima: Universidad Ricardo Palma.
- Rodrigues, Aryon Dall’Igna. 2005. As vogais orais do Proto-Tupí. In: Aryon Dall’Igna Rodrigues, Ana Suelly Arruda Câmara Cabral (orgs.). *Novos estudos sobre línguas indígenas*: 35–46. Brasília: Editora UnB.
- Rodrigues, Aryon Dall’Igna. 2007. As consoantes do Proto-Tupí. In: Ana Suelly Arruda Câmara Cabral, Aryon Dall’Igna Rodrigues (orgs.). *Línguas e culturas Tupí, vol. I*: 167–203. Campinas: Curt Nimuendajú; Brasília: LALLI/UnB.
- Rodrigues, Aryon Dall’Igna. 2009. A case of affinity among Tupí, Karíb and Macro-Jê. *Revista Brasileira de Linguística Antropológica* 1(1): 139–167. doi:10.26512/rbla.v1i1.12289
- Rodrigues, Aryon Dall’Igna. 2012. Flexão relacional no tronco linguístico Macro-Jê. *Revista Brasileira de Linguística Antropológica* 4(2): 267–277. doi:10.26512/rbla.v4i2.20691
- Rodrigues, Aryon Dall’Igna, Ana Suelly Arruda Câmara Cabral. 2010. Explorando a hipótese de Rodrigues sobre possíveis conexões genéticas Tupí e Macro-Jê. In: Rosane de Sá Amado (org.). *Estudos em línguas e culturas Macro-Jê*: 143–153. São Paulo: Paulistana.
- Rodrigues, Aryon Dall’Igna, Ana Suelly Arruda Câmara Cabral, Beatriz Carretta Corrêa da Silva. 2006. Evidências lingüísticas para a reconstrução de um nominalizador de objeto \*-mi- em Proto-Tupí. *Estudos da Língua(gem)* 4(2): 21–39. doi:10.22481/el.v4i1.1021
- Rondon, Cândido M. S., João Barbosa de Faria. 1948. *Glossário Geral das tribos silvícolas de Mato-Grosso e outras da Amazônia e do Norte do Brasil, tomo I*. Rio de Janeiro: Imprensa Nacional. (Comissão Rondon, 76.)
- Sabino, Wary Kamaiurá. 2016. *Awetýza ti?íngatú: construindo uma gramática da língua Awetý, com contribuições para o conhecimento do seu desenvolvimento histórico*. Ph.D. dissertation, Universidade de Brasília.
- Salanova, Andrés Pablo. 2011. A flexão de terceira pessoa nas línguas Jê. *LIAMES* 11: 75–114. doi:10.20396/liames.v0i11.1497
- Santana, Áurea Cavalcante. 2012. *Línguas cruzadas, histórias que se mesclam: ações de documentação, valorização e fortalecimento da língua Chiquitano no Brasil*. Ph.D. dissertation, Universidade Federal de Goiás.
- Santos, Ludoviko Carnasciali dos. 1997. *Descrição de aspectos morfossintáticos da língua Suyá (Kîsêdjê), família Jê*. Ph.D. dissertation, Universidade Federal de Santa Catarina.
- Schleicher, Charles O. 1998. *Comparative and internal reconstruction of the Tupi–Guarani language family*. Ph.D. dissertation, University of Wisconsin.
- Sekelj, Tibor. 1948. *Wordlist Aruá-Makurap-Zaboti-Arikapú-Tupari*. Ms. Available online at: [etnolingüística.wdfiles.com/local--files/sekelj%3A1/Sekelj\\_Wordlist\\_Aru%C3%A1\\_Makurap\\_Zabot%C3%AD\\_Arikap%C3%B3\\_Tupar%C3%AD.pdf](http://etnolingüística.wdfiles.com/local--files/sekelj%3A1/Sekelj_Wordlist_Aru%C3%A1_Makurap_Zabot%C3%AD_Arikap%C3%B3_Tupar%C3%AD.pdf), accessed on 24/02/2023.
- Seki, Lucy. 2000. *Gramática do Kamaiurá, línguas Tupi–Guarani do Alto Xingu*. Campinas: Editora da Unicamp; São Paulo: Imprensa Oficial.
- Seki, Lucy. No date. *Vocabulário português–Botocudo*. Ms.
- Silva, Mário André Coelho. 2015. *A coda consonantal em Maxakalí*. MA thesis, Universidade Estadual de Campinas. doi:10.47749/T/UNICAMP.2015.945929
- Silva, Mário André Coelho. 2020. *Tikmũũg yĩ ax tinã xohi xi xahĩnãg. Sons e pedaços da língua Maxakalí: descrição da fonologia e morfologia de uma língua Macro-Jê*. Ph.D. dissertation, Universidade Federal de Minas Gerais.
- Silva, Mário André Coelho. Forthc. Propostas para a fonologia do Yaathê e do Pré-Yaathê. *Amerindia*.
- Silva, Mário André Coelho, Andrey Nikulin. 2021. Uma aproximação à fonologia e morfologia do Malalí. *Cadernos de Etnolingüística* 9(1): e090107.
- Silva, Raynice Geraldine Pereira da. 2010. *Estudo morfossintático da língua Sateré-Mawé*. Ph.D. dissertation, Universidade Estadual de Campinas. doi:10.47749/T/UNICAMP.2010.769939
- Singerman, Adam Roth. 2016. Nasal harmony and phonotactic well-formedness in Tupari. *International Journal of American Linguistics* 82(4): 453–485. doi:10.1086/688603

- Singerman, Adam Roth. 2018. *The morphosyntax of Tuparí, a Tupian language of the Brazilian Amazon*. Ph.D. dissertation, University of Chicago. doi:10.6082/ns2a-cv36
- Snethlage, Emil Heinrich. 2015. *Die Guaporé-Expedition (1933–1935): in Forschungstagebuch. Aus dem Nachlass herausgegeben von Rotger M. Snethlage, Alhard-Mauritz Snethlage und Gleice Mere*. Wien /Köln / Weimar: Böhlau Verlag. doi:10.7788/9783412218591
- Storto, Luciana. 2019. Kin terms in Karitiana and how they may contribute to the reconstruction of Proto-Tupian kin terminology. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 14(1): 121–130. doi:10.1590/1981.81222019000100008
- Tremaine, Sheila. 2007. *Dicionário Rikbaktsa–português, português–Rikbaktsa*. Cuiabá: Associação Internacional de Lingüística — SIL Brasil.
- Tripp, Robert. 1995. *Diccionario amarakaeri–castellano*. Lima: Ministerio de Educación/Instituto Lingüístico de Verano. (Serie Lingüística Peruana, 34.)
- Viegas Barros, J. Pedro. 2005. *Algumas semejanzas gramaticales Macro-Guaicurú/Macro-Jê*. Talk presented at IV Encontro de Pesquisadores de Línguas e Culturas Macro-Jê (November 3–5, 2005). Available online at [www.etnolingüística.org/artigo:viegas-barros-2005](http://www.etnolingüística.org/artigo:viegas-barros-2005), accessed on 24/02/2023.
- Voort, Hein van der. 2007. Proto-Jabutí: um primeiro passo na reconstrução da língua ancestral dos Arikapú e Djeoromití. *Boletim do Museu Paraense Emílio Goeldi. Ciências Humanas* 2(2): 133–168. doi:10.1590/S1981-81222007000200007
- Weiss, Helga E. 2005. *Dicionário Kayabi–português com um glossário português–Kayabi*. Brasília: Sociedade Internacional de Lingüística.
- Wiesemann, Ursula Gojtéj. 2011. *Dicionário Kaingang–português, português–Kaingang*. Curitiba: Esperança.
- Xerente, Armando Sôpré. 2019. *Particularidades dos sons, nomes, verbos, advérbios e posições em Akwẽ (Xerente), família Jê Central, tronco Macro-Jê*. MA thesis, Universidade de Brasília.

A. В. Никулин. Лексические свидетельства в пользу макро-же–тупийской гипотезы

Гипотеза о возможном дальнем родстве двух крупнейших языковых семей восточной Южной Америки — макро-же и тупийской — высказывается уже давно, в основном с опорой на морфологические сходения. В этой статье приводятся лексические свидетельства в пользу макро-же–тупийской гипотезы. При этом сравниваются именно праформы, восстановленные для пра-макро-же и пратупийского языков. Особое внимание уделено дистрибуции рефлексов рассматриваемых этимонов внутри каждой семьи, морфосинтаксическим свойствам сравниваемых форм, а также семантическому и фонологическому правдоподобию предлагаемых этимологий. Хотя количество возможных сходений не очень велико, между ними устанавливаются регулярные звуковые соответствия, что делает макро-же–тупийскую гипотезу привлекательной и достойной дальнейшего рассмотрения.

**Ключевые слова:** макро-же языки; тупийские языки; сравнительно-исторический метод; коренные языки Южной Америки.

## Studies in Yukaghir etymology II

This paper offers a number of additions and corrections to the corpus of etymologies published in Irina Nikolaeva's *A Historical Dictionary of Yukaghir* (De Gruyter, 2006). The focus of the paper is on internal Yukaghir etymology rather than on search for loanwords or long-range cognates.

**Keywords:** Yukaghir languages; etymology; historical linguistics; protolanguage reconstruction.

### Introduction

This is the second paper in the series opened with Zhivlov 2022b, whose aim is to propose new etymologies for words that were erroneously etymologized or not etymologized at all in Irina Nikolaeva's *A Historical Dictionary of Yukaghir* (HDY). Proto-Yukaghir and pre-Proto-Yukaghir forms in this paper are reconstructed following my own revised version of Nikolaeva's Yukaghir reconstruction (Zhivlov 2022a). Unless noted otherwise, Kolyma Yukaghir forms are cited after Prokop'eva & Prokop'eva 2021 and Tundra Yukaghir forms are cited after Kurilov 2001.

#### 1. K *aranpaaj* 'mushroom (growing on earth)'

This word is analyzed in HDY (155–156) as *a-ra:n-paj* 'mushroom (growing on earth)' (lit. 'naked woman'). This etymology implies that the word is a compound, whose second component is K *paaj* 'woman'. The first component is supposed to be the attributive form in *-n* of the Kolyma stem *aruo-*, attested in K *aruo-ńe-* 'to be naked; to be bald' (*-ńe-* is a proprietive suffix, deriving verbs with the meaning 'have X' from nouns, see Maslova 2003: 122–124). This stem, together with its Tundra cognate *araw* 'naked', is derived from Proto-Yukaghir *\*cawa* 'skin' (> T *sawa(η)* 'hide, skin') with the negative prefix *\*e-*, here regressively assimilated to the vowel of the root (Zhivlov 2022b: 73). The change of word-initial *\*c-* to *-r-* in compounding is regular. The etymology proposed in HDY faces both phonetic and semantic difficulties. From the phonetic point of view, *-a-* or *-aa-* instead of *-uo-* remains unexplained. Semantically, 'naked woman' looks quite arbitrary.

The clue to the correct understanding of this word lies in the fact that it designates specifically mushrooms growing on earth. As such, it is opposed to K *šaan paaj* 'mushroom, fungus (growing on trees)' (Nikolaeva & Šalugin 2002: 85, 113). The latter form is transparently 'tree woman', cf. already Jochelson (1926: 419): "The Yukaghir call mushrooms *can-pai*, i. e. tree-girl". K *aranpaaj* is thus to be understood as 'not-tree-woman'. The change of initial K *š-*, T *s-* (< PY *\*c-*) to intervocalic *-r-* in compounds is regular, see above on *araw* 'naked'.

This etymology provides us with one more example of the negative prefix *\*e-* (Zhivlov 2022b: 73–74), here harmonized to *a-*.

## 2. T *asle* ‘sibling born immediately after the speaker’<sup>1</sup>, T *asl’ii-* ‘to have somebody as a sibling born immediately after the speaker’

The sibilant T *-s-* ~ K *-š-* (< PY *\*-c-*) is quite rare word-internally in Yukaghir words, apart from some verbal suffixes (HDY: 83). In the preconsonantal position we find variation *-s/šL-* ~ *-rTL-* (where T is a voiceless stop and L is a lateral) in the following items: T *kösl’e(η)* ‘burbot’, S <kortle> ‘Quappe’ (Schiefner 1871: 379); K *marql’uo* ~ *martl’uo* ~ *mašl’uo* ‘daughter’ from K *marqil’* ‘girl’ and K *uo* ‘child’. Knowing this, we can propose that T *asle* goes back to an earlier form *\*arqle* and is related to K *arqaa* ‘near, at’ (postposition) (Maslova 2003: 268–269), T *arqaa lalime* ‘sledge immediately following the first in a caravan of reindeer sledges’ (see other cognate forms in HDY: 113). The semantic motivation here is obvious: both T *arqaa lalime* and T *asle* refer to something/someone immediately following another.

## 3. T <arugi-mer-ignu-> ‘stammer’

This phrase is attested only in Jochelson’s manuscript of the Tundra Yukaghir dictionary, entries from which are included in HDY (170). It is to be interpreted as *aruu-gi mer-ig-nu-* ‘his/her voice is tied’, where *aruu-gi* is T *aruu* ‘language; word; voice’ with the 3 person possessive suffix *-gi*, and *mer-ig-nu-* is a form of an otherwise unattested verb with the verbal focus prefix/proclitic *mer-*. This verb is derived from T *ige-* ‘to be tied; to stammer’<sup>2</sup> with the durative suffix *-nu-* (Schmalz 2013: 125–127). Thus, there is no need to reconstruct a separate root *\*iγn-*, supplied in HDY with a question mark. Altogether, the following four entries in HDY can now be united under a single Proto-Yukaghir root *\*ige-* ‘to tie’ (Zhivlov 2022a: 52): *\*iγ-* ‘to sew’ (HDY: 170), *\*iye-* ‘to catch’ (HDY: 170), *\*iγn-* ‘to stammer’, and *\*iγč’ə-* ‘to sew; sinew’ (HDY: 174–175).

## 4. T *čuoire-* ‘to catch fire; to appear (of burning embers)’

This verb is clearly related to T *čuoتهj-* ‘to let the fire flare up more strongly; to light (a cigarette, a pipe)’, although HDY: 143 lists the latter under a separate reconstruction *\*čö:tə-* and connects the former with K *čorqə* ‘glade, clearing; tundra’, K *čorqə-* ‘firm, hard’ (K forms cited after HDY: 141). Despite the Russian gloss ‘загореться; появиться горящему углю’ (‘to catch fire; to appear (of burning embers)’), the example sentence in Kurilov’s dictionary (2001: 570) shows that the subject of this verb is *lačil* ‘bonfire’: *Eld’e mit lačil el čuoتهj?* ‘Ну как, появился уголь [горящий] в костре?’ (‘Well, didn’t burning embers appear in our bonfire?’, literally ‘Well, didn’t our bonfire acquire burning embers?’). Thus, this verb behaves exactly like other verbs with the denominal suffix *-re-* ‘to acquire X’ (Schmalz 2013: 113). The verb T *čuoتهj-* ‘to let the fire flare up more strongly; to light (a cigarette, a pipe)’ apparently contains the denominal suffix *-te-* ‘to provide with X’ (ibid.) and the semelfactive suffix *-j-* (Schmalz 2013: 128–129). Cf. for the same combination of suffixes T *od’etej-* ‘to moisturize; to wet’ from T *od’e* ‘dew; wet

<sup>1</sup> The original Russian gloss in Kurilov (2001) is ‘единоутробный брат (или сестра), родившийся в семье непосредственно за говорящим’. Nikolaeva (HDY: 113) translates this as ‘half-sibling born immediately after the speaker’. While Russian *единоутробный* ‘born by the same mother’ is normally applied to half-siblings, a look at how this word is actually applied by Kurilov (2001: 58, 59, 239, 329, 335, 579, 604) shows that he uses it as a synonym of *родной* in the sense ‘natural sibling’, as opposed to cousins.

<sup>2</sup> In the latter meaning with *aruu* ‘voice’ as a subject.



place; drop; juice (of meat)'. The deriving noun \*čuo 'burning ember' is not attested as such, but is preserved as a second part of compound in T *lači-n+d'uo* 'burning ember' and perhaps in T *jerpeje+d'uo* 'disk of the sun'. Now, it seems quite possible that T *čuo(η)* 'iron', also attested in Omok as *če* <Тче> 'copper; iron' (Matjuškin 1841: 121), is etymologically the same word, whose meaning 'iron' developed from 'burning ember'.

### 5. T *elem* 'nothing'

This form functions as a preverbal particle in such expressions as T *elem-gurilijen* 'I don't know anything' (Krejnovič 1982: 213), T *elem-wietejen* 'I won't do anything' (ibid.), T *elemjuol* 'to see nothing', T *elemkuril'iil* 'to know nothing'. It is quite transparently derived from T *leme ~ neme* 'what', K *leme ~ neme* 'what' with the negative prefix *e-* (Zhivlov 2022b: 73–74). This derivation is not recognized by Nikolaeva, who reconstructs a separate etymon \*ölem (HDY: 324). The rest of the material adduced by Nikolaeva under \*ölem is also undoubtedly cognate. Note that K *ulum* 'mad', cited in HDY, is not found as an independent word in other sources, only as a part of the expression *ulum kude-* 'to go mad'<sup>3</sup>, literally 'to become nothing' (Prokop'eva & Prokop'eva 2021: 300; Nikolaeva & Šalugin 2002: 72; Maslova 2003: 342, 399). The verb K *ulum-l'ə* 'mad; stupid' (attested only in HDY: 324) is derived from K *ulum* with the inchoative suffix *-l'e-* 'become X' (Maslova 2003: 205). Note that variants such as K <elūlum-ku'de> 'to run mad; to go mad' (Jochelson 1926: 321), K <oļu'lum-k'ude> 'to become mad, insane' (Jochelson 1926: 325), K <o'ļlum-k'ude> 'to run, to go, become mad' (Jochelson 1926: 336) contain a pleonastic negative prefix. The derivation of K *ulum-l'ə* 'mad; stupid' from 'nothing' allows us to explain previously unetymologized verb T *leml'e-* 'to be tolerable, normal in terms of quality, in terms of the intensity of the manifestation of smth.; to feel healthy; to be not the timid type' as a parallel derivative from T *leme* 'what', literally 'to be/become something'. This verb also has a nominal correlate T *leml'e* 'chief, superior; the authorities'.

### 6. T *eluojerke* 'dry female reindeer'

This word is given in HDY (154) under the reconstructed root \**el-* 2 together with T *elmelije* 'a bare (without vegetation) area on a hill; a flat terrain without holes or willows', T *elmeliñe-* 'to be even, smooth (about a terrain overgrown only with grass)'. The sequence *el-* that these words have in common is actually the negative prefix *el-*. The obvious components of T *eluojerke* are the negative prefix and the word T *uo* 'child'. The suffix *-rke* looks like T *-rqal-rke* — a suffix deriving names of quality from qualitative verbs (Kurilov 1994: 43–49), although here its function is clearly different. Note the same suffix in a semantically close word T *ličuoerke* 'female reindeer'. The element *-je-* looks mysterious, since normally the suffix *-je* is deverbal (Kurilov 1994: 10–18). However, there are other similar examples with the suffix *-je*, whose approximate meaning can be described as 'having X', where X is a noun denoting living beings. The first is T *ńorquoje* 'female of a wild reindeer with a newborn calf; reindeer herd where there are only calving females', whose first two components are T *ńorqo-* 'newborn reindeer' (see below) and T *uo* 'child'. Another case is K <o'yē> 'father' (Jochelson 1926: 326), probably derived from K *uo* 'child', a cognate of T *uo* 'id.'. One more example is T *elñiimije* 'orphan'.

<sup>3</sup> Prokop'eva & Prokop'eva (2021: 300) also give a meaning 'to turn sour (of milk)', which developed directly from 'to become nothing'.

Here *el-* is the same negative prefix we see in T *eluojerke*. The remaining part *-ñiimi-* is not attested as a word for ‘mother’ or ‘parent’, but may be tentatively etymologized as consisting of the reciprocal prefix *ñi-*, sometimes used in kinship terms (Maslova 2007: 1854), and the otherwise unattested Tundra cognate of K *emej* ‘mother’. Summing up, *eluoje-* would have meant ‘having no child’ and *-rke* was possibly added under the influence of T *ličuoerke* ‘female reindeer’. Finally, K *olujorko*, given in HDY: 326 under the reconstruction *\*olujərqa*, is a hapax, attested as <олýjopko> in but one text (Jochelson 1900: 70). Although the Kolyma word is translated by Jochelson as ‘(wild reindeer) male’, it is almost certainly cognate to T *eluojerke*, and the meaning ‘male’ is most likely an error.

### 7. K *kenkeraa* ‘bucket’

This word is listed in HDY: 206 as *kenkə-ra*: ‘amber’ (sic!)<sup>4</sup> with the following commentary “[t]he cluster *-nk-* is atypical and indicates that the word may be a recent borrowing”. Indeed, a similar form is attested in Ewen (Kolyma-Omolon dialect) *kéŋka* ‘a big cauldron’ (TMS I: 450), but given the absence of Tungusic cognates, it is certainly itself borrowed from Yukaghir. The “suffix” *-ra*: derives names of various household utensils, such as ‘trough’, ‘rack’ or ‘pitch fork’ (Maslova 2003: 132–133). In fact, it is a postpound going back to K *šaa-l* ‘tree; stick; firewood’, with the regular change *š > r* between vowels. The atypical cluster *-nk-* points to a morphological boundary. It seems probable that the word contains the suffix *\*-rkə*, which is attached to qualitative verbs to form names of objects or concepts possessing the quality in question (see Zhivlov 2022a: 49–51). The deriving root is K *kenbe-*, attested in K *kenbune-* ‘wide’, K *kenbuben* ‘width’, K *kenbegedej-* ~ *kenmegedej-* ‘to open, unfold’ (tr.), K *kenbel’eš-* ‘to spread out, to lay out, to unfold; to flatten’. Thus, ‘bucket’ is a ‘wide/flattened thing’. The simplification of the cluster on the morphological boundary follows the general rule: only the first and the last consonants (in this case, *-n-* and *-k-*) are preserved.

### 8. T *kise-* ~ *kiise-*, K *kiše-* ~ *kišše-* ‘to show’

Comparison of T *kise-* ‘to show’ with K *kiše-* ‘id.’ leads to a straightforward reconstruction of PY *\*kice-*. This reconstruction, however, does not account for the variants T *kiise-* ‘to show’ and K *kišše-* ‘id.’. The geminate variant in Kolyma results from a recent syncope: Jochelson’s records show both the simple variant *kiše-* <кíшä-> and ‘long’ variants *kičše-* <кíчäшä-> (Jochelson 1900: 104) and *kiššē-* <кíшšäш-> (Jochelson 1900: 142). Nikolaeva explains this in the following way: “[s]ome forms demonstrate the following phonetic changes: *\*kičəse-* > *kičəš-* > *kiššə-* > *kiše-/kise-*” (HDY: 213). This scenario runs into chronological problems, since ‘simple’ variants *kiše-/kise-*, supposed to result from syncope, actually predate this syncope in Kolyma. Moreover, one of the variants (<кíчäшä-> and <кíшšäш->) must be secondary, and the variant with *-č-* can be tentatively explained by the influence of K *kiš-* ‘to teach’ (*š* is an allophone of *č* in Kolyma Yukaghir). The variant <кíшšäш-> can be derived from <кíшä-> with the pleonastically added causative suffix *-še-*.

It seems probable, however, that the PY form *\*kice-* itself contains the same causative suffix: words for ‘to show’ are typically causative formations in languages of Northern Eurasia (Uralic, Mongolic, Tungusic, Itelmen, Nivkh). Since the root shape CV- with a short vowel is impossible in Yukaghir, we must assume the loss of some consonant before the causative suf-

<sup>4</sup> The erroneous gloss possibly resulted from a confusion between English *amber* and Estonian *ämbەر* ‘bucket’.

fix already in Proto-Yukaghir. Fortunately, we are able to identify this consonant thanks to the unique forms K *kigie-m* ‘he showed’ and K *kigie-č-u-m* ‘he was showing’, cited by Krejnovič (1982: 129). Since there are no other attestations of this verb, and it is suspiciously similar to the well-attested K *kigie-* ‘to prick; to butt’, one would be tempted to doubt the authenticity of these forms. Support comes from T *kiise-*, which can owe its vowel length to a loss of pre-consonantal *-g-*, cf. T *iire-* ‘to tie’, derived from T *ige-* ‘to be tied up’.

Summing up, the Proto-Yukaghir verb *\*kice-* ‘to show’ was derived from a root *\*kig-* with the causative suffix *\*-ce-*. The velar stop was lost in this form already in the proto-language<sup>5</sup>. Later, in Tundra Yukaghir a renewed causative form was created by adding the suffix *-se-* to the root *\*kig-*. This latter form gave T *kiise-*, which coexists with T *kise-* — a direct reflex of PY *\*kice-*.

### 9. T *liidej-* (intr.) ‘to scatter; to disperse (of reindeer)’

This verb is adduced in HDY under the reconstructed form *\*lintə-*. The reconstruction with a short vowel here, as in some other words<sup>6</sup>, is apparently based on an assumption that the simplification of clusters “homorganic nasal + obstruent” caused compensatory lengthening of a preceding vowel. However, many counterexamples can be cited, e.g., *\*joŋq-* > K *joyul* ‘nose’, T *joyul* ‘nose’ (HDY: 196); *\*montə-* > K *modo-* ‘to sit’ (HDY: 276); *\*noŋqə-* > K *noyo* ‘sand’, T *noyo* ‘ash’ (HDY: 309); *\*nontə-* > K *nodo* ‘bird’, T *nada* ‘wolf’ (HDY: 309), etc. Thus, the long vowel must be original here. Still, Proto-Yukaghir does not have roots of the shape (C)VVCC(V)-, i.e. roots with a long vowel before a consonant cluster; thus, we can conclude that the verb *liidej-* has a morphological boundary before *-d-* (< *\*-nd-* < *\*-nt-*). This is confirmed by the stem T *liid’i-*, the iterative counterpart of *liidej-*, which can be analyzed as consisting of the root *lii-* and the iterative suffix *-d’i-* (Schmalz 2013: 129–133). The stem *liidej-*, then, is composed of the root *lii-*, the non-iterative suffix *-de-* (Maslova 2003: 192–193) and the semelfactive suffix *-j-* (Schmalz 2013: 128–129). The same root *lii-* with the causative suffix *-te-* (Schmalz 2013: 152–156) is also found in T *liite-* (tr.) ‘to share smth. with smb.; distribute, divide’. Another derivative from the root *lii-* is T *liitterej-* (tr.) ‘to separate (reindeer from the herd)’, which contains the augmentative suffix *-tte-* (Schmalz 2013: 164–165), non-iterative suffix *-re-* and semelfactive *-j-* (Schmalz 2013: 128–129). An iterative counterpart of T *liitterej-* is T *liittes-*, where *-s-* is a causative suffix (Schmalz 2013: 152–156). T *liite-* and T *liitterej-* are given in HDY: 245 under a separate reconstruction *\*li:tə-*.

Summing up, instead of two reconstructions proposed in HDY — *\*lintə-* and *\*li:tə-* — we can reconstruct the verbal root *\*lii-*. Note that this is not a proper Proto-Yukaghir reconstruction, since no cognates are found in Kolyma Yukaghir, Chuvan or Omok. Still, it is possible that this root was present in Proto-Yukaghir and was simply lost outside of TY.

### 10. K *lomd’e* ‘dew’

This word cannot be separated from K *löude-* ‘to take off; to drop’, semelfactive K *löudij-*<sup>7</sup> ‘to fall; to go down; to descend (of fog)’, cf. a *figura etymologica* in K *lomd’e löudiš* ‘dew has

<sup>5</sup> Loss of the velar stop before *\*-c-* becomes more understandable once we recognize the secondary nature of intervocalic *\*-c-* itself. Word-initial *\*c* regularly alternates with intervocalic *\*r*, which means that intervocalic *\*c* must have been a cluster or geminate on the pre-Proto-Yukaghir level.

<sup>6</sup> E.g., *\*ončə-* > K *o:žə-* ‘to drink’ (HDY: 330).

<sup>7</sup> Given as *löudiš-* in Prokop’eva & Prokop’eva 2021: 151. The same source gives inflected forms like 1 sg. *löudiše* (< *\*löudij-je*) and 3 sg. *löudiš* (< *\*löudij-j-*), which show that the underlying stem is *löudij-*.

fallen' (Prokop'eva & Prokop'eva 2021: 147). Nikolaeva reconstructs two roots, *\*l'omč'ə/\*lomjə* (HDY: 248) and *\*löw-* (HDY: 250). The first is supported by the following data from her own fieldwork: K *l'omd'ə* 'moisture, humidity', K *l'omd'ə-* 'to shed hair, feathers', K *lomd'u:-* 'humid, damp' (HDY: 248). The palatal *l'*- is confirmed by K *l'omd'oj* 'to shed hair, feathers' in Nikolaeva & Šalugin 2002: 39 and by *l'omd'e:j* 'fade, lose color'<sup>8</sup> in Maslova 2003: 548. On the other hand, Krejnovič (1982: 63) gives K *lōmd'e* 'moisture' with non-palatal *l-*, and examples in Maslova's grammar also have *l-* (Maslova 2003: 187, 509, 511). The variant with the palatal lateral *l'*- is possibly due to assimilation to the following *-d'*-. Anyway, it should be secondary, since Proto-Yukaghir word-initial *\*l'*- regularly gives *j-* in Kolyma Yukaghir. Nikolaeva's *\*löw-* is supported by the following reflexes from her fieldwork: K *löudu:-* 'to fall down', K *löudə-* 'to drop; to take off', K *lömdijə ~ jömdijə* 'precipice' (HDY: 250). Nikolaeva notes that "[t]he alternation *-w-* ~ *-m-* is irregular" (ibid.). In fact, the alternation is governed by the rule described in Zhivlov 2022a: 51–53: pre-Proto-Yukaghir clusters of the type "obstruent + nasal + obstruent" yield Proto-Yukaghir clusters "nasal + obstruent", where the nasal has the same place of articulation as the first obstruent in the pre-Proto-Yukaghir cluster. Thus, we can reconstruct a pre-Proto-Yukaghir root *\*lop- ~ \*löp-* 'to fall, to drop': K *lomd'e* 'dew' goes back to pre-PY *\*löp-ńčə* with the participle suffix *\*-ńčə*, while K *löude-* can be derived from pre-PY *\*löp-te-* with the non-iterative suffix pre-PY *\*-te-* > PY *\*-de-*, which did not contain a nasal. The original *\*p* is preserved before *-č-* in K *löpsii- ~ löpčii-* 'to drop; to shed (leaves); to take off, remove'.

### 11. T *ńaal'uol-* 'to enter into a sexual relationship'

This verb is tentatively (under a question sign) derived in HDY: 283 from T *ńaajl* 'son-in-law, daughter-in-law'. Apart from formal problems (*-l'*- vs. *-j-*), such a derivation looks utterly improbable in the Yukaghir cultural context, where certain relatives, including in-laws, were required to perform mutual avoidance: "Persons who are mutually "bashful" should not address themselves directly to each other, or look into each other's faces, or call each other by name or by their term of relationship. ... A daughter-in-law should not look into the face of her father-in-law or her husband's elder brother, neither is a son-in-law allowed to look into the face of his father-in-law or his mother-in-law" (Jochelson 1926: 76). Cf. the Tundra Yukaghir verb *ńaajči-* 'behave like a son-in-law or daughter-in-law; behave modestly, shyly, like a son-in-law or daughter-in-law; perform mutual avoidance norms towards smb.'.

Another etymology can be suggested for T *ńaal'uol-* 'to enter into a sexual relationship'. If we suppose that this verb contains a fossilized reciprocal prefix *ńi-*, we can derive it from T *al'uol-* 'to be melted', cf. T *al'aa-* 'to melt (of snow, ice); to warm oneself; to heat up (of a house); to dissolve'. In this case, the literal meaning of T *ńaal'uol-* would be 'to melt reciprocally' or simply 'to warm each other'.

### 12. T *ńaarčuu-* 'to be bad', attributive form *ńaarčə*

This Tundra verb and its various derivatives are compared in HDY to K *ńerčəd-anil'* 'buttocks' (HDY: 290), K *ńerčəd-anil'* <нәрчәд-әңил> 'vagina' (Jochelson 1900: 158), as well as to T *ńerče(ŋ)* 'buttocks', which HDY erroneously glosses as 'nasty'. The forms meaning 'buttocks' and 'va-

<sup>8</sup> This gloss is apparently a mistranslation of Russian *линять*, which means both 'shed hair or feathers' and 'fade, lose color (of textile etc.)'.

gina' go back to a separate Proto-Yukaghir form *\*ńerčə*, which, in view of a different vocalism and different semantics, has nothing to do with T *ńaarče*.

According to HDY, T *ńaarče* has “[a]n irregular long vowel in a closed syllable”. The irregularity in question is morphophonological rather than simply phonological — long vowels in closed syllables are perfectly possible both in Tundra and in Kolyma Yukaghir, but underlying roots cannot have long vowels in closed syllables. Therefore, T *ńaarče* should be morphologically segmentable. It is tempting to compare T *ńaarče* with K *erče* in K *erče ńińieńe-* ‘to be unkind, bad (of a person)’, *erče šoromo* ‘bad person, villain’, *erčed önmeń-* ‘bad (of a person), rude, quick-tempered’. The only way these words can be related is if the Tundra word includes a fossilized reciprocal prefix *ńi-*, the original meaning being ‘bad with respect to each other’. The modern Tundra Yukaghir reciprocal prefix has an allomorph *ńiń-* in prevocalic position, but this must be an innovation, cf. the cognate Kolyma Yukaghir reciprocal prefix *ńe-*, used both before vowels and consonants.

K *erče* is related to K *eruu-* ‘to be bad’, K *eríš-* ‘to be bad, unsatisfactory, ugly’, K *erulbe-* ‘to get worse (of mood, temper); to deteriorate, to become unusable’, and K *erie-* ‘to hate; to disdain’. The latter word has a Tundra cognate *erie-* ‘to disdain; to be disgusted by smth.; to reject’. Thus, we have a Proto-Yukaghir root *\*er-* ‘bad’ with a lot of derivatives, at least two of which go back to Proto-Yukaghir: PY *\*er-čə* (K *erče* ~ T *ńaarče*) and PY *\*er-ie-* (K *erie-* ~ T *erie-*). Now, Nikolaeva (HDY: 163) compares this root with Tungusic *\*er(e)-*. This reconstruction refers to the following forms: Ewenki *erū* ‘bad’, Solon *erū* ‘bad’, Ulcha *erule-* ‘to torment; to scold; to oppress’, Nanai *erū*, *erulē* ‘torment’, *erule-* ‘to torment’, Manchu *eru-*, *erula-*, *erule-* ‘to torment, to torture’, *erun* ‘torment, torture; execution’ (TMS II: 465–466). Doerfer (1985: 39) considers all these words to be borrowed from Mongolic, cf. Written Mongolian *eregüü* ‘torture, torment, chastisement; capital punishment’ (Lessing et al. 1960: 321–322). There is no doubt that Ulcha, Nanai and Manchu forms are Mongolic loans, since their semantics is the same as that of the Mongolic original. Things are less clear with Ewenki and Solon forms. If these are also borrowed from Mongolic, we have to assume that 1) the meaning shift ‘torment, torture’ > ‘bad’ occurred in Proto-Ewenic and that 2a) either this word was borrowed at an early stage into Yukaghir languages, where the final *-ū* was reinterpreted as a Yukaghir suffix, after which other derivatives were formed from the bare root *\*er-*, or 2b) the resemblance with Yukaghir is due to chance. Both options do not look very probable. The most simple solution in my view is to separate the Ewenic forms from the Ulcha, Nanai and Manchu ones, and to consider the former borrowed from Yukaghir and the latter borrowed from Mongolic.

### 13. T *ńamne-* ‘to be wide (of a distance between the horns of a reindeer)’

This verb, given in HDY: 287 under the separate protoform *\*ńamnə-*, contains the suffix of qualitative verbs *-ne-/na-* (Schmalz 2013: 112). Verbs with this suffix typically have active intransitive counterparts with the suffix *-gej-/γaj-*: T *ńori-ne-* ‘to be yellow’ ~ T *ńoro-γoj-* ‘to turn yellow’, T *sil-na-* ‘to be dry’ ~ T *silγa-γaj-* ‘to dry up quickly’, T *wen-ne-* ‘to be stretched’ ~ T *wede-gej-* ‘to stretch out, to expand’, etc. The active counterpart of T *ńam-ne-* ‘to be wide (of a distance between the horns of a reindeer)’ is T *ńaba-γaj-* ‘to open (of door, etc.), to widen (of an opening)’ (intr.), for which HDY: 287 reconstructs *\*ńampə-*. Other words from the same root are T *ńabal’es-* ‘to open (of door, etc.)’ (tr.), T *ńabačėń-* ‘to be excessively open (about the smoke hole of the yaranga); be wide open; be wide (about horns sticking out in different directions)’. Moreover, Kolyma cognates can be proposed: K *ńabol’e-* ‘to be unbuttoned, to have unbuttoned clothes’, K *ńabol’uot* ‘being in a position with the head thrown back’. The Proto-Yukaghir root can be reconstructed as *\*ńambə-*.

#### 14. T *ńarqajewlid'e* ‘newborn reindeer’, T *ńorqomayil* ‘coat made of newborn reindeer skin’

These words are compounds with the second components T *jewlid'e* ‘reindeer calf’ and T *mayil* ‘coat’ respectively. One more related item is T *ńorqoje* ‘female of a wild reindeer with a newborn calf; reindeer herd where there are only calving females’ (see above on the morphological analysis of this word). The variable vocalism of *ńarqa-* ~ *ńorqo-* results from a relatively recent change *\*o > a* in Tundra Yukaghir, whose exact conditions so far remain unclear (HDY: 58–59). Thus, the original form must have been *ńorqo-*. Nikolaeva compares this word with Tungusic *\*ńar-gu-* ‘new, fresh’, noting that “[t]he element -qə in Yukaghir may be a derivational suffix”. A native Yukaghir etymology can be proposed for T *ńarqa-* ~ *ńorqo-*. We can start with the similarity of this stem with T *ńorqo* ‘copper’. Of course, the name for ‘newborn reindeer’ cannot go back to ‘copper’; instead, both are semantic specializations of a more general meaning. T *ńorqo* ‘copper’ goes back to pre-PY *\*ńor-rkə* from the root of T *ńori-ne-* ‘to be yellow’ and T *ńoro-γoj-* ‘to turn yellow’ (Zhivlov 2022a: 50). The suffix *\*-rkə* is attached to qualitative verbs to form names of objects or concepts possessing the quality in question (Kurilov 1994: 43–49; Schmalz 2013: 106). Thus, the original meaning of T *ńorqo* was ‘yellow thing’. The shift from ‘yellow thing’ to ‘copper’ is trivial. The connection of ‘newborn reindeer’ to ‘yellow’ is found in Tundra Nenets, where the word *tas<sup>o</sup>ko* <таско> ‘newborn reindeer’ is apparently related to *tasyey<sup>o</sup>* <тасей> ‘yellow’ and *tasyexey<sup>o</sup>* <тасахэй> ‘id.’ (Tereščenko 1965: 639–640). Moreover, the same semantic connection is found in English, where *fawn* has a secondary meaning ‘pale brown colour tinted with yellow’.

#### 15. T *ńimojie-* ‘to mix blood (though marriage)’

HDY: 301 lists this verb under a reconstructed form *\*ni:mə* together with T *elńiimije* ‘orphan’ (on which see above) and K <ni’mdietek> ‘mother’s younger sister, her younger female cousins’ (Jochelson 1900: 239; Jochelson 1926: 71). In fact, T *ńimojie-* is derived from T *mojie-* ‘to touch; to stroke; to smear; to mix; to stir up; to rub oneself’ with the reciprocal prefix *ńi-*. Cf. Russian glosses: *mojie-* ‘смешать’ and *ńimojie-* ‘смешать(ся) (по крови)’, where Russian *-ся* is equivalent to Yukaghir *ńi-*.

#### 16. T *ńiruon* ‘separately; apart’, T *ńiruose-* ‘to part (with someone); to get divorced’

These words are given in HDY: 317–318 under the reconstruction *\*ńyr-*, with a comment that “[t]he stem shows back vowel harmony”. It seems reasonable to assume that *ńi-* is a reciprocal prefix here. Intervocalic *-r-* can regularly go back to *s-*. If this assumption is correct, the words in question are related to T *suose-* ‘to miss the target’, iterative T *suorič-*. Thus, ‘to part with someone’ is ‘to reciprocally miss the aim’. The postposition T *suode* ‘except, apart from’ can also be related.

#### 17. T *ńiwien* ‘different’

This form (Krejnovič 1958: 276) is given in HDY: 304 as *niwie-* ‘different’ (with erroneous *n-*) under a reconstruction *\*niw-* together with unglossed forms *niwol*, *niuol* from Jochelson’s unpublished Tundra Yukaghir dictionary. The etymology of T *ńiwien* could not be more evident:

it is simply T *wien* ‘another, other’ with the reciprocal prefix *ńi-*. As for T *niwol* and *niuol*, I would tentatively suggest that these forms are the same as T *ńiwal* ‘next to each other’ from T *wal* ‘near’ with the same reciprocal prefix. Of course, the absence of a gloss next to a form in HDY means, in theory, that its gloss must be the same as that of the preceding word, but in practice this is not always the case (Zhivlov 2022b: 71–72).

### 18. T <niuoletile-, niwoletile-> ‘to change’

These forms from the unpublished dictionary of Jochelson are given in HDY: 318 under the reconstruction *\*nywolə-*, supplied with a question mark. In fact, this word is attested in Kurilov’s 2001 dictionary as T *ńiwal’itiil’e-* ‘to do in exchange for something’. This verb is derived from T *ńiwal’itii-* ‘to exchange’, which in its turn is derived from T *ńiwal* ‘next to each other’, on which see above.

### 19. K *ńuoduope* ‘descendants’

This word, attested only in Jochelson’s materials (Jochelson 1900: 112, 114), is treated in HDY: 305 as etymologically isolated under a reconstructed form *\*ńo:δ-/ńo:nt-*. The final *-pe* is a plural suffix. The stem K *ńuoduo-* is hard to separate from T *uoduo* ‘grandchild’. The latter form can be easily analyzed as *uo-d+uo* ‘child’s child’, from T *uo(η)* ‘child’ with the attributive suffix *-d*. The initial *ń-* in the Kolyma form needs an explanation. The only formally possible hypothesis is that *ń-* here is the reciprocal prefix. While its semantic function here is not entirely clear, it is attested in other kinship terms, although only in terms denoting relations within a generation (Maslova 2007: 1854).

### 20. K *poyožil* ‘knee’

The comparison to K *poyoŋin* ‘knee-long’ (HDY: 354), where *-ŋin* is a dative case ending, implies that the root here is *poyo-*. While *-l* can be a suffix<sup>9</sup>, there is no nominal suffix *-ž* in Kolyma Yukaghir. Therefore, the most probable hypothesis is that we are dealing here with a compound. The second part of this compound can be identified as K *qožile* ‘cavity’ (Nikolaeva & Šalugin 2002: 76). The semantics make sense if we assume that the original meaning of the compound was ‘knee pit’, cf. also K *qoži-d+elbe* ‘armpit’ (Nikolaeva & Šalugin 2002: 76). In order to get the attested form, we need to assume a haplology: *\*poyo+yožil > poyožil*.

### 21. T *qaadale* ‘armpit’

This word is reconstructed in HDY: 379 as *\*qantala*. The reconstruction with a short vowel cannot be correct (see above on T *liidej-*). The word can be analyzed as a compound *qaa-d+ale*, where *-d* is the attributive suffix and *-ale* is related to the postposition T *al*, K *aal* ‘below’. The preposition *qaa-* is most likely related to K *qaar* ‘hide; bark (of a tree); skin’. Note that the *-r* in K *qaar* must be a suffix. Krejnovič (1982: 87) lists the following examples when *-r* in this word gets

<sup>9</sup> Cf. K *poyožaaq* ‘on one’s knees’ (Nikolaeva & Šalugin 2002: 60).

ousted by the attributive suffix: K *qa-n+punbur* ‘bed’ (K *ponbur* ‘sleeping mat’), K *qa-n+murudu* ‘fur stockings’ (K *murudu* ‘stockings’), K *qa-d+ejmunde* ‘half of a skin’ (K *ejmunde* ‘half’). The same prepound *qaa-* can be found in two more words. One is K *qa-d+uo* ‘trousers with fur inside’, listed in HDY: 373 as *qa:d-o*: ‘leather trousers with fur inside’, with vowel length in the first syllable. The second part of the compound here is K *oo* ‘trousers’. Another example is K *qa-n+šogi* <xancogi> ‘leather bag’ from the manuscript dictionary of Jochelson (HDY: 378). This word literally means ‘leather bag’ and has K *šögi* ‘bag, sack’ as its second part. The reconstruction *\*qansə-*, provided with a question sign in HDY, can be safely discarded.

K *qaar* is given in HDY: 379 under a reconstruction *\*qa:r/\*qajr*, where the second variant is based on T *qajr* ‘skin from the head of an animal’ and MK *-cháär-*, *-hér-*<sup>10</sup> (unglossed in HDY). In fact, Kurilov’s dictionary (Kurilov 2001), which is supposed to be the source of all material marked as T in HDY, does not have the word *qajr* ‘skin from the head of an animal’. It only has T *ńaačayajr* ‘skin from the muzzle of an animal; facial skin’ — a compound with T *ńaače(η)* ‘face’ as the first member. It is not yet clear how the variants with *aa* and *aj* are to be explained.

## 22. T *saayare* ‘side (the space located on the sides, edges of something; not the middle); the left side of yaranga; South side’, T *saayar* ‘South’

HDY lists T *saayare* with an erroneous gloss ‘left side of a yurt; West’ under a separate reconstruction *\*sa:γər/\*sanqər*. T *-yar* is a suffix of spatial adverbs (Schmalz 2013: 203). Thus, the root here is *saa-*. It can be easily identified as the root of T *saa-l* ‘tree; forest, taiga’. For the Tundra Yukaghirs the south side is obviously the side of taiga. Cf. T *čawlaayar* ‘north’, literally ‘sea side’, from T *čawul* ‘sea, ocean’.

## 23. T *sebul* ‘tray for food (made from reindeer skins sewn to each other)’

This word is compared in HDY: 401 to K *šepid-i-čə* ‘top of a mountain’ and K *šibil’, šebil’* ‘window; door’. The comparison is both semantically and phonetically implausible. Krejnovič (1982: 89) glosses T *sebul* as ‘mat for things’. The Tundra word is also found as the second part of the compound T *lugu+rubul* ‘tray; mat for meat (reindeer skins sewn to each other)’, whose first part is T *legu-l* ‘food’. Note the sound change *\*e-u > u-u* in both parts of the compound (Zhivlov 2022b: 78). HDY lists this word under a separate protoform *\*luyur-/\*luŋkur*<sup>11</sup> (HDY: 252). Another compound with the same second component is T *juödu+rubul* ‘a board for cutting fish or meat’. Its first part is T *juödu-* ‘to chop’.

T *sebul* has a rather transparent Kolyma cognate: K *šobul* ‘bedding (from plant material, often from larch branches or grass)’. The Proto-Yukaghir form can be tentatively reconstructed as *\*cembul*. Nikolaeva (HDY: 404) gives the Kolyma form as K *šöbul, šubul* ‘branch of the larch tree; bedding made of larch branches’ and compares it with T *sebul* from Krejnovič’s work, but

<sup>10</sup> This is one more Old Tundra form in the MK wordlist. In Zhivlov 2022b: 72 I suggested that the presence of both Tundra and Kolyma forms in the MK and MU wordlists results from borrowing. Now I think that these lists simply are a mix of words from two Old Yukaghir languages – either because of code-switching by bilingual informants, or because each list has words from more than one informant. Unfortunately, we do not have a “pure” Old Tundra wordlist.

<sup>11</sup> Apart from T *lugurubul*, this entry also includes T *lugumul* from Jochelson’s unpublished dictionary. Since this word is not glossed, it must presumably have the same meaning. This is doubtful, since it is phonologically identical to T *lugumul* ‘aging, old age’.



not with T *sebul* from Kurilov's dictionary, although these are two attestations of the same word. She further compares these words to Northern Tungusic *\*seg-/\*sew-* 'to lay branches in a yurt'. This comparison, although semantically attractive, is made impossible by the Tundra Yukaghir forms, which require the reconstruction of word-internal cluster *\*-mb-*.

#### 24. T *unumed'uo* ~ *unemed'uo* 'earrings'

This word is explained in HDY: 444 as "T *unemed'-uo* ear-ring [lit. ear's child]" — a compound of T *unume(η)* 'ear' and *uo(η)* 'child'. Kurilov (2001: 486) gives only the variant *unumed'uo* with the second syllable *-u-* in the main entry, but the variant *unemed'uo* with the second syllable *-e-* is attested in example sentences (Kurilov 2001: 108, 174, 226), as well as in the compound T *moηo-d+unemed'uo* 'pendants of a fur cap' (T *moηo(η)* 'headdress'). The variation in the second syllable vowel is the same as in the word for 'ear': T *unume(η)* ~ *uneme(η)* (Krejnovič 1958: 279). The morphological segmentation proposed in HDY implies that *-d'* must somehow be a variant of the attributive suffix *-d*. There are no other instances where the attributive suffix takes the form *-d'*, and no known morphophonological process which could have caused such a change. Note, however, that the Kolyma Yukaghir word for 'earrings' is *unume ludul*, literally 'ear iron' (K *ludul* 'iron'). The Tundra Yukaghir word has the same semantic structure: T *unume+d'uo*, literally 'ear iron', cf. T *čuo(η)* 'iron'. The voicing *č > d'* is regular in compounds, cf. T *čuul* 'meat', but T *al'γα+d'uul* 'boiled fish meat as a dish' (T *al'γα(η)* 'fish'). There was also a parallel form without voicing, attested as T <u'nemečō> 'earring' (Jochelson 1926: 327) and preserved in two verbal derivatives in modern Tundra Yukaghir: T *unumečuóne-* 'to have pendants, earrings', T *unumečuore-* 'to acquire earrings'.

#### Language Abbreviations

- K Kolyma Yukaghir (Prokop'eva & Prokop'eva 2021)
- MK Old Kolyma Yukaghir materials of Mueller/Lindenau (HDY)
- MU Old Ust'-Jansk Yukaghir materials of Mueller/Lindenau (HDY)
- S Old Kolyma Yukaghir materials of Suworov (Schiefner 1871)
- T Tundra Yukaghir (Kurilov 2001)

#### References

- Doerfer, Gerhard. 1985. *Mongolo-Tungusica. (Tungusica 3.)* Wiesbaden: Otto Harrassowitz.
- HDY = Nikolaeva, Irina. 2006. *A Historical Dictionary of Yukaghir. (Trends in Linguistics, Documentation 25.)* Berlin: Mouton de Gruyter.
- Jochelson, Waldemar. 1900. *Materialy po izučeniju jukagirskago jazyka i fol'klora. I. (Trudy jakutskoj èkspedicii, snar'ažennoj na sredstva I. M. Sibir'akova. Otděl III. Tom IX. Čast' III.)* Sankt-Peterburg: Imperatorskaja Akademija nauk.
- Jochelson, Waldemar. 1926. *The Yukaghir and the Yukaghirized Tungus. (Memoir of the American Museum of Natural History IX.)* New York: G. E. Stechert.
- Krejnovič, Eruxim A. 1958. *Jukagirskij jazyk*. Moskva: Izdatel'stvo Akademii nauk SSSR.
- Krejnovič, Eruxim A. 1982. *Issledovanija i materialy po jukagirskomu jazyku*. Leningrad: Nauka.
- Kurilov, Gavril N. 1994. *Obrazovanie imennyx slov v jukagirskom jazyke*. Yakutsk: Yakutskij naučnyj centr SO RAN.
- Kurilov, Gavril N. 2001. *Jukagirsko-russkij slovar'*. Novosibirsk: Nauka.
- Lessing, Ferdinand D. et al. 1960. *Mongolian-English dictionary*. Berkeley / Los Angeles: University of California Press.

- Maslova, Elena. 2003. *A Grammar of Kolyma Yukaghir*. (Mouton Grammar Library 27.) Berlin: Mouton de Gruyter.
- Maslova, Elena. 2007. Reciprocals in Yukaghir languages. In: Vladimir P. Nedjalkov (ed.). *Reciprocal Constructions (Typological Studies in Language 71): 1835–1863*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Matjuškin, Fedor F. 1841. Sобрание слов Чуванского и Омокского языков, составленное Мишманом Матюшкиным. In: *Pribavlenija k Putešestvoju po severnym beregam Sibiri i po Ledovitomu morju, soveršenomu v 1820, 1821, 1822, 1823 i 1824 g. èkspedicieju, sostojavšeju pod načal'stvom flota lejtenanta Ferdinanda fon-Vrangelja*: 115–125. Sankt-Peterburg: Imperatorskaja Akademija nauk.
- Nikolaeva, Irina A., Vasilij G. Šalugin. *Slovar' jukagirsko-russkij i russko-jukagirskij (Verxnekolymskij dialekt)*. Sankt-Peterburg: Drofa.
- Prokop'eva, Praskov'ja. E., Aleksandra E. Prokop'eva. 2021. *Jukagirsko-russkij slovar' (jazyk lesnyx jukagirov)*. Novosibirsk: Nauka.
- Schiefner, Anton. 1871. Beiträge zur Kenntniss der jukagirischen Sprache. *Bulletin de l'Académie Impériale des Sciences de Saint-Pétersbourg* XVI: 373–399.
- Schmalz, Mark. 2013. *Aspects of the Grammar of Tundra Yukaghir*. PhD dissertation, University of Amsterdam.
- Tereščenko, Natalija M. 1965. *Nenecko-russkij slovar'*. Moskva: Sovetskaja ènciklopedija.
- TMS I = Cincius, Vera I. (ed.). 1975. *Sravnitel'nyj slovar' tunguso-man'čžurskix jazykov. Tom I*. Leningrad: Nauka.
- TMS II = Cincius, Vera I. (ed.). 1977. *Sravnitel'nyj slovar' tunguso-man'čžurskix jazykov. Tom II*. Leningrad: Nauka.
- Zhivlov, Mikhail. 2022a. Pre-Proto-Yukaghir Consonant Clusters. *International Journal of Eurasian Linguistics* 4: 41–59.
- Zhivlov, Mikhail. 2022b. Studies in Yukaghir etymology I. *Journal of Language Relationship* 20(1): 71–80.

М. А. Живлов. Исследования в области юкагирской этимологии II

В настоящей статье предлагается ряд дополнений и уточнений к корпусу этимологий, опубликованному И. А. Николаевой в *A Historical Dictionary of Yukaghir* (De Gruyter, 2006). Основной упор сделан не столько на поиске заимствований или когнатов в рамках дальнего сравнения, сколько на собственно внутриякагирских этимологиях.

*Ключевые слова:* юкагирские языки; этимология; сравнительно-историческое языкознание; праязыковая реконструкция.

**Issue 2 / Часть 2**

**Articles / Статьи**



## Hadza as Afrasian?<sup>1</sup>

In this paper, I address the issue of the genetic affiliation of Hadza, the language of a tiny tribe of Tanzanian hunter-gatherers, genetically separated from other groups of *Homo sapiens* by some hundred thousand years and occupying the same area for over 50,000 years; the language was formerly considered to be Khoisan because of its click-containing phonetic inventory, but is now regarded as an isolate. The paper provides parallels from various Afrasian (Afro-Asiatic) languages, tied together through regular consonant correspondences and fairly strict semantic criteria, to the extensive Hadza lexical material collected by American linguists with the help of their Hadza collaborators; the parallels are drawn from the standard 100-item Swadesh wordlist (including the 50-item subset of the most stable items based on the selection of Sergei and George Starostin) as well as from other semantic groups. The author analyzes other explanations for these matches (such as accidental look-alikes; borrowings into Hadza from neighboring and even geographically distant Afrasian languages; common substrate), but concludes that the most plausible explanation is genetic affinity. The position of Hadza within the Afrasian super-family is, according to lexicostatistics, more or less equally close to the Omotic and Cushitic families; glottochronology dates the separation between Proto-Hadza, Proto-Cushitic and Proto-Omotic to the turn of the 10-9<sup>th</sup> millennia BCE when a group of Afrasian speakers presumably made it to Northern Tanzania and passed on their language to the (presumably) formerly Khoisan-speaking Hadza ancestors.

*Keywords:* Hadza language; Afrasian languages; genetic affinity; sound correspondences; lexicostatistics; etymology; lexical borrowing.

The Hadza people, a group of Tanzanian hunter-gatherers, are one of the most enigmatic peoples in the world, genetically separated from other early *Homo sapiens* groups by about a hundred thousand years (Tishkoff et al. 2007). They live close to the Olduvai Gorge, sometimes called the “Cradle of Mankind”, and have occupied the same area at least since the beginning of the Later Stone Age, 50,000 years ago or perhaps even longer. Exposed to powerful influences from different cultures, confirmed historically, genetically and, what is particularly important for our research, linguistically through many lexical borrowings, the Hadza have somehow held on to their traditional way of life, preserving their astounding singularity in the midst of a rapidly changing world. Only recently their culture, language and identity have become endangered.

Their language used to be considered Khoisan because of the presence of “click” phonemes (Greenberg 1966). However, lexicostatistical analysis (Starostin 2013) did not confirm the kinship with the Khoisan languages; for the most part, the language is today considered to be an isolate<sup>2</sup>.

---

<sup>1</sup> In memory of my untimely departed beloved son Mikhail Militarev (Jan. 18, 2005 – Nov. 21, 2022) who not only helped me with this investigation but whose deeply personal worry over every language and people that is endangered has moved me to this study and whose essay (Militarev 2021) introduced me to the phenomenon of the Hadza people.

This research has been supported by the Russian Science Foundation (Project No. 20-18-00159) with the Institute of Linguistics of the Russian Academy of Sciences as the financing organization.

<sup>2</sup> One should not pay serious attention to ludicrous assumptions about its status as a “first language”, even if published in a highly respectable scientific journal (Pennisi 2004).

The assumption of Afro-Asiatic affinity was first expressed by Derek Elderkin (1982) and cautiously supported by George Starostin (2008)<sup>3</sup>. Some of the lexical matches suggested by Elderkin and Starostin are convincing, but the proposed parallels with non-obvious consonant correspondences (such as Hadza *w* vs. \**f* or Hadza *h* vs. \**p*) are not confirmed by my analysis. However, further research by the author relying on the updated Afrasian etymological database by Militarev and Stolbova and the Hadza lexicon by Kirk Miller et al. (2021)<sup>4</sup> has revealed several hundred lexical matches between Hadza and all the Afrasian branches, showing regular and unsophisticated “one-to-one” correspondences in root consonants. Curiously, it turned out that, other than the presence of nine clicks and such secondarily developed traits as weak opposition of the dentals, lack of distinction between *l* and *r* and between *h* and *h̥*, lack of post-velars (uvulars) and a few other nuances (such as pre-nasalized plosives and affricates in Hadza not postulated for Proto-Afrasian, and labialized velars in Hadza whose PAA status is disputed), the Hadza system of consonants (Table 2) is very similar to the one that was first outlined for Proto-Afrasian by Dolgopolsky (1973), further elaborated in the late 20th century by Igor Diakonoff and his team<sup>5</sup> and, more recently, amended by several leading Afrasianists (Table 1):

|           |          |           |    |    |
|-----------|----------|-----------|----|----|
| *ḗ (?)    | *p       | *b        | *f | *m |
| *ṭ        | *t       | *d        |    | *n |
| *ḥ [tsʰ]  | *c [ts]  | *ʒ [dz]   | *s |    |
| *ḥ̥ [tʃʰ] | *č [tʃ]  | *ʒ̣ [dʒ̣] | *š | *r |
| *ḥ̣ [tʃʰ] | *ĉ [tʃ]  | *ṣ̌ [sʃ]  |    | *l |
| *ḥ̣̥      | *k       | *g        |    | *y |
| *ḥ̣̥ʷ (?) | *kʷ (?)  | *gʷ (?)   |    | *w |
| *ḥ̣̥ʰ (?) | *kʰ (?)  | *gʰ (?)   |    |    |
|           | *ḥ̣      | *ḡ        |    |    |
|           | *ḥ̣ʷ (?) | *ḡʷ (?)   |    |    |
|           | *ḥ̣      | *ḡ        |    |    |
|           | *ḥ̣      | *ʔ        |    |    |

Table 1. The Proto-Afrasian consonantal system.

<sup>3</sup> “It seems that the Afro-Asiatic hypothesis for Hadza is no less, and perhaps even more likely, than the Khoisan one... None of this means that Hadza should be automatically assigned to the Omotic group; to do this, it is necessary to have a more complete corpus of etymologies with established phonetic correspondences... In order to somehow really move towards clarifying the “Hadza-Afro-Asiatic” problem ... it is necessary to conduct preliminary lexicostatistical counts with other branches of the Afro-Asiatic family” (Starostin 2008; translated from Russian). It might be claimed that all these perfectly fair conditions have been fulfilled by the present author.

<sup>4</sup> To the best of my knowledge, the dictionary still remains in the status of unpublished manuscript since it has not been yet tone-marked.

<sup>5</sup> Diakonoff et al. 1992; cf. also Takács 1999: 266–270. A specific feature of the Proto-Afrasian consonant system, which is not preserved in most daughter languages, but best explains the consonantal peculiarities of them all, is the pattern of three local consonant series each consisting of four consonants, represented by triads of a voiced, voiceless and glottalized / “emphatic” affricate (with the lacuna of a voiced affricate in the lateral triad) and a corresponding sibilant. The system is practically identical with that of Hadza; it goes without saying that, while reconstructing the Proto-Afrasian consonantal system, we had no idea about the phonology of Hadza.

|                 |                |                |         |                  |
|-----------------|----------------|----------------|---------|------------------|
| p <sup>h</sup>  | p              | b              | (p')    | f                |
| mp <sup>h</sup> | mb             |                |         |                  |
| t <sup>h</sup>  | t              | d              |         |                  |
| nt <sup>h</sup> |                | nd             |         |                  |
|                 | c [ts]         | ɟ [dz]         | ç [ts'] | s                |
|                 | nc             | nɟ             |         |                  |
|                 | č [tš]         | ǰ [dž]         | č [tš'] | š                |
|                 |                | nǰ             |         |                  |
|                 | ĉ [tl]         | ĉ [tl']        |         | ŝ [sl]           |
| k <sup>h</sup>  | k              | g              |         |                  |
| ŋk <sup>h</sup> |                | ŋg             |         |                  |
| k <sup>hw</sup> | k <sup>w</sup> | g <sup>w</sup> |         |                  |
| h/h̥            | w              | y              |         |                  |
| l/r             | m              | n              | ŋ       | ŋ <sup>w</sup> ʔ |

Table 2. Hadza consonants (elicited from Sands, Maddieson, & Ladefoged 1996: 174).

Matches between lexemes from various semantic fields could be accounted for by borrowing into Hadza (cf., for example, Elderkin 1978), chance coincidence, or genetic relationship. However, mass coincidence or mass borrowing from diverse and geographically distant Afrasian languages by the Hadza, known for their permanent presence in the same area, do not *a priori* seem like a realistic solution. To test the hypothesis of kinship, it would be required to apply the lexicostatistical method, in the process of slowly transforming from 'controversial' into 'semi-accepted' by the linguistic mainstream<sup>6</sup>. It has been applied using the Swadesh 100-word list compiled for Hadza by Kirk Miller (Miller 2021), also taking into account the annotated list by George Starostin (Starostin 2012), as well as the 50-item wordlist of the universally most stable words (Starostin 2010) as applied to the Afrasian Lexicostatistic Database, composed by the author and connected to the Afrasian Etymological Database which allows to etymologize the Afroasiatic lexemes from the lists.

The average percentages of matches obtained with languages across all Afrasian branches both for the 50- and the 100-item wordlists (see below) do not produce the impression of random numbers. In particular, the subsystem of most stable personal pronouns ('I' — 'we' — 'thou'), a solid marker of linguistic kinship upon first approximation, is practically identical with the Common Afrasian system, all but precluding the scenario of borrowings to Hadza from various Afrasian languages (although some such cases are also elicited or tentatively supposed) or by a common substrate, to say nothing of random coincidences. The least controversial explanation is that "proto-Hadza" is an early offshoot of the Afrasian macrofamily in Africa, superimposed on an unknown (most likely Khoisan) substrate.

### Hadza 100-item wordlist: selected items with AA matches

*Note:* in accordance with the rules of the Swadesh-Starostin method in lexicostatistics, only those words, both in Hadza and any Afrasian language, whose meaning fully coincides with the meaning of the correlated item on the list are scored as positive matches (i.e. the

<sup>6</sup> The author belongs to that particular school of historical-comparative linguistics which considers a certain percentage of direct matches in core vocabulary to be the most reliable criterion for establishing both the fact of linguistic kinship and its degree.

meaning in Hadza and the meaning in the compared AA language(s) must be identical). To avoid confusion, each such word or a reconstructed form representing two or more languages is preceded by the sign =.

Additional *comparanda* are adduced (after '< AA') to demonstrate the reliability of the given AA etymology; naturally, these do not add to the score.

Further conventional symbols are as follows: (1) '~' simply means 'compare'; (2) '//' is a separator between the various AA branches; (3) ';' is a separator between groups within the same branch/family; (4) '◊' introduces bibliographic references to the sources.

1. 'all' *waʔi* ~ Omot. N.: =Basketo *woy-ci*, =Doko *way-ci* (< *way-t-*) 'all' <AA \**waʔ-* 'one, whole': Sem.: Arab. *wʕy* 'to collect, put together in one place' // Berb. \**yaw-n/t* (<\**yawʕ-*) 'one' // Eg. (Pyr.) *wʕ* 'one' // Cush. S.: Ma'a *we* 'one', etc. ◊ AADB 2873.

6. 'bird' *tʰitʰi*, *titiʔ* ~ =Chad. \**diʔ(diʔ)-* 'bird (gen.)': W.: =Kiir *dot*, =Ngizim *ḍúta*, etc., C.: =Podoko *ḍī-ya*, =Mofu *ḍiy-áŋ*, E.: =Lele *tidi*, =Migama *ḍīḍú*, =Mubi *ḍīḍúo*, etc. <AA \**diʔ(diʔ)-* '(k. f) bird': Eg. (Med.) *idw* 'Art Vogel' // Sem. \**daʔy(-at)-* 'bird of prey': Hbr. *dāʔā*; Arab. *daʔy-at-*, etc. ◊ ADB 360. Cf. SED II No. 64.

9. 'blood' *átʰaʔmá-*, *átámá* [ʔatʰama]<sup>8</sup> =AA \*(*ʔa-*)*dam-* 'blood': =all Sem. (except Mandaic and MSA) // =all Berb. (except Tuareg) // =Chad. \**dVm-* > *t-dVm-* > *H/t-dVm-*: W.: Angas *toom*, Bolewa *dòm*, Tsagu *ʔiim-an*, Bade *tə-dóm*, etc.; C.: Tera *tòm* // ◊ AADB 93; Stolbova 2016 #148.

10. 'bone' *midla* (Miller et al. 2021), *mitl'a* (Sands 2012: 5) =AA \**ma/iĉ-* 'bone': =Cush. C. \**ḥaṣ-* (<\**maṣ-*); =E.: HEC: Darasa, Burji *mičĉ-o*; Yaaku *moĉ-o*; =S.: Dahalo *miĉĉ-o*<sup>9</sup> // =Omot. N.: Mao (Sezo) *málĉ-é* ◊ AADB 1269.

16. 'come' *dza* (in Miller et al. 2021: 1, commented: Bantu) =Chad. W.: Hausa *zó*, C.: Gidar *zó* 'to come' <AA \**ʒaʔ/w-* 'go (slowly), come': Eg. (OK) *izy* 'go!' (alternately < \**ʒVʔ-*) // Berb.: Tuat *e-zzu* 'to march', Mzab, Wargla *zwa* 'to go' (alternately < \**ʒVʔ-*) // Cush. E.: Dullay: Tsamai *zey-* 'go' ◊ AADB 641; Takács 1999: 176; Stolbova 2016 #788.

17. 'die'<sup>10</sup>.

21. 'ear' *fiatʰapiʔʰi* (in Miller et al. 2021: 1, commented: <\**fiatsʰape* 'leaf'<sup>11</sup>) (met. < \**haĉaĉi piʔ*) =Omot. N.: Omoto \**Hayṣ-* 'ear', Bworo *aayĉ-ē* (cf. also S.: Ongota *ʔāš*, *haš* 'hear' and 'leaf') <AA \**ʎ/ħaṣ-* 'leaf; ear' ◊ AADB 1093.

23. 'eat' *seme*, *simi* =AA \**suʔVm-* 'eat': =Eg. (BD) *sɛm* 'eat' // =Chad. \**sVm-* 'eat': W.: Angas *səm*, Sura *sum*; C. \**sal-im-* ◊ AADB 1252; Stolbova 2009 #258.

<sup>7</sup> The other generic term for 'bird', *tsiriiri* [ʔsiliʔili] (for the symbol ʔ see fn. 12) is likely a loan from Iraqw *ĉirši*.

<sup>8</sup> In addition to this match, Hadza *tʰitʰi* vs. AA \**d* (mostly with \*ʔ or ʕ in the same root?) is confirmed by Hadza 'bird' *tʰitʰi*, *titi* vs. Chad. \**diʔ(diʔ)-* 'bird (gen.)' (above); Hadza *tetʰaʔo* vs. AA \*(*y*)*daʕ-* 'know' (#45); Hadza *ʔutʰume-ko* vs. AA \**daʔVm-/ʔudum-* 'spear' (below); Hadza *bitití-bii* 'fierce' vs. AA \*(*ʔV-*)*bVd-* 'crazy' (below), etc.

<sup>9</sup> Similarity between the Hadza and Dahalo terms made various authors suspect borrowing, but lack of other known Dahalo loanwords in Hadza speaks against the idea of 'bone' (which belongs to the most stable part of the core wordlist and is borrowed extremely rarely) to be the only word borrowed into Hadza from Dahalo.

<sup>10</sup> Hadza *miši* (in Miller 2021: 1, commented: Dat[ogaʔ]), cf. Chad. W.: S. Bauchi \**miš-* 'die': Kir *muse*, Polchi *miši*, etc. ◊ AADB 4274. In spite of a full coincidence, not scored (perhaps, super-cautiously). The verb, naturally, evokes associations with \**mwot*, the main NAA (Sem.-Eg.-Berb.-Chad.) term for 'die', but there seems to be no other cases of AA/Chad. \**-ti* > \**-si* in S. Bauchi. Stolbova's position is ambivalent: the section on Chadic Phonological Reconstruction (Stolbova 2016) contains no *-t* > *-s/š* transition, but in the Dictionary section, S. Bauchi \**miš/ši* is included into the entry \**mV(wV)t-* 'to die' (Stolbova 2016 #579 with the comment "all < \**mVti*").

<sup>11</sup> Another comment (Miller et al. 2021: 184): "jj vs zz personal variation [between various speakers - AM]... Starostin notes widespread connection between 'ear' and 'leaf' in African languages, thus perhaps hajjapitchi < hazzape 'leaf' (perhaps historically meaning \*ear), w[ith] sibilant assimilation of \*zz > jj before tch (itself \*iti > itci??)."



24. ‘egg’ *usle-ko* [ʔuʔeko] =Chad. \**šay(šay)-* / \**ʔi(n)š-* ‘egg’: W.: Bolewa *d-inš-a*, Karekare *ʔins-à*, Gera *is-à*, etc.; C. \**š/žay(š/žay)-*: Mandara *šay-a*, Logone *ʔens-e*, Munjuk *ež-e*, etc.; E. \**is/š-*: Migama *d-èès-è*, Mubi *d-íss-ó*, etc. ◇ AADB 1028.

30. ‘fly’ (v.) *pururu* [puʔulu]<sup>12</sup> ‘fly off, overhead’ =AA \**pir-* ‘to fly’: =Cush. N.: Beja *fir*, C.: Bilin *fir y-*; E.: Saho *fire*, LEC: Dirayta *fir-*; Yaaku *peri*; S.: Ma’a *puru* // =Omot. \**f/pir-*: N.: Wolaita *pir-ad-*, Gemu *fir-*, etc., Bworo *pirap-*; S.: Dime *far* // =Chad. W.: Kupto *firò*; C.: Gude *pár*, Podoko *pərərərə*, etc. // =Berb.: Semlal *firri*, Shawiya *fərfər*, Siwa *əm-fər*, Ahaggar *fərət-t*, etc. // =Eg. *pʔ* (<\**pVr*) // =Sem.: Tna. *näfärä*; Mhr. *farr*, Soq. *fer*, etc. ◇ AADB 692; Stolbova 2016 #617.

31. ‘foot’-1 (also ‘leg’) *a/uphukwa* [ʔa/up<sup>h</sup>uk<sup>w</sup>a] =Chad. W.: Hausa *kaƴàà* (Abraham 1962); E.: Kera *kámp-a* ‘foot, leg’ (metathesis?; the comparison is quite vulnerable both phonetically and semantically) <AA: Cush. S. \**fank<sup>w</sup>-*: Iraqw *fuknoo*, Burunge *faanķu* ‘hoof’, Alagwa *faanķu* ‘fleshy part of foot or leg’, Dahalo *funk-* ‘to walk rapidly’.

‘foot’-2 *asena-ko* [ʔasenako] ‘toes (people or animals)’ (Miller 2013: 23), ‘foot, toe, hoof’ (Sands 2012: 3)<sup>13</sup> =AA \**ʔa-sin-* ‘foot, leg’: =Eg. (20 Dyn.) *sn.wy* ‘zwei Füße’ (less likely < ‘two’) // =Chad. \**ʔa-sin-* ‘foot’: W.: S. Bauchi \**ʔa-sinH-*; C.: Higi-Nkafa *šini*, Banana *asénú*, etc.; E.: Migama *ʔásín*, etc. // =Cush. E.: LEC: Boni *saʔán* // Omot. N.: Dizi (Sheko) *šaanu* ‘foot’ ◇ AADB 216.

32. ‘full’<sup>14</sup> *furu-ne* ‘to be many, plenty, to be full’ =AA \**ʔVpVr-* ‘full’: =Berb.: Zenaga *tu-fur-t* // =Eg. (MK) *ʔpr* // =Chad. W.: Hausa (Katsina dial.) *fál*, *far* ◇ AADB 2965.

33. ‘give’ *kwe-* ~ AA \**kaʔ-* / \**ʔVk-*: =Berb. \**ʔawk* ‘to give’: Siwa *ūš* (<\**ūk*), Fodjaha *aš*, Zenaga *yaka* // =Chad. C.: Gidar *ki*; E.: Sokoro *áka* ‘to give’ (Stolbova 2011 #11: \**kVʔVw/y-* ‘to bring, carry’): ◇ AADB 3651.

34. ‘good’ *muta-na*, *mta-na*<sup>15</sup> =AA \**mVʔ-* ‘good’: =Chad. E.: Mubi *mḍé* ‘good’ // =Omot. N. \**mVʔt/ḍ-* ‘good’: Gofa *modē*, Koyra *modē*, Mao (Sezo) *maṭṭi* ◇ AADB 4123.

36. ‘hair’ *hadle* ~ AA \*(*Ha-*)*č/šVw-* ‘hair, feather’: Cush. S.: =Ma’a *ašú* ‘hair’ (\**č* and \**š* both render *š* in Ma’a, according to Takács 2011) // Eg. (Pyr.) *šw.t* ‘feather’ ◇ AADB 1284.

37. ‘hand’ *ukhwa-bii* [ʔuk<sup>w</sup>h<sup>a</sup>-] (also ‘fingers’) ~ Sem.: =Akk. *kaṭ-u* (< \**kaṭ-t-*) ‘hand’<sup>16</sup> // Cush. N.: =Beja *kákwi-i* ‘hand, arm’ (Reinisch 1895: 139) <AA \**kaṭi-* ‘claw, fingernail, hand’: Eg. (MK) *kaḥ* ‘(upper) arm, shoulder, elbow’ (<\**kaḥ-h* with nominal class marker *-h*, cf. Takács 1997) // Chad. W.: Mburku *kaṭwi*, etc.; E.: Mokilko *kóok-e* ‘claw, fingernail’ // Cush. E.: LEC: Konso *kaṭ-aa*, HEC: Darasa *kaṭy-a* ‘claw’, Dullay \**kaṭ-* ‘claw, nail, finger, hoof’ // Omot. N.: Oyda *kaṭ-a* ‘palm of hand’, etc. ◇ AADB 2206.

42. ‘T’ *ono* [ʔono, ʔono-ko] =AA \**ʔani/u* ‘T’: =Sem. (except MSA) \**ʔanV-kV* // =Eg. *ink*, Copt. (all) // =Berb. \**na/ikk* // =Chad. \*(*ʔa-*)*na(n)-* // =Cush. \*(*ʔ*)*ani* // =Omot. \*(*ʔ*)*inu/a* ~ \*(*ʔ*)*in-ta* ~ \**ta-ʔana* (metathesis) ◇ AADB 2762.

<sup>12</sup> In Miller et al. (2013), the symbol *l* conventionally stands for either *r* or *l* and may correspond to AA \**r* or \**l*. The other term for ‘to fly’ quoted in Miller 2021 is *hukwa*, but in Miller et al. 2021: 220 it is translated as ‘to fly away, take off’.

<sup>13</sup> Both terms show the distribution of meanings as ‘upper part of foot’ and ‘lower part of foot’ (including ‘toes and hoof’), allowing both terms to represent the notion ‘foot’ in the list.

<sup>14</sup> The other term for ‘full’ is a word with a click.

<sup>15</sup> In Miller 2021, quoted as \**mata-* (Bantu), but in Miller et al. 2021: 302 commented as “Bantu -tana, but not attested in neighboring languages” which makes borrowing from Bantu highly improbable. The synonym, Hadza *zzi’e* [ts’iʔe] ‘to be good, fine’ (Miller et al. 2021: 554), matches Omot. N.: \**c/ča/uy-*: Gimira (Bench) *soy*, (She) *cea*, *sea* ‘good’ <AA \**c/čuʔ-* ‘good, sweet’, but, in view of Iraqw (and Burunge) *cuuʔ* ‘(taste) sweet’ (<\**c/čuʔ-*, AADB 3046), the Hadza verb is suspicious as a possible borrowing from Iraqw and not scored.

<sup>16</sup> With strictly regular consonant correspondences (Beja *k* < \**ķ*) and trivial semantic changes, it is one of the examples of disregard by mainstream Semitologists of non-Semitic Afrasian comparative data: the Akkadian term for ‘hand’ is still considered “of uncertain origin”.

45. ‘know’<sup>17</sup> *tetha’o* [tet<sup>h</sup>aʔo] (also ‘understand’) =AA \*(y)daʕ- ‘know’: = Sem. \*yVdVʕ- (Akk., Ugr., Hbr., Aram.) // =Chad. C.: Mandara, Podoko *diya* // Cush. E.: =Saho-Afar \*d/ḍi/aʕ- // Omot. N.: =Bworo *daa-n* <AA: Cush. E.: LEC: Oromo *yaada*, Konso *yaat*- ‘think’; S.: Iraqw *daʕ-ati* ‘witchcraft, magic’, etc. // ◊ AADB 3472.

46. ‘leaf’ *fiats’a-pe* (pl.) =AA \*h/hʕaʕ- ‘leaf’: Sem.: =Palest. Aram. *hūš-*; =Soq. *heš* (‘palm leaf’ in PSem.) // =Cush. C. \*h/hʕaʕ- // =Omot. N. \*Hayʕ-: Omoto, Chara \*hayʕ-, Mao \*y/wac/ʕ- ◊ AADB 1093.

49. ‘long’ *thas-* [t<sup>h</sup>as-] =AA \*tays- ‘long’: Chad. C.: =Daba *tšim* // Omot. N.: =Hozo *-tiš-ti* // Cf. Sem.: Arab. *tys* ‘grandir et avoir les cornes comme un bouc (se dit d’un petit de chèvre)’ ◊ AADB 4265.

52. ‘many’ *ʔaso* =Cush. S.: Asa *-šāʔi* // =Omot. N.: Yemsa *šoyo* ‘many’ <AA \*w/yasaʕ- ‘big, many’: Cush. S.: Maʔa *-ša* ‘very (much)’ // Omot. N.: Dizi (Maji) *šwe-t* ‘all’ // Chad. W. \*was/c- ‘swell’; E. \*waHas/c- ‘swell, become bigger’: =Kujarke *wáašā* ‘many’ // Sem.: Arab. *w/yʕ* ‘be spacious, big’ ◊ AADB 2002.

56. ‘mouth’ *awani-ka* [ʔawani-ka] (also ‘lip, beak’)<sup>18</sup> =Omot. N. \*waan- ‘mouth’: Chara, Gimira, Yemsa, Kafa, Anfillo, Bworo, Mao (Ganza), (Sezo) *waani* <AA \*ʔawin- ‘tongue; (part of) mouth’: Cush. E.: LEC: Somali *ʕan* ‘cheek; inside of mouth next to cheek’; S.: Dahalo *ʕén-a* ‘tongue’ // (?) Chad. C. \*ʕan- ‘tongue’: Tera *ʔən-a*, Fali-Kirya, Higi-Nkafa *n-ʔàn-á* // (?) Sem.: Arab. *ʕinān-* ‘muscle longitudinal de chaque côté du cou’) ◊ AADB 1822; cf. Stolbova 2019 #331a.

61. ‘nose’ *intawe* [ʔiŋt<sup>h</sup>awe] =Chad. \*ʔi/a(wa)nti/an- ‘nose’: W.: Hausa *hánčì* < \*Hanti), Ankwe *wantín*, Bolewa *wunti*, Karekare *ʔwantín*, Kulere *ʔatánán*, etc.; C.: Gisiga *hətaŋ*, Musgu *nčìŋ*, etc.; E.: Migama *ʔítín*, Bidiya *ʔetēno*, etc. ◊ AADB 3037. Cf. Stolbova 2016 #767a.

62. ‘not’ *ukhuwa*, *ukuwa* [ʔu<sup>k</sup>huwa, ʔu<sup>k</sup>kuwa] =AA \*(ʔa-)k<sup>w</sup>ay- ‘not’: =Berb.: Ghadames *ak*, Fodjaha *ənk* // =Chad. \*kway: W.: Sura *ka*, Miya *kwa*, etc.; C.: Mandara, Sakun *ka*, Musgu *kai*, etc. // =Cush. \*ka-: N.: Beja *ka-*; E.: Tsamai *-kaka*; S.: Iraqw *-Vká* // =Omot. \*(ʔV-)kway: N.: Basketo *-kkaye*, Bworo *káyay*, Mao (Sezo) *ke:wε*; S.: Dime *ʔəkai* // Cf. Sem.: Akk. *akû* ‘destitute, weak, powerless’, Arab. *ʔawak-at-* ‘mal; malheur’ ◊ AADB 137.

63. ‘one’ *itçhâme* [ʔiçt<sup>h</sup>aame]<sup>19</sup> =AA \*ʕist(-an)- ‘one’: Sem.: =Akk. m. *ištēn*, f. *ištīat*, =Sab. *ʕs₁tn* // =Chad. \*sVt(-an)- ‘one’ (\*sVt- in Stolbova 2016 #722): =C.: Mafa *sta-d*, Podoko *taŋá*, Mofu *té-d*, Logone *səyədiya* (met. <\*ʕistən), etc. // =Omot. N. \*(H)ist(an)- ‘one’: Wolaita *issō*, *istā*, *issinō*, Yamma *isson*, Bworo *issa*, Dizi (Nao) *isn*, etc. ◊ AADB 2800.

64. ‘person’ *unu* [ʔunu] ~ AA \*ʔa/inay(-n)- ‘man, person (also elder kin)’: =Chad. W.: Pero *nìy-é*, Ngizim *nón* ‘person’ // Cush. E.: LEC: =Rendille *éneŋy-et* ‘person’ // =Omot. S. \*ʔani(n)-: Ari *aŋg*, Banna *aŋi*, Ubamer, Galila *aŋin-a* ‘man, person’, Ongota *ʔin-ta* ‘person’ ◊ AADB 1472.

<sup>17</sup> Another term, *zzahi* [ts’ahi] ‘know a person’, is not included as it is probably borrowed from Iraqw *caah* ‘recognize, understand’.

<sup>18</sup> Miller et al. (2013: 29–30) wonder if \**awa* (mouth?) can be a common element of *awanika* and *awati* ‘upper lip, both vermilion and area of moustache & philtrum’, which they compare, following Elderkin and Starostin to Rift \**ʔafa* ‘mouth’ and related Agaw and S. Omot. forms (in fact, < AA \*ʔap- ~ \*pay- ‘mouth’ AADB 245). The above comparison to AA \**ʔawin-* is valid only if *ʔawani-* and *ʔawati-* are not related.

<sup>19</sup> With the comments (Miller et al. 2021: 54): “Maybe itçhV- ‘alone, ʔanother’ plus msg -me?” and “Can we get ’itçhâko ~ ’itçhakoko for fem?”. The connections look very likely and resemble the AA structure (see, e. g., Akkadian and Wolaita). The Hadza form is a metathesis <\*ištā-me: the tš cluster is avoided in Hadza. Miller et al. (2013: 54) also quote Qwadza *itame* ‘one’, which does not match AA \**ʕis-tV-n-* and does not seem to have any etymology at all, but oddly looks somewhat similar to the Hadza term. Can it be a borrowing from Hadza? See a somewhat similar case with ‘two’ in fn. 25.

65. ‘rain’-1<sup>20</sup>.

‘rain’-2 (verb) *sa* =Cush. C. *\*siw-* ‘rain’ // Chad. C.: =Buduma *ha* (<*\*sa*) ‘to rain’; E.: Sokoro =*was* (met.) ‘rain’ <AA *\*sawi?*- ‘rain’: Beja *siw* ‘spurt’ // Omot. N.: Gimira *so?* ‘water’. AADB 3959.

67. ‘road’ *yeye* (also ‘way, path’) =Chad. E.: Kera *kákáy* (redupl.) ‘road’ // =Omot. N.: Dizi (Maji) *kook*, (Nai) *kuu-t-* ‘road’ <AA *\*kaw/y(k)-* / *\*yVhk-* / *\*hVwk-* ‘go, walk, come; road’: Omot. S.: Ari, Ubamer *kay-* ‘go, walk’ // Sem. *\*yhk/\*hwk*: Aram. Off. *yhk*, Samaritan *hwk* ‘go, walk’, etc.; Gz. *hwk* ‘stir, move’ // Berb. *\*?yVkk-* ‘come, go, walk’ ◊ AADB 3832.

71. ‘say’-1 *î* (<*\*?iy?*)<sup>21</sup> =AA *\*ya-* / *\*?iy-* ‘say’: Cush. =C.: Bilin, Kemant *y-*; =E.: Saho *iy-*, Afar *-iy(y)-*, LEC: Somali *ii-*, Dasenech *y-*, HEC: Sidamo *i-*, *y-*, etc.; =S.: Ma’a *-yo*, Dahalo *žō-m-* (<*\*yo-*) // =Chad. E.: Lele *yàá*, Somrai *yé* // =Eg. (Pyr.) *î* ◊ AADB 849.

‘say’-2 *he* (also ‘to tell’) =Cush. N.: Beja *hay-*; E.: LEC: Somali *hay-*, HEC: Darasa *hiy-*, Burji *hay-en-* // =Omot. N.: Koyra, Kachama, Ganjule *hii-*, etc. <AA *\*hay/w-* ‘say; speak, shout’: Sem.: Akk. *awû* ‘speak’, Ugr. *hw-t* ‘word, statement’ // Eg. (Pyr.) *îhy* ‘shout’ // Cush. S.: Dahalo *hwayu* ‘voice, noise’ ◊ AADB 856.

74. ‘sit’ *hama* [hama] (also ‘stay, lodge’) =Cush. S.: Asa *ʔamim-* ‘sit’ (*\*h* > Asa *ʔ*, acc. to Takács 2011) <AA *\*hVm-* ‘stay still (stand, sit, sleep)’: Cush. S.: Ma’a *huma* ‘stand’ // Chad. W.: Gurunum *yemmu*; E.: Birgit *ʔamí* ‘to sleep’ ◊ AADB 3075.

76. ‘sleep’<sup>22</sup> *ʔase* (also ‘lie’ Miller 2021) =Berb.: Audjila *iša* ‘sleep’ // =Omot. N.: Mao (Ganza) *šo-* ‘sleep’ <AA *\*say?*- / *\*ʔays-* ‘sit, sleep, rest’: Omot. N.: Dizi (Nao) *aša* ‘sit’ // Chad. W.: Hausa *sàyyáayaa* ‘take a rest’; C.: Zime-Batna *sóʔó* ‘to rest’ // Cush. N.: Beja *sāʔ*, *ʔassa*; E.: LEC: Arbore *siye*, Elmolo *asíya* ‘sit’ ◊ AADB 2182.

78. ‘smoke’ *ts’ikx’o* [ts’ik<sup>h</sup>o]<sup>23</sup> =AA *\*ç/Vg<sup>w</sup>-* ‘smoke’: Cush. C.: =Waag *çəg-a* (< Omot.?) ; S.: =Dahalo *ʔoggw-a* // =Omot. N.: Oyda *çugg-ə*, Kafa *çug-ō*, Anfillo *çug-o* ◊ AADB 1516.

79. ‘stand’ *ikha-* (also ‘to stop’) [ʔik<sup>h</sup>à] =Omot. N. *\*?ik(k)-* ‘stand (up)’ <AA *\*kaʔ/w-* / *\*?Vk(k)-* ‘rise, be high, stand (up)’: Chad. *\*kVʔVw-* ‘rise’ (‘to go up; top’ in Stolbova 2011 #293) ◊ AADB 659.

80. ‘star’ *ntsa-ko* [nt̪sako], *sa-ko* =AA (*t/wV-n-*)*ciʔ(ciʔ)-* ‘star’: =Chad. *\*(ti-)ca(c)-* ‘star’: W.: Hausa *tàçūniy-ā* (likely < *\*ta-çun-*): C.: Zime-Batna *ʔícíú*, Peve *čiču*, E. *\*tisaw/y-*: Lele *tèsé*, Kabalai *tás-a*, etc. // =Cush. S. *\*(ti-)ca(ca)ʔ-*:<sup>24</sup> Iraqw *cacēʔ*, Alagwa *ticaʔ-oo*, etc. // =Omot. *\*(wu-n-)cay(c)-*: N.: Male *šaci*, *wuçaci*, Chara *wonč/ʒe-ta*, etc.; S.: Ari (Banna) *wonč-o* ◊ AADB 4267.

82. ‘sun’ *isho-ko* (*isho* ‘sunlight’) [ʔiʃoko] =AA *\*ʔa/is-* ‘sun’: Berb.: =Ghadames *īsi* (Berb. *\*a/uss* ‘day’) // =Chad. C.: Logone *sə*, Makeri *sí* // =Cush. S.: Qwadza *as-o* ◊ AADB 1622.

84. ‘tail’ *zzaho* [ts’aho] =Chad. *\*çVH-* ‘tail’: W.: Hausa *wúçiy-àa* (met.), Diri *çúw-ā*; C.: Masa *čátú*, Dzepaw *čyáw*, etc. ◊ AADB 3352; Stolbova 2009 #942.

85. ‘that’-1 (distal) *\*na* =AA *\*nV/\*?Vn* ‘that’: =Eg. (PT) *p-n* (m.), *t-n* (f.), *n-n* (pl.), Copt. *nē* // =Berb.: Ayr *we-n*, Ahaggar *wî-n*, Zenaga *a-n*, Shawiya *wi-n*, etc. // =Chad. W.: Kiir *nòð-náñ*;

<sup>20</sup> Also ‘water’. Compatible with AA *\*ta?*- ‘flow, pour’ (AADB 991), but not scored because Proto-Khoe *\*tú*, Sandawe *\*to* ‘water’ probably fit better.

<sup>21</sup> Miller et al. (2019: 44) also cite Dat. *yi* ‘say’, but if the idea of the Hadza-Afrasian kinship holds water, the Afrasian origin of this verb seems more likely, given that verbs, especially belonging to the core vocabulary, are usually not borrowed.

<sup>22</sup> The other term for ‘sleep’ is a word with a click.

<sup>23</sup> Cf. *zzoko* [ts’oko] and *zzogô* (acc. to Obst) [ts’ogo] ‘fire, campfire, coal from fire’ with the comment “Cf. possible cognates *zziggo* ‘smoke’...” (Miller et al. 2021: 558). In view of the meanings of *zziggo-bee* (pl.) ‘smoking fires, streams of smoke’ and *zziggo-ko* ‘distant smoke’ (Miller et al. 2021: 555–6), the identification of the two terms, ‘fire’ and ‘smoke’ seems correct. The variation of velars (*k/k/g*) in oral comprehension (vs. *\*-g* in AA) may be due to dissimilation from *ç-* in the Anlaut. It should be noted that there are alternative Khoisan matches (Proto-Khoe *\*cáo* and Sandawe *\*çwā*).

<sup>24</sup> Borrowing into Hadza from the reduplicated form in W. Rift seems unlikely.

C.: Fali-Kirya *ní-ítá*; E.: Ubi *ʔà-n-am* // =Cush. C.: Aungi *á-n* // =Omot. N.: Gofa *in-ay-ssi*, Dizi (Nayi) *ne-ás* (m.), *ne-yin* (f.), (Hozo) *za-ɲaɲà* ◇ AADB 2880.

‘that’-2 \*b- =Cush. N.: Beja *ba-* // =Omot. N. \*bV-: Kafa *ebi*, Mocha *ɛbe*, Anfillo *bɛ-ni* <AA \*b- demonstrative and locative pronoun: Cush. E.: LEC: Konsoid \**-pa* (<\*ba) locative case ending // Sem.: Mhr., Jib., Soq. *bo* ‘here’ ◇ AADB 3863.

86. ‘this’ \*ha =AA \*ha(y): Sem. \*hā =Ugr., Aram., Arab. // Berb. =Ahaggar *wah* // Chad. =C.: Sakun *hayu*, Logone *ha-ma* // =Cush. S.: Alagwa *hee*, Burungi *ha, hi* // =Omot. N.: Omoto \**hay(-t)* ◇ AADB 3349.

87. ‘thou’-1 *t<sup>h</sup>e* (m.), *t<sup>h</sup>eko* (f.) =A \*(ʔan-)tV ‘thou’: =Sem. \*ʔantV // =Eg. (PT) *nt-k* // Berb.: =Ghat *ti-unti* (f.) // Cush. =C. \*ʔanti; =E. \*ʔatu/i; S.: =Dahalo *ʔatta* // Omot. N.: =Dizi \**yeta*; S.: =Ongota *ʒaa-melʒa-nta* (cf. AA \*ʒV ‘this, that’) ◇ AADB 2833.

91. ‘two’ *pi<sup>h</sup>e*, *pie-(be)* =AA \*(hV)pV(p)- / \*H/yV(m)p/b- (<\*hVp-?) ‘two’: =Chad. W.: Hausa *bíyú*, Sura *bap*, Fyer *poo*; C.: Mandara, Glavda *buw-a*, Peve *hwōb*, Zime-Batna *hōb* // =Cush. S.: Qwadza (m)be-a<sup>25</sup> // =Omot. N.: Yemsa *hep, yep*, Mao (Bambeshi) *yúmbó*, Ganza *mambu* (redupl.) // Cf. Sem.: Arab *ʿabb-* ‘pareil, égal à...’ ◇ AADB 1967. Cf. Stolbova 2019: 198.

92. ‘walk’ *haka* (*haka* ‘go’, *ʔetlhikwa* ‘walk’) =AA \*hw/yk- / \*kw/yH- ‘go, walk’: =Sem.: Aram. Off. *yhk*, Samar. *hwk* // =Berb.: Fodjaha *yuke*, Ahaggar *akk*, etc. // =Chad. E.: Lele *kàw* // =Omot. S.: Ari, Ubamer *kay-* ◇ AADB 1967.

93. ‘warm’-1 *piti-* ‘hot’, *pitipiti* ‘tepid’<sup>26</sup>.

‘warm’-2 ‘to warm (oneself by a fire)’ *sifi* [sihi] =Sem. \*šVhan- (likely <\*šVh-an-) ‘(be) warm, hot; warm oneself’: Akk. *šahānu*, Arab. *shn*, Gz. *saḥana*, Tna. *sāhanā* (all verbs), Jib. *šhan-ún* ‘warm, hot’ ◇ AADB 1871.

95. ‘we’ (inclusive) *uni-bii* [ʔuni-] =AA \*nV-(h/k)n- / \*naḥ- / \*(h/lan)a/in- ‘we’: =Sem.: \*(ʔa)naḥnu ~ \*naḥ- // =Eg. (Pyr.) *n* (dependent), (NE) *inn* (independent, Coptic *anon*), -*n* (oblique) // =Berb. \*nVkni // =Chad. W.: Pa’a *hnaà* (f.), *tinà* (m.); C. \*na(na); E. (all) // =Cush. N.: Beja *hanín*; C. \*(ʔ)ya/in(n)a; E. (all) \*ʔin-/ \*nV(-n), S.: Iraqw *aten*, Ma’a *nine*, Dahalo *naṇi* // =Omot. N. \*nu-/ \*nuni ◇ AADB 2005.

96. ‘what?’ *akwiʔa* [ʔak<sup>wi</sup>ʔa] =Chad. C.: Daba *ki-n*, Logone *ɣwa-ni* ‘what?’ // =Omot. N.: Mao (Hozo) *ki-nda*, (Sezo) *kí-nà*, (Bambeshi) *kò-mísijà* (Blažek 2008) ‘what?’ <AA \*k<sup>w</sup>ay- ‘what? who?’ ◇ AADB 2021.

97. ‘white’ *pedla* (*petl’a-* in Sands 2012) =Sem. \*pays̄- (AA \*payç-) ‘white’<sup>27</sup>.

98. ‘who?’ *akwaza* [ʔa<sup>k</sup>waɖza] (<\*ʔak<sup>w</sup>-) ~ AA \*k<sup>w</sup>ay- ‘what? who?’: Chad. C.: =Tera *kiya* ‘who?’ // =Omot. N.: Kafa *ko-ni*, Mocha *ko*, Anfillo *ko-nne*, Bworo *ko-nni*, Mao (Sezo) *kí-nà*; (Bambeshi) *kiya* ‘who?’ ◇ AADB 2021.

<sup>25</sup> Cf. Miller et al. 2021: 347: “Qwadza (m)bea does not seem to be Cushitic..., so if Qwadza bea is related to Hz *pie*, the direction of loan is not clear, assuming it’s even a loan... Hadza *itchâme* ‘1’ also similar to Qwadza *itâme*.” Note that there is an alternative source of borrowing -βwī ‘two’ in the Bantu Nyaturu language spoken in the Singida region of Tanzania (with matches in some more distant Bantu languages). See a somewhat similar case with ‘one’ in fn. 18.

<sup>26</sup> Cf. (not scored, of course) AA \*fVṭ- ~ \*wVfVṭ- ‘to process by heating’ (AADB 485): Sem. \*wpt̄ ‘to burn, cook, make pottery’: Sab. *wft̄* ‘to burn’, Gz. *wafaṭa* ‘to burn, cook, ignite’ // Eg. (NK) *fty* (<OK \*fdy?) ‘smth. referring to the fabrication of metal weapon’, (Med.) *wft̄* ‘to drill, bore’ // Chad. W.: Hausa *fyāḍā* ‘to hit with smth. flexible’; C.: Mada *vād* ‘to forge’, *ávàḍ* ‘choffer, forger, percer’, Mafa *viḍ-* ‘forger, fabriquer’; E.: E. Dangla *pāḍé* ‘marteler le fer rougi au feu, forger à chaud’ (Cf. Takács 2001: 593).

<sup>27</sup> About the much-debated problem of relating Arab. *ʔabyaḍ-* ‘white’ either to Arab. *bayḍ-at-* ‘egg’ (with Sem. and Chad. cognates) or Akk. *pešû* ‘white’ see SED I No. 43 and II: 338. For me, as a proponent of Semitic (and Afra-sian) \*p̄, the argument of the main term for ‘white’ coinciding in the two Semitic languages to some extent outweighs its apparent lack of Sem. and AA cognates.

In the 100-item wordlist for Hadza, there are 17 words with clicks; 8 unequivocal or highly probable loans from W. Rift (most likely Iraqw); one from Bantu; one from Dat.; three items can perhaps be better explained as borrowings from Khoe and Sandawe. I rank all of them as loans and, in accordance with Sergei Starostin's method (Starostin 2000), remove them from scoring — that is, the Hadza-Afrasian 100-item wordlist is reduced by 30 units while augmented by 9 items representing full synonyms ('bird', 'full', 'good', 'know', 'to rain', 'say', 'that', 'warm', 'we'), i.e. 79 items remain for comparison, of which thirty find no matches in AA, while 49 have matches in one or more AA languages which receive positive scores for presumed affinity<sup>28</sup>.

Table 3 lists the percentages of matches between Hadza and the various subgroups of AA on the 100- and 50-item wordlists.

As can be seen from the table, nearly all Hadza matches with different Afrasian groups and individual languages show a higher percentage in the 50-item (more stable) wordlist than in the 100-item wordlist. This is a strong argument for kinship. Of course, the table also shows that not everything is so smooth. In addition to uneven distribution of the Hadza matches with the individual Afrasian languages and groups closely related to each other,<sup>29</sup> there remain a number of hard-to-explain cases, such as an extremely low percent of matches with Qwadza and Asa. Moreover, if the surge in some Hadza-Cushitic matches can be explained by geographical proximity (undetected loans or influences?), a similar surge with some Chadic languages is hard to explain by the same logic, to say nothing of several standalone and strikingly similar Hadza-Chadic zoonyms.

However, the average percentage of matches between Hadza and Afrasian in both lists seems to imply kinship, with a higher proximity between Hadza and the Cushitic and Omotic branches; this provides us with some grounds to hypothetically separate the Hadza language into a third — alongside Cushitic and Omotic — subbranch of the South Afrasian branch of the AA macrofamily.

At the same time, of course, it is impossible not to pay attention to the high percentage of coincidences with individual Chadic languages (Tera, Mubi, etc.), which is not easy to explain. Even more enigmatic is the coincidence, which is difficult to recognize as accidental, of such a unique grammatical phenomenon as the infixation of *-k-* with the meaning of plurality, intensity of action, etc. in both Hadza and some of the Chadic languages.

---

<sup>28</sup> Hadza-AA matches representing a common AA (like #42 'I') or at least a common AA branch root (like #24 'egg' in Chad.) are, of course, of much better *quality* than Hadza matches with a few isolate and disperse AA terms (like #32 'full'). However, lexicostatistics is a *quantitative* method relying on a relative percentage of coincidences; it must meet the requirements (at least in the Starostin version) of regular sound correspondences and representative semantics, whereas the probability of some percentage of random lookalikes is the same when comparing both related and unrelated languages. For comparison: lexicostatistics that I have applied to AA vs. 11 Nubian and to AA vs. 3 Kuliak languages on the basis of both 100- and 50-item wordlists shows no more than 3% of matches, which is the level of random noise, excluding the possibility of kinship despite the fact that in both Nubian and Kuliak languages there are quite a few loanwords from Afrasian outside the basic vocabulary. The same result — no genetic affinity — is for the lexicostatistical comparison between AA and Sumerian, or between AA and Elamite; both of these extinct West Asian languages have a lot of common areal lexical items with AA, including non-Semitic Afrasian branches (which *inter alia* supports my theory of the Proto-Afrasian Urheimat in West Asia.)

<sup>29</sup> E. g., in the 50-word list: 10 in Mashile vs. 3 in Dasenech; 11 in Iraqw vs. 5 in Burunge; 17 in Gemu vs. 6 in Zaisse (Ometo); 20 in Chara vs. 9 in Mocha; very low percent in S. Omot. languages vs. 9 in Ongota; 14 in Amharic vs. 6 in Soddo; very low percent in MSA; 20 in Tera vs. 5 in Bachama; 0 in Masa vs. 13 in Peve, etc.

|                       | 100-word list       | 50-word list                    |
|-----------------------|---------------------|---------------------------------|
| Omotic:               |                     |                                 |
| N. Omotic (average)   | 11                  | 14                              |
| Ometo (average)       | 8.5                 | 12.5                            |
| Chara                 | 12                  | 20                              |
| Gimira (average)      | 6.5                 | 10                              |
| Yemsa                 | 10                  | 11                              |
| Gonga (average)       | 9                   | 13 (Kafa 17)                    |
| Dizoid (average)      | 6.5                 | 12                              |
| Maoid (average)       | 11                  | 17.5 (Bambeshi 20)              |
| South Omotic          | 3                   | 4                               |
| Cushitic:             |                     |                                 |
| Beja (after Reinisch) | 6                   | 12                              |
| C. Cush. (average)    | 9 (Bilin, Qwara 12) | 15 (Bilin, Falasha 17, Waag 18) |
| E. Cush. (average)    | 6                   | 9.5                             |
| Saho-Afar (average)   | 8                   | 9                               |
| LEC (average)         | 5                   | 7 (Diraita 10)                  |
| HEC (average)         | 6                   | 10.5 (Burji 12)                 |
| Dullay (average)      | 5                   | 12.5 (Tsamai 15)                |
| Yaaku                 | 4                   | 9                               |
| S. Cush.              | 4                   | 7                               |
| W. Rift (average)     | 4.5 (Iraqw 5)       | 7 (Iraqw 11)                    |
| Asa                   | 1                   | 0                               |
| Qwadza                | 2                   | 3                               |
| Ma'a                  | 6                   | 9                               |
| Dahalo                | 8                   | 17                              |
| Chadic:               |                     |                                 |
| West (average)        | 4.5 (Hausa 9)       | 9 (Hausa 14)                    |
| Central (average)     | 5.5 (Tera 10)       | 9.5 (Tera 20, Musgu 16)         |
| East (average)        | 6 (Mubi 10)         | 11 (Mubi 17, Migama 16)         |
| Berber:               |                     |                                 |
| North (average)       | 7                   | 7                               |
| Zenaga                | 8                   | 9                               |
| East (average)        | 7.5                 | 12                              |
| Tuareg (average)      | 4.5                 | 6.5                             |
| Egyptian              |                     |                                 |
|                       | 4                   | 6                               |
| Coptic (average)      |                     |                                 |
|                       | 3.5                 | 6                               |
| Semitic:              |                     |                                 |
| Akkadian              | 8                   | 13                              |
| Hebrew                | 5                   | 10                              |
| Aramaic (Syr.)        | 9                   | 11                              |
| Arabic (Qur'an)       | 8                   | 12                              |
| Ethiopian (average)   | 5                   | 10                              |
| Modern S. Arabian     | 3.5                 | 3                               |

Table 3. Percentages of Hadza matches with the various subgroups of AA.

In Hadza (Miller et al. 2021: 7–8):

- “ $\langle kV_1 \rangle$  [Same as  $-kV-$  A.M.] (after first CV, echoes first V.) pluractional. (very common)
- (1) on verbs: multiplicity of
    - (a) subject, esp. on intransitive verbs
    - (b) action, esp. on intransitive verbs with singular subjects
    - (c) object, esp. on transitive verbs with singular subjects....
  - (2) on nouns, adjectives:
    - (a) multiplicity of detail...
    - (b) greater or lesser extreme...

In Chadic (Leger & Stolbova 2021: 42–45):

“The semantics of the Kupto verbs... suggest that there is an infix pluractional morpheme  $-k-$  in their basic... form... (42). In Maha: pluractional verbs  $kàay-áayò$  ‘to plant’ pl.  $kàky-áayò$  ‘to plant many time’ ...  $wèen-áayò$  ‘to observe’ pl.  $wèkn-áayò$  ‘to observe often’... (43). ‘FROZEN  $-K-$  PLURACTIONAL VERBS IN THE ANGAS-SURA GROUP (44). MASS AND LIQUID NOUNS WITH  $-K-$  INFIX

a. PCh  $*dVm-$  ‘blood’: P-AS  $*tVkvVm$  (pl.)  $< *tVm$  (regular devoicing of initial Chadic  $*d-$ ): ...Kofyar  $tàgàm$ , Mushere  $tekem$ , Mwaghvul  $tỳòm$ , Ngas  $toom$ ... (45).”

Glottochronology<sup>30</sup>, again applied according to Sergei Starostin’s version of the method<sup>31</sup>, has yielded the turn of the 10<sup>th</sup>–9<sup>th</sup> millennia BCE as the time of separation between Proto-Hadza, Proto-Cushitic and Proto-Omotic (vs. circa 10 300 BCE as the Proto-Afrasian time of separation into SAA and NAA represented by Semitic, Egyptian and Berber-Chadic).

### Non-basic lexicon matches

The abundance of matches in the non-basic vocabulary is also an argument in favor of relationship, if indirect. Below I list some of the best ones, grouped by various semantic fields.

#### I. Animals.

##### I.1. Ungulates.

##### I.1.1. Bovids and pigs.

Hadza  $biso-ko$  ‘wildebeest, gnu’ (Miller et al. 2021: 89) ~ AA  $*bus-$  ‘goat, bushbuck’: Cush. E.: LEC: Oromo  $buš-aa$  ‘goat’; S.: Iraqw  $baʒas-a$  ‘bushbuck’ // Omot. N.: Bworo  $bush-aayá$  ‘Kleinvieh’ // Chad.  $*bus-$  ‘he-goat’  $\diamond$  AADB 4270; Stolbova 2019: #135.

Hadza  $geweda-ko$  ‘dikdik’ (Miller et al. 2021: 151) ~ AA  $*guday-$  ‘k. of bovid’: Cush. E.: LEC: Oromo  $gad-am-sa$  ‘greater kudu’, Dirayta  $gad-an/m-sa$  ‘antelope’, HEC: Sidamo  $god-a$  ‘deer, gazelle’; S.: Iraqw  $gwand-a$ , Alagwa  $gwand-o$  ‘ram’, etc. // Omot. N.: Zaise  $gaid-é-endo$  ‘buffalo’ // Chad.  $*galuday-$ : W.: Hausa  $gàd-áa$ , Ngizim  $gád-ùwà$  ‘duiker’, etc.; C.  $*gVday-$ : Zime-Batna  $gódàyy$  ‘buck’ // Sem.  $*gaday-$  ‘kid’: Ugr.  $gdy$ ; Hbr.  $gadi$ ; Aram.: Syr.  $gady-ā$ ; Arab.  $žady-$  ‘kid’,  $žadāy-at-$  ‘gazelle; petit de gazelle’  $\diamond$  AADB 2490; SED No. 76. Cf. Militarev & Nikolaev 2020 #1.13a.

<sup>30</sup> Another method that is at best recognized by the mainstream as “controversial”. My experience of using glottochronology in Sergei Starostin’s version gives quite plausible results, especially with the Semitic family (in which it is easier to find historical parallels to linguistic divisions), perhaps due to the ‘sharp’, one-time separation of the speaking groups from each other, due to large distances preventing from immediate contacts, which “blur” the picture of linguistic split as in the case of Slavic languages. The most convincing example of coincidence of the proto-language split time in glottochronology with archaeological and genetic dating (and even with the ‘legendary’ historical tradition) is the Ethiopian Semitic case (see Militarev 2023).

<sup>31</sup> Yielding much more plausible proto-language split dating wherever these dates are verifiable extralinguistically.

Hadza *ndama* ‘calf’ (Miller et al. 2021: 321)<sup>32</sup> ~ AA *\*damaʔ/ɣ-* ‘k. of bovid, antelope’: Cush. E.: HEC: Hadiya *damal-iččo* ‘antilope dekkula’ (with the fossilized *-l* suffix); S.: Iraqw *damaʔ-amo*, etc., Asa *domo-k*, Qwadza *damay-ituko* ‘eland’ // Omot. N.: Male *damm-ə* ‘calf (bovine)’, Yemsa *dum-ā* ‘antilope dekkula’ // Chad. *\*dVm-*: W.: Mburku *dumun* ‘duiker antelope’; C.: Masa *dimi*, Afade *dümm-o*, etc.; E.: Somrai *demai*, Tumak *d̄m-ā* ‘sheep’ // Berb.: Tlit *á-dmū*, Zenaga *dāmi*, Ahaggar *e-demi*, etc. ‘gazelle’ ◇ AADB 2464; Militarev 2022 #32.

Hadza *nja* [nɔ̄ʒa] ‘(Bohor) reedbuck’<sup>33</sup> = *minza* (Miller et al. 2021: 328), both having AA matches, obviously related to each other:

(a) Hadza *nza* ~ AA *\*ʕi(n)ɣ-* ‘k. of smaller bovid’: Cush. *\*ʕiɣ-* ‘k. of lesser bovid’: E. *\*ʕiɣ-* and *\*ʕiɣ-*: Saho *ʕid-oo*, Afar *ayd-ā* ‘sheep’, LEC: Somali *id-o* ‘flock of sheep’, Arbore *ʕizz-e* ‘gazelle’, Elmolo *ʕéd-i* ‘goat’ // Sem. *\*ʕalinz-/ʔVɣ-* ‘(she-)goat’: Akk. *enz-u*; Hbr. *ʕēz*; Aram.: Off. *ʕnz*, *ʕz*; Sab. *ʕnz* ‘goats’; Arab. *ʕanz(-at)-*; Jib. *ʕɔz*, Soq. *ʕoz* ◇ AADB 2471. Cf. Militarev & Nikolaev 2020, #1.4a; SED II No. 35.

(b) Hadza *minza* ~ AA *\*ma-ʕi(n)ɣ-* ‘k. of bovid’: (?) Cush.: S.: Dahalo *máʕaq-e* ‘female topi’<sup>34</sup> // Omot. N. *\*miHi(n)z-* ‘cow, bull’: Wolaita *mízz-a*, Yemsa *miž-à* ‘cow’, Kafa *minž-o* ‘cattle’, Dizi (Sheko) *mīnz-a*, etc. // Chad. *\*maHa(n)z-* ‘k. of bovid; male-goat’: W.: Hausa *màaz-o* ‘harnessed antelope’, Mburku *maaz-u* ‘antilope sp.’; C.: Logone *máz-ā* ‘Antilopa hamariya’, Matakam *múz-àk* ‘he-goat’; E.: Barein *mú-z-o* ‘ox’ // Sem. *\*miʕaz-*: Aram.: Jud. *mēʕazz-ē* ‘goats’;<sup>35</sup> Minean *mʕz-y* (du.) ‘chèvre’; Arab. *miʕāz-* ‘chèvre ou bouc’ ◇ SED II No. 148; Takács 2008: 156–157, 796; AADB 2695; Militarev & Nikolaev 2020, #1.4b.<sup>36</sup>

### 1.1.2. Equids.

Hadza *dongo-ko* ‘zebra’<sup>37</sup> ~ AA *\*da(n)gʷ(-ay/r)-* ‘k. of equid’: Cush. E.: Oromo (Barareta dialect) *dongorr-a* ‘donkey’; S.: Alagwa *ndaḡʷai*, Qwadza *daḡʷagwai-ko* ‘donkey’ (Dolgopolsky 1973: 275) ◇ AADB 3318.

### 1.1.3. Large herbivores<sup>38</sup>.

Hadza *begḡawu-ko* [be'kx'au-ko] ‘a bull elephant’ (Miller et al. 2021: 87)<sup>39</sup> ~ AA *\*bakʰaw-*<sup>40</sup> ‘bull, cow, large cattle’: Eg. (Late) *bḥ* ‘Buchisstier’ // Berb. *\*baḡaw*: Ayr, E. Tawllemmet *e-bāyāw* ‘vieux boeuf’, Nefusa *byu* ‘veau’ // Chad. W.: Hausa *ḡakwáa-nè* (pl. of *ḡauna*) ‘dwarf buffalo’

<sup>32</sup> Commented as [Sands; maybe Sukuma? identical to Swahili].

<sup>33</sup> Commented as [Isanzu nja, Sukuma njaa]. Could the perfect Hadza match with the AA terms imply a borrowing from Hadza?

<sup>34</sup> *d* is usually considered to reflect only *\*d*, but a few other cases of Dahalo *d* < *\*ɣ/\*ʕ*, including the present reflex in a triconsonantal root, do not seem to be haphazard.

<sup>35</sup> Oddly translated in some Aram. dictionaries as ‘from goats, goats-hair, horn, etc.’.

<sup>36</sup> Looks like a unique case of derivation with the prefix *m-* from *\*ʕi(n)ɣ-* ‘k. of smaller bovid’ as early as the PAA level. With all the problems in each of the proposed roots, a rather rare combination of the radicals *\*ʕ* and *\*ɣ* makes chance coincidence highly unlikely.

<sup>37</sup> Commented in Miller et al. (2021: 130): Cf. Dat. *diḡèdà* ‘donkey’, < *\*dakee* > pWRift *\*dakeetu* ‘zebras’, sg *\*dakeetiya*; but Dat. itself perhaps from pWR *\*daḡway* ‘donkey’. Sandawe *doro zebra*, *dak'we* ‘donkey’.

<sup>38</sup> Cf. also a striking and hard-to-explain case: Hadza *wezza'i-ko* [ʔwe'ts'aʔiko] ‘hippopotamus’ (Miller et al. 2021: 506) ~ Chad. W.: Angas *wūžāi*, C.: Muskum *wūzil* ‘hippo’ (Blažek 1994: 201). The stunning similarity between the Hadza and Angas forms (the Cush. terms quoted by Blažek are hardly related) do not look haphazard. Can it be an Angas loan in Hadza? A common substratum term? See also fn. 45.

<sup>39</sup> Dr. Sands (personal communication) suggested for this zoonym *bek'ahu-ko*, the general term for ‘elephant’.

<sup>40</sup> In the alternative notation, *\*baḡaw-*. A phonetically unique, if debatable, case: Hadza *kx'* (in the version by Miller et al.) neatly matches the hypothetical AA *\*kʰ* or *\*q* reconstructed by Diakonoff’s team on the evidence of a very limited number of examples with *\*k* in Berber, Chadic, most Cushitic and Omotic supposedly corresponding to *\*ḥ* in Semitic and Egyptian (perhaps also *ḥ*). Semantically, ‘elephant’ vs. ‘bull’ is normal.



(Abraham 1962: 90); E.: Kera *bèke*, Saba *boko* ‘cow’ // Cush. E.: Yaaku *baḥbaḥ* ‘small dikdik antelope’ ◇ AADB 2593; cf. Stolbova 2021 #108.

Hadza *rôsho* [lo:fo] ‘rhinoceros’ (Miller et al. 2021: 400),<sup>41</sup> likely <\*roš- ~ AA \*warš-<sup>42</sup> ‘rhinoceros’: Cush. E. \*worš- ‘rhinoceros’: LEC \*worš-: Oromo *worse-sa*, Konso *orš-ayta*, Gidole, Dirayta *orš-ayt*, HEC \*worš-aŋ-: Hadiya *oršaŋ-aḍo*, Burji *wórš-a*, Dullay \*wVrs-Vŋ-: Gollango, Harso *oršaŋ-aḍo*, Yaaku *órse?* // Chad. C. \*warž-: Mbara *wí(r)ž-ĩ*, (?) Gidar *walš-ya* (-l < \*-r and \*ž > š influenced by l?) ‘bull’ // Sem.: Eth. \*ḥawriš-<sup>43</sup> ‘rhinoceros’: Gz. *ḥariš*, Tna., Tgr. *ḥariš*, Amh. *haris*, *awraris* ◇ AADB 2529. Cf. Blažek 1994; Militarev and Nikolaev 2020 #3.7.

## I.2. Predators.

### I.2.1. Canines.

Hadza *hadehade* ‘wild dog’<sup>44</sup> ~ AA \*Hayd- ‘dog’: Berb. \*a-yd/di ‘dog’ (all except Siwa and Aujila) // Chad. \*hid-, *hadd-* ‘dog’: W.: Bolewa, Ngamo *ḥàdà*, etc., C.: Tera *yìdà*, Gudu *hídà*, etc. // Cush. S.: Ma’a *idiḥe* ‘dog’ (hardly a lw.) (met.) ◇ AADB 2939; Stolbova 2019 #116; Militarev 2022 #1.

Hadza *biriri-ko* [bililiko] ‘bat-eared fox’ (Miller et al. 2021: 89) ~ AA \*ba(y)r- ‘k. of wild canine’: Cush. E.: Yaaku *bari-e* ‘jackal’ // Chad. E. \*bar- ‘jackal’: Migama *bòr-ú*, Birgit *bâr-á*, etc. // Berb.: Shilh *a-bayrr-u* ‘renard’ (Naït-Zerrad 1998: 146) // Sem. \*ba(r)bar- (redupl.): Akk. *barbar-u* ‘wolf’; Arab. *babr-* (pl. *bubūr-*) ‘espèce de chacal qui conduit le lion sur la piste de la proie’ ◇ Cf. PMasai \*-barie ‘jackal’. AADB 1863. Cf. Militarev & Nikolaev 2021: #7.4; Takács 2001: 22, 148.

Hadza *gondera* [gondela] ‘dog’ (not the main term) ~ Chad. \*gVHVd- ‘dog’: W.: Hausa *góòžèè* (< \*gVHVd-); C.: Dghwede *gdè*, Mofu *gədəy*, etc.; E.: Somrai *dò-gádà*, Mokilko *gédè*, etc. <AA \*gV(ŋ)d-Vl/r- ‘k. of canine’: Sem.: Arab. *ḣaŋd-at-* ‘loup’, *ḣadlā?* ‘chienne’ // Cush. E.: Oromo *gedall-o* ‘jackal’ // Omot. S.: Ari *gudrí* ‘hyena’ ◇ AADB 353; Stolbova 2016 #196.

### I.2.2. Felines<sup>45</sup>.

Hadza *mondo* ‘a large male lion’ (Miller et al. 2021: 291)<sup>46</sup> ~ AA \*mandaw- ‘k. of predator’: Chad. C.: Afade *maud-á* ‘hyena’; E.: Mubi *ḥàmdàw-ùt* ‘cat’ (isolated and disparate terms, not quite reliable) // Sem. \*mV(n)dīn- ‘a large wild cat’: Akk. (OB on) *mindin-u* (*middin-u*, *mandin-u*) ‘tiger (?)’; Arab. *ḥal-madīn-* (met.) ‘lion’ ◇ AADB 4254; cf. SED II No. 151.

### I.2.3. Hyenas.

Hadza *uzame-ko* [ʔudzameko] ‘spotted (laughing) hyena’ (Miller et al. 2021: 76) ~ Chad. \*zVm- ‘lion’: C.: Musgu *zenīm* (met. < \*zVm-n), Masa *zími*; E.: Kwang *zém-ki* (cf. Muktele *ḣim-dàlí* ‘leopard’)<sup>47</sup> ◇ Stolbova 2009 #499.

## I.3. Rodents.

Hadza *yondo*, a general term for mouse, rat (Miller et al. 2021: 543)<sup>48</sup> ~ AA \*ḥandaw- ‘mouse’: Cush. E. \*ḥandaw-: Saho *andaw-a*, Afar *andaw-aa* ‘mouse, rat’ // Chad. C.: Mofu *m-*

<sup>41</sup> Derivation from roko>shô ‘stooped, ready to charge’ seems to be a folk etymology.

<sup>42</sup> The reconstruction of \*-š (not \*-ĉ) is based on the Mbara form alone.

<sup>43</sup> Initial ḥ- absent in the other languages has to be explained; there is also Syr. *ḥars-ūm-ā* ‘proboscis; labia bovis’ which matches the Eth. noun phonetically, but the chain of semantic shifts is hard to imagine.

<sup>44</sup> *Lycan pictus*, according to Dr. Sands (p.c.).

<sup>45</sup> Another case of striking and hard-to-explain resemblance with Chadic forms (see fn. 39): Hadza *sigwazi*, *sigwasi* [siḡ<sup>w</sup>adzi, siḡ<sup>w</sup>asi], an admiring name for *seseme* ‘lion’ (Miller et al. 2013: 419) ~ Chad. W.: Karekare *cágázáú*, *cágáḣaw*, Ngizim *ḣágázáú* ‘lion’.

<sup>46</sup> Compared in Miller et al. 2021: 291 to Burunge *moondo* ‘jackal’, from pBantu \*mondo. Semantically hardly compatible, especially in a presumed borrowing.

<sup>47</sup> Cf. also Chad. W. \*n-zVm- ‘griffon’: Bolewa *n-zimò-kì*, etc. (Stolbova 2009 #500). If related, the common proto-meaning is ‘a scavenger’. Another peculiar Hadza-Chadic zonym?

*àndùw-àṅ*, Gisiga *m-onduw-aṅ*, Muktele *m-ádàw-á*, Matakam *m-ùdùw-à* ‘rat’ ◇ AADB 285; Militarev & Nikolaev 2021: #7.2.

Hadza *giririba-ko* [gililibaʔuko] ‘striped grass rat, or striped mouse’ (Miller et al. 2021: 153) ~ Chad. E.: Mokilko *gàrb-ô*,<sup>49</sup> Mubi (met.?) *gùmbùr-ò* ‘mouse, rat’ ◇ Jungrauthmayr & Ibrișimow 1994.

Hadza *wajjo* [watʃo], unidentified sp. mongoose (Miller et al. 2021: 503 after Sands) ~ AA \*ʔV(n)čaw- ‘k. of smaller carnivoran or rodent’: Cush. C. \*ʔi(n)čaw-: Bilin *inšuw-aa*, Khamir *iečuw-aa*, etc. ‘mouse’; E.: LEC \*wawač- (met.): Oromo *wawwač-oo* ‘mongoose’ // Omot. N. \*ʔi(n)čaw- ‘rat’: Wolaita *uça-a*, Kafa *ičo-o*, Bworo *inčo-o*, Gimira *uč/ç*, Dizi (Sheko) *iičč-o* // Sem. \*ʔay(n)šaw-: Akk. (OB on) *ayāš-u* ‘weasel’; Gz. *ʔanšaw-ā* ‘mouse, weasel’, Tna *ʔančəw-a* ‘rat, mouse’, etc. // (?) Berb. S. \*-dVway: Ayr. E. Tawllemmet *e-dəwi* ‘jerboa’ // Chad. W.: Hausa *číy-ò* ‘field rat’, Sayanchi *ààcá*; C.: Musgu *ausi* ‘mouse’ ◇ AADB 379; SED II No. 26; Militarev & Nikolaev 2021 #7.1.

#### I.4. Birds.

Hadza *garaani*<sup>50</sup> ‘heron’ (Miller et al. 2021: 150) ~ AA \*gawir- and \*garay- ‘heron, crane, ostrich’: Cush. E. \*garay- ‘ostrich’: Saho *gàraay*, Somali *goray*, etc.; S.: Dahalo *ngára* ‘crested crane’ // Chad. C. \*girw-: Mandara, Padukwo *žirrw-e*, Uzam *žirw-e* ‘ostrich’; E. \*gawr- ‘heron’: Kera *gúgur-i*, Somrai *gwár-a* // Sem. \*ʔVgVr-: Akk. (OB) *igir-ú* ‘heron’, etc. ◇ AADB 1931; cf. Militarev and Nikolaev 2021 #8.6.; SED II No. 29.

#### I.5. Reptiles.

Hadza *jjowa-* [tʃoa] ‘gecko’ (Miller et al. 2021: 230) [čowa] ~ Sem. \*ʔiṭāy- (< AA \*ʔiçay-) ‘k. of lizard’: Akk. *iššû* ‘gecko’; Arab. *ʔiḏāy-at-* ‘sorte de lézard’ ◇ SED II No. 46.

Hadza *miro* ‘spitting cobra’ (Miller et al. 2021: 287) ~ AA \*mayr- ‘k. of snake’: Cush. C. \*mVr- ‘snake’: Bilin *mər-aaw-aa*, Kemant *mer-ewa*, *mär-āwaa*, Aungi *muri*, etc. // Chad. W. \*mīr- ‘python’: Gwandara *mēr*, Angas *myirm*, Sura, Ankwe *mir*, etc. ◇ AADB 371.

#### I.6. Worms and insects.

Hadza *duma-ko* ‘a bug that plays dead’ (Miller et al. 2021: 131) ~ AA \*dV(?)m- ‘k. of insect’ (see also \*dVm- ‘worm’): Sem.: Akk. *dim-ī-tu* ‘locust’, *dimān-u* ‘insect’; Arab. *dimmat-* ‘louse, ant’ // Eg. (NK) *dm* ‘worm’ // Chad. W.: Hausa *dàman-ā* ‘the red velvety spider’; C.: Mofu *ḍa-ḍamíy-daw* ‘k of locust’ ◇ AADB 2696; cf. SED II No. 69.

Hadza *gaga* [gaga] ‘grasshopper (general term)’ (Miller et al. 2021: 150) ~ AA: Sem. \*gūg- ‘spider; flea’: Aram. Syr. *gəwāgay* ‘aranea’; Gz. *gug-ā* ‘flea’ ◇ SED II No. 77.

## II. Body parts, functions and states

Hadza *ʔakwisiti-ko* ‘sinew that runs along the spine and neck’ (Miller et al. 2021: 13) ~ AA \*kac/sw- ‘back with shoulders’: Cush. C. \*kas- ‘shoulder’: Bilin *kas*, Khamir *kesaa*, Qwara *kaš*; E. \*kas- ‘shoulder’: LEC: Konso *ħaš-ito*, HEC: Burji *káččoo*, Dullay: Gawwada *ħeššé* // Sem. \*ka/iskalis- (redupl.) and \*kVsʔ/w- ‘back, side or front part of the body (of a sheep?)’: Akk. *kaskas-u* ‘soft part of the sheep's breastbone’; Arab. *kusʔ-* ‘derrière, partie postérieure d'une chose; derrière de la tête’; Tgr. *kəšw-üt* ‘sheep's side’ // Chad. W. \*kawis-: Ankwe *kəšat* ‘shoulder’, Sha *ʔakweš*, Kulere *kašáw*, Daffo-Butura *kúnḡwíši* ‘buttocks’; E. \*kasy-: Kabalai *kāsi* ‘back’, Kera *kasi*, Kwang *kósi* ‘shoulder’ ◇ AADB 1718; cf. SED I No. 152.

Hadza *dushu* ‘distended, big stomach’ (Miller et al. 2021: 131) ~ AA \*dVs(-Vm/n)- ‘fat belly’: Sem. \*dašm/n-: Hbr. *dāšān* ‘fatness’; Aram. Jud. *došn-ā* ‘fat’; Arab. *dsm* ‘être gras’ // Berb.

<sup>48</sup> Cf. *unda’unda* [ʔunḍaʔunḍa] ‘hedgehog’ (Miller et al. 2021: 76), probably related.

<sup>49</sup> Note one more triconsonantal (!) Hadza-Chadic zoonym.

<sup>50</sup> Can *-ni* be a fossilized suffix?

\*-dis- ‘belly’: Ghadames *ta-dis-t*, Ntifa *a-dis*, etc. // Chad. W.: Hausa *dūsās-ā* (pl.) ‘corpulent’ ◇ AADB 2555; SED I No. 60.

Hadza *muguga* ‘crop of a bird’ (Miller et al. 2021: 294) ~ AA *mugaŋ*- ‘head with the neck’: Cush. N.: Beja *máge* ‘neck; nape of neck’; E.: HEC: Burji *muga* ‘head’, Dullay: Tsamay *múgaŋ-te* ‘head’ // Berb.: Siwa *ta-məʒa* (<\*mVga) ‘neck’, Timimun *ta-məg-n-a* ‘head’ ◇ AADB 3817.

Hadza *shububu-bi* ‘lungs’ (Miller et al. 2021: 432) ~ AA *\*ci/anp-* ‘lungs’: Sem. *\*si/anp-* ‘lung’: Akk. *sinib-t-u*, *sinip-t-u* ‘part of sheep’s lung’; pB Hbr. *simpōn* ‘ramified blood vessel, artery; bronchiae’; Gz. *sanbuʔ*, *sambuŋ* ‘lung’, etc. // (?) Eg. (MK) *šnb.t* ‘breast of man, chest’; cf. (Gr.) *snb* ‘breath’ // Chad. C.: Gude *atsemb-á*; E.: Dormo *ká-sibiŋ* ‘breast’ // Cush. *\*sanb-* ‘lungs’: N.: Beja *šamb-ut*; C. *\*sanb-*: Bilin *sānb-ii*, Aungi *saamb-ii*, etc.; E.: LEC: Somali *sambab*, Oromo *somb-a*, Arbore *soñb-ot*, HEC: Sidamo *šamb-oo*, etc. // Omot. N. *\*šVmb/p-* ‘heart, lung’: Dawro *šenfo* ‘heart’, *šemp-* ‘breathe’, Male *šempi* ‘soul’, Kafa *šomboo* ‘lung’, etc. ◇ AADB 164; SED I No. 235. Cf. Ainu *\*sanpe* ‘heart’ and Sino-Tibetan *\*sin(V)b-* ‘heart’.

Hadza *zze’a* [ts’eʔa] ‘to shit’ (Miller et al. 2021: 552) ~ Sem. *\*ṭayVʔ-* (*ṭ* < AA *\*ṭ*) ‘excrement, excreta’ ◇ SED I No. 286.

Hadza *bokê*, *bukî-ko* ‘sick, a sick person or animal’ (Miller et al. 2021: 91) ~ AA *\*baw/yVk-* ‘disease, illness’: Cush. E.: Afar *biyak* ‘illness’, LEC *\*buk-* Somali *buk-*, Boni *buš-i* ‘become ill’, Oromo *bokok-* ‘swell (of stomach)’ // (?) Sem.: Gz. *bak*, *bok* ‘scab, wound’ // Eg. (MK) *bṭ.w* (< *\*bVk-*) ‘Bez. einer unheilbaren Krankheit’ // Chad. E.: Bidiya *bàak*, pl. *bàkàw* ‘attraper un rhume, avoir de la fièvre’ ◇ AADB 2623; cf. Takács 2001: 353.

Hadza *nkoro-ko* [ŋkʰo.loko] ‘epilepsy’ (Miller et al. 2021: 330) ~ AA *\*kurVy-* ‘altered state of mind, ecstasy, trance, epilepsy’: Berb. Hgr. *tā-karaww-at*, Ayr *i-kerker-āni* ‘épilepsie’ (cf. E. Tawlemmet *a-tkar* ‘possessed (by an evil spirit)’ < *\*t-kVr?*), Qbl. *kerrer* ‘faire des sorcelleries, des sortilèges’ // Eg. (Dyn. 22) *kʔ* (likely < *\*kur-*) ‘soul’ // Chad. W.: Hausa *kūr-w-ā* ‘soul; ghost’ (Abraham 1962), ‘the soul or personality which is supposed to leave a sleeper, returning when he awakes’ (Bargery 1934; first compared to Eg. *kʔ* by C. Hodge), C.: Bura *kir* ‘the self; the will’; Sem.: Arab. *kry* ‘sommeiller’, Tgr. *kərri belä* ‘talk nonsense’ (cf. *sāb kəyar-āt* ‘soothsayers’) // Cush. C. *\*ʔan-kir-* ‘soul’: Bilin *ʔankəra*, Qwara *enker*, Aungi *enk-* (*\*ʔank-[ər]* in Appleyard (2006: 126); all the comparanda starting with Oromo are incompatible); E.: LEC: Oromo *e-keer-a* ‘ghost’ ◇ AADB 2700.

Hadza *kumba-ne* ‘to have a cold (stuffed nose, cough, sore throat)’ (Miller et al. 2021: 250) ~ AA *\*ka(m)b-* ‘cold’: Cush. N.: Beja *kanba*, *kamba*; C. *\*kalimb-* ‘cold, be cold’; E.: HEC: Burji *kabb-* ‘become cold’ ◇ AADB 3838.

### III. Varia.

Hadza *ʔutʰume-ko* ‘spear’ (Miller et al. 2021: 75) ~ AA *\*daʔVm-* ~ *\*ʔudum-* ‘spear’: Sem.: Akk. (YB) *daʔimu* // Chad. C.: Gude *ʔuduma*, Fali-Jilbu *wudumi*, etc. ◇ AADB 1918.

Hadza *slaa* [šaʔa] ‘to love’ (Miller et al. 2021: 434) ~ AA *\*šaʔ/w-* ‘want, wish, like, love’: Sem. *\*šxVyʔ-* (*\*šx* based on Hrs. *š*) ‘wish, want, need’: Aram. Jud. *swy* ‘to wish’, Syr. *səwē* ‘cupidus, studiosus’; Arab. *šyʔ* (also *šhw* and *šwh*) ‘to desire, aspire, wish’; Amh. *ša*, *eša* ‘to wish, want’; Hrs. *šáwweš* ‘want or need badly’ // Chad. C. *\*zV* ‘be happy, loving’: Podoko *ža* ‘rejoice’, Muyang *žayay* ‘to like’, Mbara *žà* ‘cher (affection)’ // Cush. S. *\*šaʔ-* ‘to like, love’: Iraqw, Alagwa *šaaʔ-*, Asa *šaʔ-at* ‘to like’, Dahalo *šaw-* ‘to love, like’ ◇ AADB 1962; Stolbova 2007 #396 (compared C. Chad. to S. Cush.)<sup>51</sup>.

Hadza *ts’ukuts’uku* [çukuçuku] ‘to smile’ (Miller et al. 2021: 560) ~ AA *\*çVhVk-* ‘to laugh’: Sem.: Ugr. *zħk*, Hbr. *šħk* ‘to laugh’; Hatra *šħk* ‘to laugh, smile kindly on so.’, Arab. *ḏħk*; Soq.

<sup>51</sup> A Hadza borrowing from S. Cush. is unlikely; borrowing verbs is a rare phenomenon in AA.

*dáhak* ‘to laugh’, etc. // Chad. \**šVVk/k-* ‘mock, laugh at’: W.: Ankwe *swak* ‘to exite, irritate, entice’, Tsagu *kàašá-n* (met.) ‘to laugh’; C.: Daba *šàk šàk* ‘to tease’  $\diamond$  AADB 2348; Stolbova 2007 #334 (compared to Sem.).

Hadza *bititî-bii* [bititi:bi?] ‘fierce’ (Miller et al. 2021: 88) ~ AA \**ʔVbVd-* ‘crazy’: Sem.: Eth. \**ʔVbVd-* ‘to be crazy’ // Berb.: Mzab *beddu* ‘perdre la raison, ê. fou’, Wargla *a-beddiw* ‘faible d’esprit, idiot, fou’, Nefusa *beddu* ‘ê. fou’  $\diamond$  AADB 3769.

Hadza *ihî’a* ‘thing’ [ʔihîʔa] (Miller et al. 2021: 46)<sup>52</sup> ~ AA \**ʔih-* ‘thing’: Eg. (PT) *ih.t* ‘thing’; Chad. W.: Bolewa *ʔya* ‘thing’; C.: Hwona *ʔya* ‘thing’  $\diamond$  AADB 1586; cf. Takács 1999: 42.

Hadza *akhana* [ʔak<sup>h</sup>ana] ‘name’ (Miller et al. 2021: 11) ~ AA \**kVnVy-* ‘call by name’: Sem.: Akk. *kunnû* ‘to treat a person kindly, to honor a deity’; Hbr. \**kny* (pi.) ‘to give so. a name of honor’; Aram. Syr. *kena?* ‘give a name’; Arab. *kny* ‘donner un surnom à qqn.’ // Eg. *kny* ‘call’ // Chad. \**kwa/un-*: W.: Ankwe *kun* ‘tell’, Zar *kwan* ‘say’; C.: Logone *kāwún* ‘speech’  $\diamond$  AADB 859.

Hadza *okoiya-bii* ‘whirlwind, dustdevil’ (Miller et al. 2021: 62) ~ AA \**kiw/yVH-* ‘wind, to blow’: Sem.: Arab. *kwh* II ‘souffler pour allumer (le feu)’ // Eg. (Pyr.) *tʔw* (if <\**kVʔVw*) ‘wind, air’ // Chad. \**kVwV(H)-* ‘to blow (wind)’: W. Tangale *kiu* ‘blowing (wind)’, Miya *kay* ‘to fan’; C.: Mulwi, Munjuk *kì* ‘to blow (wind)’; E.: Mobu *kwe* ‘souffler (vent)’, Ndam *ʔúgâ*, Somrai *kū* ‘to blow’ // (?) Cush. C. \**kuy-* ‘cloud’: Qwara *kuya*, Falasha *kúya*  $\diamond$  AADB 1580; Stolbova 2016 #411.

Hadza *thimbothimbo* [t<sup>h</sup>imbot<sup>h</sup>imbo] (redupl.) ‘dusk, twilight’ (Miller et al. 2021: 473) ~ AA \*(*ʔV*)*ʔum-* ‘darkness, night’: Sem.: Arab. *ʔʔm* V ‘devenir sombre, se couvrir de ténèbres (se dit de la nuit)’ // Cush. E.: Saho *ɖum-* ‘become dark’, LEC: Konso, Dirayta *ɖum-* ‘to set (of sun)’, HEC: Hadiya *ʔum-o* ‘darkness’ // Omot. \**ʔum-* ‘night’: N.: Male *ɖuum-i*, Kafa *ʔuum-o*, Gimira (Bench) *ʔum-*; S.: Ari *ɖuum-i*.

Such examples could be significantly increased.

## Analysis

Let us again list a number of arguments that would allow for a different interpretation than the alleged Hadza-Afrasian affinity:

(1) Hadza-Afrasian matches (attested only in African branches of AA) with parallels in non-Afrasian African languages from which these matches could have been borrowed by the corresponding AA languages and Hadza independently;

(2) detected borrowings into Hadza from the neighboring non-Afrasian African languages which, in turn, had earlier been borrowed from Afrasian;

(3) parallels that possibly go back to a much deeper chronological level than the one corresponding to the postulated Hadza affiliation with the Afrasian macrofamily, perhaps even reflecting certain “universal” roots (such as #30 ‘fly’: Hadza *pururu* ~ AA \**pir-*, with similar forms also attested in Kartvelian, North Caucasian, Sino-Tibetan, etc.);

(4) words in Hadza that meet both the phonetic and the semantic requirements of kinship with AA while specifically having reflexes in South Cushitic languages (above all, Iraqw), other cases of borrowing from which have been firmly established (the most etymologically challenging type of cases, making the choice between common origin and borrowing quite difficult);<sup>53</sup>

<sup>52</sup> Cf. also Proto-Central Khoisan \**xuu* ‘thing’.

<sup>53</sup> What is meant are not obvious cases, but those where differences in vocalism or word base structure or shades of meaning may indicate a common origin rather than borrowing (in which the borrowed lexeme is more often “copied” without changes).

(5) words in Hadza that meet both the phonetic and the semantic requirements of kinship with AA, but have "competing" matches in Sandawe and/or other Khoisan languages which may be interpreted as substratum lexemes.

To me, it seems that genetic affinity with AA explains such a large number of matches more plausibly (and economically) than even an unlikely combination of all these situations.

### Acknowledgments

To come to such a conclusion would be unrealistic without the high-quality research on the vocabulary of Hadza, conducted in recent years by Bonny Sands and Kirk Miller with the help of their wonderful Hadza co-authors and informants. I want to express my deep gratitude to these authors for sharing their unpublished materials, and also to Bonny Sands, Kirk Miller, George Starostin and Sergei Nikolaev for critical and very useful discussions via e-mail. It would be impossible to massively compare Hadza words to the vocabulary of languages of almost all branches and groups of the Afrasian macro-family without a large Afrasian etymological database compiled by Olga Stolbova and the present author. It would also be impossible to carry out the Hadza-Afrasian lexicostatic comparison without Kirk Miller's carefully calibrated Swadesh lists (Miller 2019) and the Afrasian lexicostatic database compiled by the author and making heavy use of the works of countless field researchers and of the advice of colleagues specializing in ancient texts. The author expresses his deep gratitude to all of them. And, of course, this is another occasion to remember with love and gratitude my late friend and informal teacher, the great linguist Sergei Starostin.

### Abbreviations

AA = Afrasian; Akk. = Akkadian; Amh. = Amharic; Arab. = Arabic; Aram. = Aramaic; BD = Book of the Dead; Berb. = Berber; Chad. = Chadic; C. = Central; Copt. = Coptic; Cush. = Cushitic; Dat. = Datoga; E. = East; Eg. = Egyptian; Eth. = Ethiopian; Gz. = Ge'ez; Hbr. = Hebrew; HEC = Highland East Cushitic; Hrs. = Harsusi; Jib. = Jibbali; Jud. = Judaic Aramaic; LEC = Lowland East Cushitic; Mhr. = Mehri; MK = Middle Kingdom; MSA = Modern South Arabian; N. = North; NE = New Kingdom; OB = Old Babylonian; Off. = Official Aramaic; OK = Old Kingdom; PAA = Proto-Afrasian; S. = South; Sab. = Sabaic; Sem. = Semitic; Soq. = Soqotri; Syr. = Syriac; Tgr. = Tigre; Tna. = Tigrinya (Tigray); Ugr. = Ugaritic.

### Transcription and transliteration

$\dot{p}$  = bilabial emphatic voiced stop;  $\dot{t}$  = dental emphatic voiceless stop;  $\dot{d}$  = dental emphatic voiced stop;  $\dot{d}$  = voiced interdental fricative (in Egyp., a conventional symbol conveying  $\mathfrak{ḏ}$ );  $c$  = alveolar voiceless affricate [ts];  $\mathfrak{z}$  = alveolar voiced affricate [dz];  $\mathfrak{c}$  = palato-alveolar voiceless affricate [tʃ];  $\mathfrak{z}$  = palato-alveolar voiced affricate [dʒ];  $\mathfrak{s}$  = hissing emphatic voiceless fricative;  $\mathfrak{c}$  = emphatic voiceless affricate;  $\mathfrak{c}$  = palato-alveolar emphatic affricate;  $\mathfrak{s}$  = lateral voiceless fricative;  $\mathfrak{c}$  = lateral voiceless affricate;  $\mathfrak{c}$  = lateral voiceless emphatic affricate;  $\mathfrak{z}$  = lateral voiced sibilant;  $\mathfrak{g}$  = voiced velar fricative (in Berb.);  $k$  or  $q$  = emphatic velar stop;  $\mathfrak{g}$  = uvular voiced fricative (Arabic "ghain");  $\mathfrak{h}$  = uvular voiceless fricative;  $\mathfrak{h}$  = pharyngeal voiceless fricative;  $\mathfrak{ʔ}$  = pharyngeal stop ("ayin");  $h$  = laryngeal voiceless fricative;  $\mathfrak{ʔ}$  = glottal stop ("aleph", "hamza");  $y$  = palatal resonant;  $\mathfrak{z}$  and  $i$  = conventional transcription symbols (in Egyptology).

### Conventions for reconstructed proto-forms

$V$  renders a non-specified vowel, e.g.  $*bVr-$  to read = "either  $*bar-$ , or  $*bir-$ , or  $*bur-$ ".

$H$  renders a non-specified laryngeal or pharyngeal.

S renders a non-specified sibilant.

/ when separating two symbols means “or”, e.g. \*ʔi/abar = “either \*ʔibar– or \*ʔabar–”.

() a symbol in round brackets means “with or without this symbol”, e.g. \*ba(w)r– to read = “\*bawr– or \*bar–”.

## References

- AADB = Militarev, Alexander, Olga Stolbova. 2020. Afrasian etymological database. Available online at: [starlingdb.org](http://starlingdb.org) [accessed 07.20.2020].
- Abraham, Roy Clive. 1962. *Dictionary of the Hausa Language*. London: University of London Press.
- Appleyard, David. 2006. *A Comparative Dictionary of the Agaw Languages*. Köln: Rüdiger Koppe.
- Bargery, George Percy. 1934. *A Hausa-English dictionary and English-Hausa vocabulary*. London: Oxford University Press.
- Blažek, Václav. 1994. Elephant, Hippopotamus and Others: On Some Ecological Aspects of the Afroasiatic Homeland. *Asian and African Studies* 3(2): 196–212.
- Blažek, Václav. 2008. A lexicostatistical comparison of Omotic languages. In: J. Bengtson (ed.). *In Hot Pursuit of Language in Prehistory*: 57–148. Amsterdam / Philadelphia: John Benjamins.
- Diakonoff, Igor, Olga Stolbova, Alexander Militarev. 1992. Proto-Afrasian Phonology. In: Diakonoff, Igor. *Proto-Afrasian and Old Akkadian*. *Journal of Afroasiatic Languages* 1991–92: 7–15.
- Dolgopolsky, Aron B. 1973. *Sravnitel'no-istoricheskaja fonetika kushitskix jazykov* [Comparative- Historical Phonetics of Cushitic]. Moscow: Nauka.
- Elderkin, Edward. 1978. Loans in Hadza: Internal evidence from consonants. *Occasional Paper, Dept. of Foreign Languages*, 3. University of Dar es Salaam.
- Elderkin, Edward. 1982. On the Classification of Hadza. *Sprache und Geschichte in Afrika* 4: 67–82.
- Greenberg, Joseph H. 1966. *The Languages of Africa*. Bloomington, Indiana University; Mouton & Co., The Hague.
- Jungraithmayr, Herrmann, Dymitr Ibrizimow. 1994. *Chadic Lexical Roots. Vols. I–II*. Berlin: Dietrich Reimer.
- Leger, Rudolf, Olga Stolbova. 2021. Internal -k- Plurals in Kupto and in Chadic. In: Joseph A. McIntyre, Henning Schreiber (eds.). *Topics in Chadic Linguistics VIII. Papers from the 7<sup>th</sup> International Colloquium on the Chadic Languages, Hamburg, 12-14 September 2013*: 37–52. Köln: Rüdiger Köppe Verlag.
- Militarev, Alexander. 2023. Non-mainstream considerations by a comparative linguist of the Biblical story about the ‘Sons of Israel’ in Egypt and internal Biblical chronology. Ms. Available at [academia.edu](http://academia.edu), [ResearchGate](http://ResearchGate).
- Militarev, Alexander. 2023. Forthc. Common Afrasian (Afro-Asiatic) terms related to the magic, supernatural, spiritual and mythic. Etymologies and reconstructions. *Kervan — International Journal of Afro-Asiatic Studies* 27 (Special Issue).
- Militarev, Alexander. 2022. Reconstructing a cultural lexicon for pre-history: Berber zoonyms of Afrasian (Afro-Asiatic) origin. *Asian and African Studies* 31(1): 1–47.
- Militarev, Alexander, Sergei Nikolaev. 2020. Proto-Afrasian names of ungulates in light of the Proto-Afrasian homeland issue. *Journal of Language Relationship* 18(3): 199–226.
- Militarev, Alexander, Sergei Nikolaev. 2021. Proto-Afrasian names of non-ungulate animals in light of the Proto-Afrasian homeland issue. *Journal of Language Relationship* 19(4): 233–262.
- Militarev, Mikhail. 2021. *Hadza, a genetically isolated group of Tanzanian hunter-gatherers and the problem of their linguistic affiliation*. Competitive project submitted to 2021 Davidson Institute Fellows Scholarship. Ms.
- Miller, Kirk (ed.), w. Mariamu Anyawire, G. G. Bala, Bonny Sands. 2021. *A Hadza Lexicon. II: Lexical Dictionary. III: Grammatical Dictionary*. Ms.
- Miller, Kirk. 2021. *A Hadza Lexicon. Appendix IV: Swadesh lists*. Ms.
- Nait-Zerrad, Kamal. 1998–2002. *Dictionnaire des racines berbères*. Paris-Louvain: Peeters.
- Pennisi, Elizabeth. 2004. The First Language? *Science* 303: 1319–1320. doi: 10.1126/science.303.5662.1319
- Sands, Bonny. 2012. *Hadzabe Vocabulary*. Draft Ms. (compiled in 1991–1992).
- Sands, Bonny, Ian Maddieson, Peter Ladefoged. 1996. The phonetic structures of Hadza. *Studies in African Linguistics* 25(2): 171–204.
- SED I = Militarev, Alexander, Leonid Kogan. 2000. *Semitic Etymological Dictionary*. Vol. 1: *Anatomy of Man and Animals*. Münster: Ugarit-Verlag.
- SED II = Militarev, Alexander, Leonid Kogan. 2005. *Semitic Etymological Dictionary*. Vol. II: *Animal Names*. Münster: Ugarit-Verlag.

- Starostin, Sergei. 2000. Comparative-historical linguistics and lexicostatistics. In: Colin Renfrew et al. (eds.). *Time Depth in Historical Linguistics*: 233–259. McDonald Institute for Archaeological Research: Oxford Publishing Press.
- Starostin, George. 2008. K voprosu o geneticheskoj prinadlezhnosti jazyka hadza [On the issue of the genetic affinity of Hadza]. In: V. F. Vydrin (ed.). *Afrikanskiy sbornik 2007*: 262–278. Saint Petersburg: Nauka. 262–278.
- Starostin, George. 2010. Preliminary lexicostatistics as a basis for language classification. A new approach. *Journal of Language Relationship* 3: 79–116.
- Starostin, George. 2012. *Annotated Swadesh wordlists for the Hadza group (Hadza family)*. [Text version of database, created 19/12/2012]. Ms., available online at: [starlingdb.org/new100/hdz.pdf](http://starlingdb.org/new100/hdz.pdf).
- Starostin, Georgij S. 2013. *Jazyki Afriki. Opyt postrojenija leksikostatisticheskoy klassifikatsiji. T. I. Metodologija. Kojsanskije jazyki* [Languages of Africa: an attempt at a lexicostatistical classification. Volume 1: Methodology; Khoisan languages]. Moscow: Jazyki slav'anskoj kultury.
- Stolbova, Olga. 2007. *Chadic lexical database. Issue II*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Stolbova, Olga. 2009. *Chadic lexical database. Issue III*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Stolbova, Olga. 2011. *Chadic lexical database. Issue IV*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Stolbova, Olga. 2016. *Chadic Etymological Dictionary*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Stolbova, Olga. 2019. *Chadic lexical database. Issue V*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Stolbova, Olga. 2021. *Chadic lexical database. Issue VI*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Takács, Gábor. 1997. The Common Afrasian nominal class marker \*h. *Studia Etymologica Cracoviensia* 2: 241–273.
- Takács, Gábor. 1999. *Etymological Dictionary of Egyptian. Vol. I*. Leiden: Brill.
- Takács, Gábor. 2001. *Etymological Dictionary of Egyptian. Vol. II*. Leiden: Brill.
- Takács, Gábor. 2008. *Etymological Dictionary of Egyptian. Vol. III*. Leiden: Brill.
- Takács, Gábor. 2011. *Studies in Afro-Asiatic Comparative Phonology: Consonants*. Berlin: Dietrich Reimer Verlag.
- Tishkoff, Sarah, A., Mary Katherine Gonder, Brenna M. Henn, et al. 2007. History of Click-Speaking Populations of Africa Inferred from mtDNA and Y Chromosome Genetic Variation. *Molecular Biology and Evolution* 24 (10): 2180–2195. doi:10.1093/molbev/msm155

A. Ю. Милитарев. Хадза — афразийский язык?

В статье рассматривается вопрос о генетической принадлежности хадза, языка крошечного племени танзанийских охотников-собирателей, генетически отделившегося от других групп Homo sapiens несколько сот тысяч лет назад и занимающего одну и ту же территорию более 50 000 лет; ранее язык считался койсанским из-за его фонологического инвентаря, содержащего так называемые «кликсы», но теперь рассматривается как изолированный. В статье к обширному лексическому материалу, собранному американскими лингвистами с помощью их сотрудников-хадза, приводятся параллели из различных афразийских (афроазиатских) языков, основанные на установленных автором регулярных хадза-афразийских консонантных соответствиях и строгих семантических критериях. Проводится сравнение слов хадза с предлагаемыми афразийскими параллелями из стандартного стословного списка Сводеша (включая разработанное Сергеем и Георгием Старостинскими подмножество из 50 пунктов наиболее устойчивых элементов), а также из других семантических групп. Автор анализирует альтернативные объяснения этих совпадений (например, случайные совпадения; заимствования в хадза из соседних и даже географически удаленных афразийских языков; общий субстрат), но приходит к выводу, что наиболее правдоподобным объяснением является генетическое родство. В соответствии с полученными лексикостатистическими данными

ми положение хадза в афразийской макросемье оказывается в примерно одинаковой близости к омотской и кушитской семьям; глоттохронология датирует разделение между прото-хадза, прото-кушитским и прото-омотским языками рубежом X–IX тысячелетий до нашей эры, когда группа носителей праюжноафразийского языка, по-видимому, добралась до Северной Танзании и передала свой язык предположительно койсаноязычным предкам современных хадза.

*Ключевые слова:* хадза язык; афразийские языки; генетическое родство языков; звуковые соответствия; лексикостатистика; этимология; лексические заимствования.



Bonny Sands<sup>†</sup>, Andrew Harvey<sup>‡</sup>, Maarten Mous<sup>§</sup>, Mauro Tosco<sup>††</sup>

<sup>†</sup>Northern Arizona University; bonny.sands@nau.edu

<sup>‡</sup>University of Bayreuth; andrew.harvey@uni-bayreuth.de

<sup>§</sup>Centre for Linguistics, Leiden University; m.mous@hum.leidenuniv.nl

<sup>††</sup>University of Turin; mauro.tosco@unito.it

## Why Hadza is (probably) not Afroasiatic: a discussion of Militarev’s “Hadza as Afrasian?”

Problems with the lexical evidence used to posit Hadza as an Afroasiatic language are discussed. The failure to identify certain loanwords and the misanalysis of certain Hadza morphemes are problems rectified by having greater familiarity with Hadza and its linguistic contact history. Other problems are more general to the specific methodology employed. The overly wide semantic range of meanings often allowed in establishing form-meaning sets increases the likelihood of chance resemblances. The use of certain words that are likely onomatopoeic also reduces the impact of the proposed cognate sets. Ultimately, it is the lack of regular, repeated sound correspondences between Hadza and Afroasiatic that makes the proposal of their familial relationship unconvincing.

*Keywords:* language isolates; Hadza language; Khoisan languages; Cushitic languages; Afroasiatic languages; Chadic languages; long-distance relationship.

### 1. Introduction

There have been three main linguistic genetic classifications of Hadza that have been proposed: 1) linguistic isolate, 2) Khoisan language, 3) Afroasiatic language. There is currently a mainstream acceptance among Africanists in labeling Hadza as an isolate (Sands 1998, Güldeman 2014, Starostin 2012). Honken (2013: 20) states that “Hadza is currently regarded as an isolate by nearly everyone”, though the Khoisan classification for Hadza continues to be used by some researchers (e.g. Chebanne 2017, Elderkin 2014, Ehret 2013a). Suggestions that Hadza has affinities to Afroasiatic have been made over time and include: Tucker (1966, 1967), Elderkin (1982), Starostin (2008) and now Militarev (this volume).

In this paper, we review the evidence proposed by Militarev (this volume) but in the interest of space do not include a critique of the similarities noted by Tucker (1966, 1967) and Elderkin (1982). We argue that the arguments put forth by Militarev are not sufficiently convincing proof of a linguistic genetic relationship between Hadza and Afroasiatic. We fail to see evidence of regular, repeated sound correspondences; nor do we see a convincing number of grammatical morphemes whose similarities suggest relatedness.

We provide examples of alternative analyses of many of the comparison sets, drawing upon our extensive first-hand experience with Hadza (Sands, Harvey) and Cushitic languages (Tosco, Mous, Harvey) as well as our knowledge of Bantu and Nilotic languages. First, our knowledge of Hadza allows us to show that certain Hadza terms are incompatible with the Proto Afroasiatic (AA) forms that Militarev has connected them to. In many cases, we provide alternative citation forms and in some cases we disagree with his morphological analysis. Note that our transcriptions fail to mark tone in many cases though tone is contrastive in the language<sup>1</sup>.

---

<sup>1</sup> Jeremy Coburn (Indiana University) is currently researching tone in Hadza and we (Harvey & Sands) hope to collaborate with him in producing a fully tone-marked Hadza dictionary (cf. Coburn et al. forthcoming). As

Second, our knowledge of languages surrounding Hadza often allows us to identify loan sources that provide better etymologies than the AA forms suggested.

Our critique also discusses the sound correspondences and semantic correspondences proffered by Militarev, arguing that these do not meet the standard of evidence necessary to conclude that Hadza *must* be an Afroasiatic language. Problems of morphological analysis, semantic laxity and sound correspondences cannot be evaluated entirely separately because these factors all interact. In many instances, only a portion of a Hadza root, or only a portion of a reconstructed AA root are directly comparable yet no analysis is proffered to explain the differences in the roots. We analyze many of these cases as simply being due to chance resemblance. Furthermore, we think that Militarev has underestimated the role of contact in contributing to sets of similar words.

Given these methodological challenges of chance resemblances it is vital to restrict ourselves in how to go about establishing cognate sets. One sound restriction would be to compare only to solid reconstructed families. Militarev uses the impressive and extensive Afroasiatic reconstructions in an openly accessible database (available at the Tower of Babel website: <https://starlingdb.org/>). Reconstructions by others at various levels of families and subfamilies are not used nor cited in the article. In this contribution, we cannot assess the reconstructions from his database and we can only work under the assumption that the cited reconstructions are solid. Restricting the cognate sets to only reconstructed forms in (sub)families would reduce the danger of chance resemblance. The amount of cognate sets would be reduced although presumably it would be possible to propose new (sub)family reconstructions for some of the sets.

## 2. Problems with the Data

### 2.1. Citation Forms

A multilateral comparison involving hundreds of languages necessitates the use of sources that the author does not have first-hand experience with. This is not to say that such comparisons should not be done, rather, care should be taken to use the most recent and authoritative sources on each language; sources cited should withstand scrutiny from specialists on those individual languages.

For Hadza, Militarev uses a draft dictionary compiled by Kirk Miller (Miller et al. 2021); we have this same draft dictionary but it has a 2013 date and has not been revised to 2021. Miller et al.'s orthographic system has not been explained by Militarev, though he does present phonetic transcriptions in some cases. Miller et al. represent ejectives with symbols for voiced consonants (e.g. *pedla* 'white' (*petl'a-* in Sands 2012) (more properly cited as /**pecʎa-**); *jjowa* ['tʃoa] 'gecko') which might be distracting for readers.

Forms in this paper are modified from Miller et al. (2013) draft dictionary. We provide phonemic transcriptions based on the system used in Coburn, et al. (2023). In some cases, translations are provided based on our own Hadza research. We also refer to Peterson (2013) because his plant and animal identifications have been made in consultation with biologists.

Just as the Hadza data cited leaves something to be desired, so too, the Afroasiatic sources cited by Militarev are not always the most recent or authoritative. For a large part, the data on individual languages are from old sources. For instance, Militarev primarily sources Somali

---

tone does not enter into the discussion of Afroasiatic comparanda, we will not discuss it further here, apart from noting that it would need to be accounted for in any historical reconstruction.

data from Reinisch (1902) rather than use Zorc & Osman (1993) & *Dizionario Somalo-Italiano*. For Beja data, Reinisch (1895) is used rather than Wedekind et al. (2007).

It is not always possible for us to evaluate the data because the original sources (and therefore, the phonological interpretations) are not always cited. For instance, a language name Darasa is cited, without reference to the original source. This is an older, derogatory name for the language that has been consistently referred to as Gedeo (or Gede'o) since the 1970s. We consider Wedekind's Gedeo Dictionary (2008) and/or Hudson's Gedeo-English Index (1989: 229–265) of his *Highland East Cushitic dictionary* to be more authoritative sources for this language.

More generally, it is evident that there are degrees of reliability using primary sources. In particular, one needs to be hesitant to use words collected from languages that were no longer comfortably spoken and for which the speakers had already shifted to another language. This is the situation for the Cushitic languages Aasáx, Qwadza, Yaaku. Unfortunately, adding to lack of reliability, the earliest collections for these languages were done by non-linguists.

## 2.2. Alternative Morphological Analyses

There are a number of cases where words are cited without sufficient morphological analysis. We present some cases where we disagree with Militarev on the identification of the morpheme(s) in question.

**#62 'not' 'ukuwa** (more properly cited as /ʔuk<sup>h</sup>u-/ 'to not exist') ~ AA \*(ʔa-)k<sup>w</sup>ay- 'not'

Militarev glosses this form as 'not', whereas Miller (2013: 37–38) lists the forms [ʔuk<sup>h</sup>u] as 'to not exist', and [ʔuk<sup>h</sup>uwa] as 'to lack' or 'to not have'. Not only are the semantics of Hadza 'ukuwa different from the characterisation given in Militarev, but 'ukuwa is morphologically complex, formed of /ʔuk<sup>h</sup>u/, plus =a, a form of the copula. As such, the form to be used in comparison would be /ʔuk<sup>h</sup>u/, which seems rather more different to the forms offered by Militarev here (Proto-Afroasiatic \*(ʔa-)k<sup>w</sup>ay- 'not', and the \*kway forms of Proto-Chadic and Proto-Omotiic).

It should also be added that the Iraqw suffix offered by Militarev here (written in his work as -Vkã, where the diaeresis on the a is unexplained), probably refers to the verbal negative suffix -Vkã. This can in turn be internally reconstructed in Iraqw to the lexical verb kãhh [ká:h] 'to be absent', and itself cannot be given as cognate to the Proto-Cushitic reconstructed prefix \*ka- offered here.

**#96 'what?' akwiʔa and #98 'who?' akwaza** (more properly cited as /ʔaku-/ 'which?')

Militarev presents the forms akwi'a 'what?' and akwaza 'who' as morphologically simple, when, in fact, both are morphologically complex. The form akwi'a [ʔak<sup>w</sup>iʔa] is formed of the interrogative hadza aku [hadza ʔaku] 'person', and ihia [ʔiʔiʔa] 'thing', (cf. Miller 2013: 8) and could just as easily be translated as 'which thing?'. The form akwaza [ʔak<sup>w</sup>adza] is formed of the same interrogative aku, and haza 'person', and could just as easily be translated as 'which person?'. As such, the form to be used in both comparisons would be aku.

This does not seem to challenge the forms offered as cognates (e.g. Proto Afroasiatic \*k<sup>w</sup>ay- 'what? who?'), but does weaken the overall argument – after all, the cognacy here rests on one consonant, [k].

**#34 'good' muta-na, mta-na** (more properly cited as /mtana ~ mutana/ 'fine, good')

While Miller (2013:177) presents this form as mut<sup>=</sup>ana<sup>2</sup>, it is unclear as to why Militarev presents this form as multimorphemic muta-na in this way. At any rate, the form mutana (also

<sup>2</sup> The diacritic after the stop indicates that it has been confirmed as being unaspirated.

commonly heard as *mtana*) provides no evidence that it is (or was at any point in Hadza) a multimorphemic form. We analyze this root as a Bantu loan, discussed below in Section 3.

### 2.3. Evidence from Internal Reconstruction

It appears to be the case that some body part words in Hadza feature a fossilized form of the possessive enclitic =*kwa* (the 1st person singular possessive in contemporary Hadza).

This may possibly be the motivation for the presentation of #56 ‘mouth’ *awani-ka* [ʔawani-ka], where the *-ka* is parsed from the rest of the word and disregarded in the offered cognates (e.g. the Proto-Afroasiatic reconstruction \*ʕawin- ‘tongue; (part of) mouth’). This word is more properly cited as /ʔawanika-ko/. If *ka* is a separate morpheme, it would probably derive from enclitic =*kwa*.

Where this fossilized =*kwa* seems to have been taken into account for ‘mouth’, Militarev seems to have missed it in both #31 ‘foot, leg’ *aphukwa* [ʔap<sup>h</sup>uk<sup>w</sup>a] or *uphukwa* [ʔup<sup>h</sup>uk<sup>w</sup>a], as well as in #37 ‘hand’ *ukhwa-* [ʔuk<sup>w</sup>h<sup>a</sup>-] (and ‘fingers’ with the plural suffix *-bii* [biʔi], *ukhwabii* [ʔuk<sup>w</sup>h<sup>a</sup>-biʔi]). For ‘foot, leg’, the form to be used in comparisons should probably be something like /ʔap<sup>h</sup>u/ or /ʔup<sup>h</sup>u/, which differs rather greatly from the cognates offered. For ‘hand’ (and ‘fingers’), the form to be used in comparisons should probably be something like /ʔufu/ – again, entirely different from the cognates offered (e.g. the reconstructed Proto-Afroasiatic \*kawiʕ- ‘claw, fingernail, hand’).

### 2.4. Imprecise Translations

‘rhino’ *losho* (more properly cited as /lo:ʃo/ ‘sighting/hunting name for the rhinoceros’ (when it is stooped and ready to charge)) ~ AA \*warš- ‘rhinoceros’

Militarev glosses this form as ‘rhinoceros’, but Miller (2013: 158) gives this form not as ‘rhinoceros’, but as the sighting or hunting name for the rhinoceros (the more typical name of the rhinoceros in Hadza is *tlhakate* /cʔ<sup>h</sup>akate/). This distinction is crucial here, as these special hunting names of animals in Hadza are not nouns, but verbs (see also Blench 2013a). In this case, the source verb is /lo:ʃo/ ‘to stoop’. This would be consistent with other animal hunting names, which often characterize the animal in some way (the hunting name for ‘ostrich’ comes from a verb meaning ‘to puff up’, for example, and the hunting name for ‘baboon’ comes from ‘thirst’ (in reference to a baboon’s concave stomach)). In this case, the verb ‘to stoop’ would seem to evoke the posture of the rhinoceros, especially its stance when about to charge. The verbal nature of /lo:ʃo/ is further reinforced in that this term has no plural form. As such, any comparison of the (verbal) hunting name for ‘rhinoceros’ /lo:ʃo/ with (nominal) words for ‘rhinoceros’ in Afroasiatic (e.g. reconstructed Proto-Afroasiatic \*warš-) is misguided.

#32 ‘to be many, to be plenty, to be full’ *furu-ne* (more properly cited as /furune/ ‘to be many, to be plenty, to be a lot’) ~ \*ʕVpVr- ‘full’

The basic meaning of the Hadza root is ‘to be many’. The meaning ‘to be full’ is a secondary one, and would not be used to refer to a full cup of water or a full belly (the verb /ʔoʔa/ ‘to be full’ would be used in these cases). It appears that the Afroasiatic terms compared all mean ‘be full’ and not ‘be many’, but it is difficult to confirm this as the AA \*ʕVpVr- ‘full’ root does not appear in our searches of the StarLing Database. Unless it can be shown that a reconstructed AA form shares the basic meaning ‘to be many’, we do not think the roots can be considered to be related.

There are additional issues with the comparison. Militarev presents this form as *furu-ne*. Phonologically, it is important to note that Hadza does not feature the phoneme [r], and the form actually has a tap/flap (thus: [furune]). Morphologically, the word *furune* (or *furuni*) is

monomorphemic in Hadza, and though Miller offers the Iraqw and Proto-West Rift form *xooroo* 'crowd, community, people in group' for comparison (Miller 2013: 83–84), there is no evidence that the word can or ever could be analyzed as some morpheme *furu* and some morpheme *-ne* in Hadza. As such, the form to be used in comparison ought to be *furune* [furune] (or the alternative form *furuni* [furuni]). This makes comparisons with forms offered by Militarev (such as the reconstructed Proto-Afroasiatic \*ʕVpVr- 'full') rather less convincing.

The Hadza and Afroasiatic forms are not directly comparable because no analysis is presented explaining why the initial syllable is missing in the Hadza form. An AA \*ʕ corresponds to a (medial) Hadza glottal stop in #1 'all'. A medial AA \*ʕ (AA \**ma-ʕi(n)ʕ*- 'k. of bovid') appears to be deleted in Hadza *minza* (more properly cited as /mínza/ [míndza] 'reedbuck'). No explanation is given as to why there might be different reflexes of AA \*ʕ in Hadza.

In addition to this, it has been argued that the voiceless labiodental fricative [f] is a recent addition to the Hadza consonant inventory, and that most if not all Hadza words containing [f] can be shown to be borrowings or mimetic (Harvey 2021). If this were the case, any comparison with Afroasiatic forms would be inapplicable.

## 2.5. Ma'a Data

There are languages that one should use with utmost care in building cognate sets. A prima example is Ma'a / Mbugu. This language is known as a mixed language. Mous (2003) uncovers how this "strange case of Mbugu", as Goodman (1971) has put it, came about. Briefly, the language is a Bantu language, with an extra parallel lexicon. This parallel lexicon has partly been built consciously in an attempt to re-create a Cushitic language that was lost due to shift. The extra lexicon shows words from two different Cushitic sources, from Maasai, from Taita Bantu and manipulated words from the regular lexicon (Pare). Taking words from Ma'a / Mbugu requires that one knows from which of the two lexica it comes and, if from the mixed parallel lexicon, from which source. An additional challenge is that even the basic Pare lexicon contains borrowings from Cushitic as all Bantu languages in the area do. No evidence from Ma'a should be used to build a cognate set on; it can only be taken as additional evidence.

## 3. Lexical Sets Better Analyzed as Loans

A number of Militarev's lexical sets arguing for an Afroasiatic membership of Hadza are better analyzed as instances of loans from different languages and language families (especially Bantu, Nilotic and West Rift Cushitic). As such, they are certainly interesting for the history of Hadza, but they can hardly be taken as proof of genetic relationship. We expect that there may be more cases of loans that we cannot yet confirm due to currently available, limited resources on languages such as Ihanzu (Bantu) and Datooga (Nilotic), and the timing of the loans is not just of a recent date.

### 3.1. Bantu

#### #16 *dza* 'come'

Hadza *dza* < Ihanzu *dza*

The Hadza term may be straightforwardly interpreted as a loan from the neighboring Bantu language Ihanzu (Harvey 2019). The concept 'come' does colexify with 'go' in CLICS<sup>3</sup> (Rzymiski et al. (2019), so it is possible that an Afroasiatic term meaning 'go' might be related to a term meaning 'come', yet the borrowing analysis is, in our opinion, the most likely scenario.

There is a different Hadza root *botʃ<sup>h</sup>o* ‘come’ used in imperative forms that has no connection to Afroasiatic (that we are aware of), though it can be plausibly linked to a similar irregular root used in the imperative in Bantu languages (c.f. Ihanzu *nzuu* ‘come!’, or Swahili *njoo* ‘come!’).

**Section 1.3.2. ‘large male lion’ *mondo*** (more properly cited as /móndo/)

Miller (2013: 173) notes that this form is ultimately from a Bantu source. *\*mondo* ‘tiger-cat, serval’ (Bastin et al. 2002: 6665) is a form with distribution across northeast Bantu. This seems to be more plausible than the cognates offered by Militarev.

**Section 2 ‘epilepsy’ *nkoro-ko*** (more properly cited as /ŋk<sup>h</sup>oro-ko/)

The word for ‘epilepsy’ in the neighbouring Bantu language Ihanzu is *nkólo* (Harvey 2019). Cf. also *\*kódò* ‘heart’, with distribution across northeast Bantu (Bastin et al. 2002: 1889).

**#34 ‘good’ *muta-na, mta-na*** (more properly cited as /mtana ~ mutana/ ‘fine, good’)

As mentioned above, there is no synchronic (or diachronic) evidence for the morpheme breaks in this word proposed by Militarev. From a semantic perspective, Militarev provides the gloss for *mutana* as ‘good’ (adjectival), whereas Miller (*ibid.*) provides the gloss as both ‘good’ (adjectival) and ‘fine’ (adverbial). In fact, in the author’s work (Griscom and Harvey 2020), the use of *mutana* in an adjectival sense is absent, and only the adverbial sense is recorded. The root used for ‘good’ in an adjectival sense is *nube-* ~ *nubi-*.

Phonetically, a preponderance of words in Hadza beginning with prenasalized stops have been shown to be borrowings from Bantu languages. Despite Militarev’s assertion to the contrary (fn. 15), it is highly likely that *mtana* or *mutana* also has a Bantu source.

**Hadza *minza* ~ AA *\*ma-ʕi(n)ʒ-*** ‘k. of bovid’ (more properly cited as /mínza/ [míndza] ‘reedbuck’)

Even if we were to accept the dubious semantic correspondences between a wide-range of cloven-hoofed ungulates: ‘cow’, ‘he-goat’, ‘female topi’<sup>3</sup>, etc. and Hadza *mínza* ‘reedbuck’, there is a very strong likelihood that the Hadza form is a loan, as is demonstrably the case for many words which contain a prenasalized obstruent (cf. Elderkin 1978). Another word for ‘reedbuck’ /ndofeda/ (Peterson 2012) is also likely a loanword for the same reason.

**Hadza *ndama* ‘calf’**

This is from Cushitic (Nurse and Hinnebusch 1993: 304), also borrowed into Bantu, and Hadza must have gotten it from Bantu due to the prenasalized stop; cf. *ndama* ‘calf’ Sukuma, Swahili.

### 3.2. Nilotic

**Hadza *gondera* ‘dog’ < *gudeeda* ‘dog’ in Datooga.**

This is clearly a loan from Datooga *gùdéd* (Griscom 2019: 93). Note that the Hadza word is not the most commonly used word for dog, /||<sup>h</sup>aʔano/, nor is it the second most commonly used word /ti<sup>h</sup>inggi/.

### 3.3. West Rift (Southern Cushitic)

**#80 Hadza *ntsa-ko* ‘star’ ~ AA *\*(t/wV-n)ciʕ(cíʕ-)***

Militarev gives Iraqw *cacēʕ* as evidence for the Hadza form to be Afroasiatic. The Iraqw root is *ts’atsa’* (m) ‘stars’ (Mous et al. 2002); PWR *\*ts’its’aʕu* (m) ‘stars, starlit sky’ (Kießling &

<sup>3</sup> To this list might be added Melo (Omoti) *mintsa* ‘buffalo’ (Seibert & Caudwell 2002) and Shabo (isolate) *mijad* ‘buffalo’ and *mindza* ‘female cow’ (Kibebe 2015: 443)..

Mous 2003) but an even closer cognate would be Proto West Rift *tsaaʔa-s* (causative verb) 'shine, shed light'. The initial nasal in Hadza reminds us of a Bantu class 9 noun class prefix but there is no Bantu source in the vicinity with a similar word for 'star'. It is possible that this root came into Hadza from Cushitic via a Bantu intermediary.

**Hadza *slaa* [ʃaʔa]** (more properly cited as /ʃaʔa/) 'to love' ~ AA \*ʃaʔ/w- 'want, wish, love, like'

This root is indeed attested in Proto West Rift *ʔaʔ* 'live, like, want' (Kießling & Mous 2003) and we agree with Militarev that borrowing from West Rift cannot be ruled out. The suggested cognates with Iraqw work equally well with Proto West Rift (Kießling & Mous 2003) and there is very little current lexical transfer between Iraqw and Hadza. One should keep in mind that it was only a few decades ago that the Iraqw moved into the vicinity of the Hadzabe. It is not unlikely, however, that the speakers of Pre-Proto West Rift were in contact with Hadza during their movements into Tanzania.

#### 4. Problems with Semantic Correspondences

Semantic correspondences for alleged cognates are sometimes baffling. For instance, set #56 includes meanings as diverse as: 'mouth', 'tongue', 'cheek' and 'muscle in the neck'. These are all compared without further justification and we find these semantic correspondences to be rather implausible. By allowing such wide semantic comparisons, Militarev increases the likelihood that chance resemblances occur.

We also call into question the semantic correspondences in set #74 which include diverse meanings: 'to sit', 'to stand' and 'to sleep'. These are all quite different things. While the correspondence between 'hand' and 'finger' may be accepted, it still looks rare (cf. #37). The correspondences in the second part of the paper between different types of animals are also concerning.

We would recommend that the types of allowable semantic comparisons be limited to the prototypical types of semantic change mentioned in introductory textbooks on historical linguistics or historical semantics (e.g. widening, narrowing, metonymy, etc.). It is not sufficient to propose cognate sets based on two glosses occurring in the same semantic field. We do not find it terribly plausible that a term for a gazelle or a ram somehow semantically changes meaning to allow reference to a buffalo, or that terms for herons or cranes might somehow end up referring to ostriches.

#1 is a typical example in many ways. We doubt that the correspondence between 'all' (Hadza) and 'one' (all the rest) is semantically plausible.

**#1 Hadza *waʔi* 'all'** (more properly cited as /waʔi-/ 'all of X')

The Hadza root *waʔi-* is not a noun and must take an affix.

The concept 'all' may colexify with 'every' or 'many' but does not colexify with 'one' in any language in the CLICS<sup>3</sup> Database of Cross-Linguistics Colexifications (Rzymiski et al. 2019, accessed 27 June, 2023)

**#56 *awani-ka* 'mouth'** (more properly cited as /ʔawanika-ko/ 'mouth') ~ AA \*ʔawin- 'tongue; (part of) mouth'

As discussed above in Section 2.3, we do not agree with the segmentation of the word into a root *awani*<sup>4</sup>. Besides our disagreement over the morphological segmentation, there is an is-

<sup>4</sup> Miller internally reconstructs \*ʔawa 'mouth' for Hadza based on /ʔawanika/ 'lower lip', /ʔawanika-ko/ 'mouth, beak' and notes the similarity to Proto West Rift \*ʔafa 'mouth'. It is possible to internally reconstruct \*ʔawa-ni-k(w)a (along with \*ʔawa-ti 'upper lip') but not \*awani.

sue with the equivalence of the semantics. The concept ‘mouth’ colexifies with a number of different concepts in CLICS<sup>3</sup> (Rzymiski et al. (2019) including: ‘lip’, ‘tooth’, ‘beak’, ‘language’, ‘edge’, ‘door’, ‘face’, ‘chin’, ‘word’, ‘mother’, ‘jaw’, ‘hole’, ‘throat’. We are dubious that the concepts ‘mouth’ and ‘tongue’ (or ‘inside of cheek’) might be connected through semantic change from a common meaning.

**Hadza *geweda-ko* ‘dikdik’** (more properly cited as /géwéda-ko/ ‘dik-dik’)

The Hadza word is compared to Afroasiatic words for various ungulates including ‘greater kudu’, ‘buffalo’ and ‘duiker’. Dik-diks are tiny antelopes, very visually and functionally distinct from the kudu and the buffalo, one of the most dangerous animals on the continent. Although it is not impossible for a word to come to refer to different ungulates over time, we are concerned that the great semantic laxity in such a comparison *greatly* increases the likelihood that a chance resemblance may have occurred. The AA form *\*guday* provides a match for two of the three consonants in the Hadza root, but readers are left to imagine for themselves how the two roots might be related through a series of sound changes. The sound changes that would be necessary to connect the roots seem to us to be implausible and not supported by other correspondence sets as a regular, repeated correspondence. The shape of the Hadza root (especially the final *da* syllable) immediately bring to mind the possibility that the word may be a Southern Nilotic loan<sup>5</sup>. The relatively poor documentation of Datooga lects in the present day makes it impossible to conclude that the Hadza word could not be a loan. Although not proveable at this time, we feel this alternative hypothesis is no more unlikely than the one set forth by Militarev.

**Hadza *garaani* ‘heron’ ~ AA ‘ostrich’** (more properly cited as /garaʔani-ko/ [galaʔani-ko] ‘black-headed heron’)

There is very little to semantically connect a heron with an ostrich other than the fact that both birds have relatively long legs.

**Hadza *gaga* ‘grasshopper’** (more properly cited as /gagá/)

There is very little to semantically connect a grasshopper with a spider or flea. This word is best analyzed as a loan from the Bantu language Ihanzu; the Ihanzu word for ‘grasshopper’ is *ngaga* (Harvey 2019).

## 5. Onomatopoeia

In this section, we discuss some sets that we feel are better analyzed as being due to onomatopoeia.

**#6 Hadza *tʰítʰi*, *tīti* ‘bird’** (more properly cited as /ti:ti-/ ‘bird’ (generic term for birds excluding ostriches)) ~ AA *\*diʔ(diʔ)* ‘k. of bird’

**#30 Hadza *pururu* ‘fly’ (v.)** (more properly cited as /pururu-/ ‘to fly off’) ~ AA *\*pir-* ‘to fly’

Militarev compares the Hadza form *garaani* ‘heron’ with AA roots *\*gawir-* and *\*garay-* ‘heron, crane, ostrich’. We note that the Kanuri (Saharan) word *garagára* ‘heron’ (Cyffer 1994: 88) nicely matches the first CVC of the Hadza, presumably completely due to chance or to onomatopoeia. The Afroasiatic roots also only match an initial voiced velar, low vowel and liquid.

<sup>5</sup> The Hadza root resembles Datooga *géewèdà* ‘shoe’ (Roland Kießling, p.c. 23 June 2023). Though the semantics do not match, they are one of the most abundant antelopes in the area and the leather is quite likely used in making shoes.



The Hadza is specifically *Ardea melanocephala* 'black-headed heron' /galaʔani-ko/ (Peterson 2012: 2018), the term used when referring to the bird's habit of swallowing snakes (or, the name *lomolo* may be used when referring to its long neck, Peterson, *op. cit.*). Hadza does not contrast r/l but it does contrast VV and VʔV. None of the proposed cognates explain the lack of a consonant corresponding to the Hadza glottal stop in this word. If the Hadza root is bimorphemic, it would be possible for the glottal to be part of a second root (\*gala-\*ʔani-) but we have no internal evidence for this. The proposed correspondence set is unsatisfactory both from the standpoint of the phonological correspondences (which ignore half of the Hadza word and the correct spelling of the Hadza) and from the semantic correspondence (which equates herons and ostriches). If any two types of birds might be compared, then it is unsurprising to find at least one root with a voiced velar and a liquid in one of over a hundred possible languages. The fact that two different AA roots have been proposed is unsatisfactory as well, as it introduces an indeterminacy in the historical scenario being proposed.

## 6. Evaluation of Selected Sound Correspondences

In this section, we discuss correspondences involving a subset of Hadza consonants.

### 6.1. Glottal stop

The presence of a glottal stop is not always noted in the orthographic representations cited by Militarev though it is shown in the phonetic transcriptions (when provided). Glottal stops are not merely phonetic in Hadza, but phonemic.

Glottal stops in Hadza roots have been compared to AA \*ʔ in several sets:

#41 'I' *ono* [ʔono, ʔono-ko] (more properly cited as /ʔono-/) ~ AA \*ʔani/u 'I'

#31 'foot' -2 *asenako* [ʔasenako] (more properly cited as /ʔasena-ko/) ~ AA \*ʔa-sin- 'foot, leg'

#9 'blood' *átʰaʔmá-*, *átàmaʰ* [ʔatʰama] (more properly cited as /ʔátʰamá/) ~ AA \*(ʔa-)dam- 'blood'

#64 'person' *unu* [ʔunu] (more properly cited as /ʔunu/) ~ AA \*ʔa/inay-(n)- 'man, person (also elder kin)'

#76 'sleep' *ʔase* (also 'lie') (more properly cited as /ʔase/) ~ AA \*sayʔ- / \*ʔays- 'sit, sleep, rest'

#71 'say' -1 *ʔ* (<\*ʔiyʔ) (more properly cited as /ʔi:/) ~ AA \*ya- / \*ʔiy- 'say'

#79 'stand' *ikha-* (also 'to stop') [ʔíkʰà] (more properly cited as /ʔíkʰà/) ~ AA \*kaʔ/w / \*ʔVʔ(k)- 'rise, be high, stand (up)'

#82 *isho-ko* (*isho* 'sunlight') [ʔiʃoko] (more properly cited as /ʔiʃo-ko/) ~ AA \*ʔa/is- 'sun'

At first glance, these sets might appear to be a sufficient collection to demonstrate a regular, repeated sound correspondence. But, it is necessary for all of the segments in roots to regularly correspond, and not just a single segment. First, however, note that #9 'blood' and #31 'foot' cannot be considered independent of each other as they appear to reconstruct to a single morpheme \*ʔa- in Afroasiatic. These correspondences are dubious, however, since there is no evidence that these syllables are separate morphemes in Hadza. The corresponding segments in #41 'I' are not demonstrated to be regular and there are difficulties with comparing the generally restricted set of consonants that occur in pronominal systems with one another; the small number of consonants that tend to be used greatly increases the likelihood of a chance resemblance (Gordon 1995). Although we do not examine vowel correspondences in this paper, those shown in these sets do not appear to be regular. We do analyze voiceless stops and

fricatives and cannot consider the correspondences shown here involving these to be regular. If only one segment in a word has a regular correspondence, the entire word cannot be presumed to be cognate.

We question the first impression that the /ʔ/~ \*ʔ correspondence is regular.

There is no glottal stop comparable to AA \*ʔ in:

#33 ‘give’ *kwe-* ~ AA \*kaʔ- / \*ʔaVʔk-

No conditioning environment for this is discussed, nor is an explanation given to account for the metathesis.

There also does not appear to be a segment in AA corresponding to the glottal stop in:

#24 ‘egg’ *usle-ko* [ʔúʔe-ko] (more properly cited as /ʔúʔe-ko/) ~ Chadic \*ʂay(ʂay)

#31 ‘foot’ -1 (also ‘leg’) *aluphukwa* [ʔa/ʔup<sup>h</sup>uk<sup>w</sup>a] ~ Cush. S. \*fank<sup>w</sup>-; Kera *káma-a* ‘foot, leg’ (met.), etc.

There are a few cases where a Hadza glottal appears to correspond to a pharyngeal in AA \*ʕ:

#56 *awani-ka* ‘mouth’ (more properly cited as /ʔawanika-ko/ ‘mouth’) ~ AA \*ʕawin- ‘tongue; (part of) mouth’

#52 ‘many’ *ʔaso* (more properly cited as /ʔáso-/ ) ~ AA \*w/yasaʕ

#1 ‘all’ *waʔi* ~ AA \*waʕ ‘one’

#45 ‘know’ *tetha’o* [tet<sup>h</sup>aʔo] (also ‘understand’) (more properly cited as /tet<sup>h</sup>aʔo/) ~ AA \*(y)daʕ- ‘know’

63 ‘one’ *itchâme* [ʔit<sup>h</sup>a:me] (more properly cited as /ʔit<sup>h</sup>a:me/) ~ AA \*ʕist(-an)- ‘one’

Since Hadza does not have a pharyngeal fricative, it might be presumed that such a sound may have merged with the glottal stop. Also, the correspondence between Hadza ʔ ~ AA \*ʕ in #52 is only allowable if one presumes metathesis, yet no explanation is given to motivate the metathesis. Furthermore, this analysis would not explain why other AA roots with \*ʕ have no corresponding glottal stop in Hadza, e.g.:

#23 ‘eat’ *seme, simi* ~ AA \*suʕVm-

#32 ‘to be many, to be plenty, to be full’ *furu-ne* (more properly cited as /furune/ ‘to be many, to be plenty, to be a lot’) ~ \*ʕVpVr- ‘full’

#80 ‘star’ *ntsa-ko* ~ AA \*(t/wV-n)ciʕ(ciʕ-)

To summarize, we see irregular rather than regular, repeated correspondence sets involving the Hadza glottal stop. We do not see sufficient evidence to hint at a linguistic genetic relationship between Hadza and Afroasiatic.

## 6.2. Voiceless Stops

The distinction between aspirated and unaspirated stops (and other obstruents) in Hadza is contrastive and regular sound correspondences should be established involving each type of consonant.

Hadza /t<sup>h</sup>/ is compared with AA \*d:

#9 ‘blood’ *át<sup>h</sup>aʔmá-*, *átámá* [ʔat<sup>h</sup>ama] (more properly cited as /ʔát<sup>h</sup>amá/) ~ AA \*(ʔa-)dam- ‘blood’

#45 ‘know’ *tetha’o* [tet<sup>h</sup>aʔo] (also ‘understand’) (more properly cited as /tet<sup>h</sup>aʔo/) ~ AA \*(y)daʕ- ‘know’

But #49 shows /t<sup>h</sup>/ is compared with AA \*t, with no explanation proposed for this seeming irregularity:

#49 'long' *thas* [t<sup>h</sup>as-] (more properly cited as /t<sup>h</sup>as-/) ~ AA \*tays- 'long'

Hadza /t/ is also compared with AA \*d:

#6 'bird' *t<sup>h</sup>it<sup>h</sup>i, t<sup>h</sup>iti* (more properly cited as /ti:ti-/ 'bird' (generic term for birds excluding ostriches)) ~ AA \*di?(di?) 'k. of bird'

Hadza /t<sup>h</sup>/ and /t/ are compared with AA \*d yet there is no comparable collapse (or alternation) of voiceless (aspirated and unaspirated) labial and velar phonemes with their voiced counterparts as shown by forms such as:

————— p

#30 'fly' (v.) *pururu* (more properly cited as /pururu-/ 'to fly off') ~ AA \*pir- 'to fly'

#91 'two' *pi<sup>h</sup>e, pie-(be)* ~ AA \*(hV)pV(p)- / \*H/yV(m)p/b- (< \*hVp-?) 'two'

————— p<sup>h</sup>

————— k

#33 'give' *kwe-* ~ AA \*ka?- / \*ʔaVk-

#92 'walk' *haka* (*haka* 'go', *ʔetlhikwa* 'walk') (more properly cited as /haka/ 'to go', /ʔic<sup>h</sup>ikwa/ 'to walk, to go, to leave') ~ AA \*k<sup>w</sup>VH- / \*HVk<sup>w</sup>- 'go, walk'

#67 'road' *yeke* (more properly cited as /jeke/ 'path') ~ AA \*kaw/y(k)- / \*yVhk- / \*hVwk- 'go, walk, come; road'

————— k<sup>h</sup>

#62 'not' *'ukuwa* (more properly cited as /ʔuk<sup>h</sup>u-/ 'to not exist') ~ AA \*(ʔa-)k<sup>w</sup>ay- 'not'

#79 'stand' *ikha-* (also 'to stop') [ʔik<sup>h</sup>à] (more properly cited as /ʔik<sup>h</sup>à/) ~ AA \*kaʔ/w / \*ʔV<sub>k</sub>(k)- 'rise, be high, stand (up)'

We find it implausible that the stop series would not pattern together in a similar way. A sufficient number of regular, repeated sound correspondences have not been presented for the sets involving pulmonic stops to be convincing.

Since it is difficult to find repeated correspondences with any particular consonant, we might instead compare consonants grouped into natural classes in order to detect correspondence patterns that might otherwise fall below the level of significance. However, in grouping together the voiceless plosives, we see that voiceless unaspirated and unaspirated velars and labials pattern differently with respect to coronals; the coronals are posited to be cognate with voiced stops in AA while the other stops are posited to correspond to voiceless plosives. Not only is there no explanation for the contrast between aspirated and unaspirated voiceless stops in Hadza, there is no explanation posited by Militarev to explain why the natural class of voiceless stops does not pattern together. Grouping consonants together into natural classes is one way to test the phonological naturalness of the sound correspondences proposed.

### 6.3. Fricatives

The only fricative with more than a few proposed cognate sets is /s/. The other sets involving fricative cannot be said to show regular, repeated sound correspondences. (The sets involving /ʃ/ at least have one repetition without a contradictory sound correspondence, but only one of these ('to love') reconstructs to AA and it is almost certainly a loan from West Rift (South Cushitic)).

\_\_\_\_\_ f

#32 ‘to be many, to be plenty, to be full’ *furu-ne* (more properly cited as /furune/ ‘to be many, to be plenty, to be a lot’) ~ \*ʔVpVr- ‘full’

\_\_\_\_\_ †

#24 ‘egg’ *usle-ko* [ʔúʔeko] (more properly cited as /ʔúʔe-ko/) ~ Chadic \*ʂay(ʂay)

Hadza *slaa* [ʂaʔa] (more properly cited as /ʂaʔa/) ‘to love’ ~ AA \*ʂaʔ/w- ‘want, wish, love, like’

\_\_\_\_\_ ʃ

#82 *isho-ko* (isho ‘sunlight’) [ʔiʃoko] (more properly cited as /ʔiʃo-ko/) ~ AA \*ʔa/is- ‘sun’

Hadza *dushu* ‘distended, big stomach’ (more properly cited as /dúʃu/<sup>6</sup> ‘distended stomach (from malnutrition)’ also /dúʃu-ko/ ‘mole cricket’) ~ AA \*dVs(-Vm/n)- ‘fat belly’

Hadza *shububu-bi* ‘lungs’ (more properly cited as /ʃububú-/ ~ AA \*ci/anp- ‘lungs’

‘rhino’ *losho* more properly cited as /lo:ʃo/ ‘sighting/hunting name for the rhinoceros’ (when it is stooped and ready to charge) ~ AA \*warʂ- ‘rhinoceros’

\_\_\_\_\_ z ~ dz<sup>7</sup>

Hadza *uzame-ko* [ʔudzameko] ‘spotted (laughing) hyena’ (more properly cited as /ʔudzame-ko/ ‘spotted hyena’) ~ Chadic \*zVm- ‘lion’

Hadza /s/ corresponding to AA \*s is one of the few consonantal correspondences with more than 2 examples. However, in many of these cases, there are elements in the AA forms which have no correspondence whatsoever in the Hadza forms. For instance, there is no explanation for the missing second syllable in Hadza *sa* ‘to rain’ compared to AA \*sawiʔ- ‘rain’, or of the difference in order of segments in Hadza *ʔase* ‘sleep’ compared to AA \*sayʔ-/ʔays- ‘sit, sleep, rest’.

#23 Hadza ‘eat’ *seme-* compared to AA \*suʃVm- ‘eat’

#31.2 Hadza [ʔasena-ko] ‘toes’ compared to AA \*ʔa-sin- ‘foot, leg’

#49 Hadza *thas-* [t<sup>h</sup>as-] ‘long’ compared to AA \*tays ‘long’

#52 Hadza *ʔaso* ‘many’ compared to AA \*w/yasaʔ- ‘big, many’

#65 Hadza *sa* ‘to rain’ compared to AA \*sawiʔ- ‘rain’

#76 Hadza *ʔase* ‘sleep’ compared to AA \*sayʔ-/ʔays- ‘sit, sleep, rest’

#93 Hadza ‘to warm *sifi-* (oneself by a fire)’ compared to Semitic \*šVhan- ‘(be) warm, hot; warm oneself’

Hadza *biso-ko* ‘wildebeest’ compared to AA \*bus- ‘goat, bushbuck’

Hadza *ʔakwisiti-ko* ‘sinew that runs along the spine and neck’ compared to AA \*kac/sw- ‘back with shoulders’

Militarev’s correspondence sets are curated to some extent. For instance, #17 Hadza *misi* ‘die’ is not scored even though it is similar to S. Bauchi \*mis- ‘die’ and AA \*mwɪ, because the S. Bauchi form cannot be shown to be a regular reflex of the AA form, and the s~t correspondence between Hadza and AA would also be irregular.

Despite Militarev’s attempts to show regular, repeated sound correspondences, we would need to see an explanation of the vowel and approximant correspondences in these same words to make sense of these as related forms.

<sup>6</sup> This word is a rare instance of a clickless word that appears to be synchronically related to another word in Hadza with a click, i.e. /!<sup>h</sup>uʃu:/ ‘have a big/stretched stomach’ (Miller 2013). Also suspiciously similar is: /!<sup>h</sup>uʃu-ko/ ‘navel and/or umbilical cord’, ‘fat critter with visible innards’ (such as some toads or crickets) (Miller 2013).

<sup>7</sup> Note that although we prefer to consider that this root has a /dz/ phoneme, there is no contrast between /dz/ and /z/ in the language so it is possible to analyze this as a fricative.

## 6.4. Central Ejective Affricates

Hadza distinguishes phonemes /tʃ/ and /tsʰ/ and there is no synchronic evidence or evidence from internal reconstruction to connect the two phonemes. Militarev compares Hadza /tʃ/ to AA \*ç and \*č, and Hadza /tsʰ/ is compared to AA \*ç, \*č and \*ĉ. There is no explanation proffered for the irregular correspondences. Looking at the sets below, we see other issues with the comparanda. Evidence proposed based on central ejective affricates does not support the hypothesis that Hadza is an Afroasiatic language.

In the cognate set #21 ‘ear’ /tʃ/ is connected to the Hadza phoneme /tsʰ/ in set #46 ‘leaf’ (compared to AA \*h/ħaç- ‘leaf; ear’). The Afroasiatic Database (AADB) provides two Proto-Omotoc reconstructions for ‘leaf’: \*Hyaç- (supported by forms Chara: *yēčā*; East Mao (Diddesa): *yaç-*) and \*Hwayš- (supported by: Koyra (Badditu, Amuru): *wāšē*; Ganjule: *waše*; Janjero (Yamma, Yemsa): *t-oša*; Gimirra (Benesho, She): She *aisi*). Neither of these reconstructions matches the Omoto reconstruction cited by Militarev: \*Hayç-. Although there are a number of similar-looking roots in Omoto languages, these do not contain ejectives: *haytta* (Wolayta), *hayḏa* (Gofa, Dawro/Kullo), *hayḏ* (C’ancha) ‘leaf’ (Alemayehu Abebe 2002: 8). Note that in many of these languages, the same word may also refer to ‘ear’: *haytta* (Wolayta), *hayḏa* (Gofa, Gamo, Dawro/Kullo), *hayḏ* (C’ancha) (*op. cit.*: 7). A correspondence in form and meaning between ‘(h)ear’/‘leaf’ cannot be presumed to be evidence of common inheritance. Not only can the connection between ‘leaf’ and ‘ear’ occur in languages unrelated to Afroasiatic, as shown in Figure 1, but colexifications may spread within a linguistic area, and be more indicative of geography than inheritance (Urban 2012, Campbell et al. 1986).

| Language     | Family       | Form    |
|--------------|--------------|---------|
| Dera         | Afro-Asiatic | kumo    |
| Rasawa       | Lakes Plain  | ura     |
| Diyari       | Pama-Nyungan | tharlpa |
| Ngamini      | Pama-Nyungan | talpa   |
| Wulguru      | Pama-Nyungan | bina    |
| Yandruwandha | Pama-Nyungan | thalpa  |

Figure 1. Colexifications for ‘leaf’ and ‘ear’ in CLICS<sup>3</sup> Database of Cross-Linguistics Colexifications (Rzymiski et al. 2019, accessed 27 June, 2023)

The Hadza consonant /tʃ/ is proposed by Militarev to have a repeated correspondence with AA \*č in two words. In the word *tʃowa* ‘gecko’, which is compared to the AA root \*ĉiçay-, no explanation is given to explain why the Hadza word does not have a syllable preceding the /tʃ/ to correspond to AA \*č. There is a similar problem with the other Hadza word containing this affricate: *watʃo* ‘slender mongoose’ (Peterson 2012: 216). The Hadza word is said by Militarev to correspond to AA \*č in words with varied meanings (‘rat’, ‘mouse’, ‘weasel’, ‘mongoose’, ‘jerboa’). Hadza speakers distinguish between five species of mongoose (Peterson 2012: 216) and there is no sense in which these carnivores would be referred to as if they were mice or rats. Furthermore, there is an initial syllable in most of the Afroasiatic forms (\*ʔV(n)çaw-) that has no correspondence with Hadza.

In set #78 ‘smoke’, the AA \*č consonant is connected to Hadza /tsʰ/ rather than /tʃ/. No explanation for the different correspondence is noted (compared to ‘gecko’ and ‘slender mongoose’), however, the AA form may alternatively be reconstructed with the \*ç consonant. The

Hadza consonant /ts'/ in the word *ts'eʔa-* 'to shit' is also proposed to correspond to AA \*č̣, but the same consonant in 'to smile' is proposed to correspond to \*č̣ (*ts'ukuts'uku* 'to smile' ~ AA \*č̣Vh/Vk 'to laugh'). No explanation is given for the irregularity of the proposed correspondences. We are not convinced by evidence consisting of a single repeated sound correspondence, especially when a set that violates the proposed regularity has also been proposed.

### 6.5. Lateral Ejective Affricates

Sets involving lateral ejective affricates do not show regular, repeated sound correspondences though the comparanda here are quite striking and are deserving of some discussion.

Sets #10 'bone' and #97 'white' appear to show a single repetition of a sound correspondence between the Proto-Afroasiatic sound \*č̣ and Hadza /c̣ʔ/. For a sound correspondence to count as evidence of a linguistic genetic relationship, however, a sound correspondence must be both regular and repeated – linguists have not quantified the number of times a correspondence must occur to count as evidence, but certainly, the difference between one occurrence and two occurrences is not statistically significant. Furthermore, set #36 'hair' shows Hadza /c̣ʔ/ does not repeat this same correspondence with AA. In this set, /c̣ʔ/ corresponds to \*č̣/š and no explanation is proffered to account for the different correspondence.

**#10 'bone' *midla*** (Miller et al. 2021), *mitl'a* (Sands 2012: 5) (more properly cited as /mic̣ʔa-/) =AA \**malič̣-* 'bone': =Cush. C. \**ḥač-* (<\**mač-*); =E.: HEC: Darasa, Burji *mičč-o*; Yaaku *moč-o*; =S.: Dahalo *mičč-o*<sup>1</sup> // =Omot. N.: Mao (Sezo) *mālt-é* ∅ AADB 1269.

**#97 'white' *pedla*** (*petl'a-* in Sands 2012) (more properly cited as /pec̣ʔa-/) =Sem. \**payš-* (AA \**payč̣-*) 'white'.

**#36 'hair' *hadle*** (more properly cited as /*hac̣ʔe-*/) ~ AA \*(*Ha-*)č̣/šVw- 'hair, feather': Cush. S. =Ma'a *ašu* 'hair' (\*č̣ and š both render š in Ma'a according to Takács 2011) // Eg. (Pyr.) *šw.t* 'feather' ∅ AADB 1284.

Using Ma'a to establish a sound correspondence is problematic as we explained above. Both Egyptian (and Ma'a) have a rounded segment (*w* and *u*, respectively) that has no regular correspondence with the Hadza front vowel. No other comparison sets with Hadza /c̣ʔ/ are proposed by Militarev. Two sets of corresponding sounds is not a high enough number to be considered by us to be a 'regular, repeating sound correspondence'; rather, this is more typical of a chance resemblance or resemblance due to borrowing.

Even though the sound correspondence involving the sets 'bone' and 'white' is striking, we must still ask whether or not one of these forms might be due simply to chance resemblance. Unrelated Kanuri (Saharan) has a form *bûl* 'white' (adj.; Cyffer 1994: 212) which strongly resembles the set, particularly when one considers that Kanuri does not have lateral obstruents.

There is another Hadza root *petla* /pec̣ʔa-/ 'to shine, glitter, gleam' with a pulmonic (not ejected) lateral affricate that is conceivably connected to /pec̣ʔa-/ 'white', though not through any known synchronic alternation.

Even within Afroasiatic, it is unclear which forms the Hadza form(s) should be compared to. Militarev connects Hadza to a Proto-Semitic form \**payš-* rather than to the Proto-Omototic form. A Proto-Omototic form may have yielded reflexes without lateral ejective obstruents, e.g. Yemsa *bitfà* 'yellow' (Aklilu Yilma et al. 2002: 26), Melo (Omototic) *bó:ts* 'white' (Siebert & Caudwell 2002); *botta* (Wolayta), *bođa* (Gofa, Dawro/Kullo), *buđ* (C'anacha, Dorze) (Alemayeh Abebe 2002: 12). Takács (2011: 185–186) connects Proto-Omototic \**bōč* to Semitic \**byđ* (Arabic *byđ* : *bāda* 'to grow yellow', *bayyada* 'to make white', etc.) and to Chadic forms such as Mafa-Mada

\**baḍ* 'white', all derived from AA \**b-ĉ* 'white' (which is comparable to the AA form \**ḥayĉ* cited by Militarev). If the rounded vowels in Omotic might be due to coloring from the initial labial, it is not out of the question, then, that Hadza might have borrowed a form with an ejective from the same type of source as 'bone'.

Another issue with the sound correspondence proposed for 'bone' and 'white' is that AA \**ĉ* has also been proposed by Militarev to have other reflexes in Hadza: *ts'ukuts'uku* 'to smile' (Miller et al. 2021: 560) is compared with AA \**ĉVḥ/Vḥ* 'to laugh'. Even disregarding the semantic differences between these two words, there is no explanation of the different sound correspondence between this set and 'bone'/'white' (not to mention different correspondences involving Hadza /*ts'*/). If a word has a sound correspondence that is irregular and cannot be explained as a conditioned sound change, then it cannot count as evidence for a linguistic genetic relationship.

## 7. Role of chance not fully evaluated

In this section, we discuss some comparison sets that raise the question of whether the similarities are due to chance rather than common inheritance from Afroasiatic.

### #31.2 *asena-ko* [ʔasenako] 'toes'

The Hadza root /ʔasena/ is compared to AA \**ʔa-sin-* 'foot, leg'. Some Chadic forms have the initial syllable, but not the Omotic, Cushitic or Egyptian forms. The paper provides no explanation of the presence or absence of purported *ʔa-* prefix nor of its function in Hadza and Afroasiatic. The second and third consonants of the Hadza root are also reasonably similar to those in /*s̄n:ntʰ*/ 'foot' in Northern Koma (Koman, Nilo-Saharan) (Bender 1971). We must stress that even if Hadza *were* Afroasiatic, it is still possible for chance resemblances to occur, particularly when involving relatively high-frequency consonants such as /*s*/ and /*n*/.

#32 *furu-ne* 'to be many, to be plenty, to be full' (more properly cited as /*furune*/ 'to be many, to be plenty, to be a lot')

The first part of the Hadza root is phonologically similar to English 'full', which reminds us that resemblances due purely to chance may certainly occur.

**Hadza** *uzame-ko* [ʔudzameko] 'spotted (laughing) hyena' (more properly cited as /ʔudzame-ko/ 'spotted hyena')

The Hadza root is as similar both in terms of phonology and semantics to Kanuri (Saharan) *zazárma* 'leopard' (Cyffer 1994: 106) as it is to any of the Chadic forms proposed (\**zVm-* 'lion', etc.). None of the forms attempt to show a correspondence to the initial *ʔu-* syllable in the Hadza form.

It is not difficult to find two words in even a very short wordlist that have a single repeated sound correspondence that are roughly comparable to the correspondence sets that Militarev proposes (i.e. where not all segments or even syllables have corresponding segments or syllables). For instance, in comparing Hadza with the Nigerian language Akye<sup>8</sup> (Benue-Congo, Plateau) (Decker et al. 2021), similar forms involving *k ~ k* include:

<sup>8</sup> This language was randomly chosen for comparative purposes.

|  |                          |
|--|--------------------------|
| Hadza  | Akye <sup>9</sup>        |
| <i>yeke-</i> / <i>jeke-</i> / ‘road, path’                       | <i>ókée</i> ‘road, path’ |
| /ʔaku/ ‘what kind, which’  | <i>āké</i> ‘what?’       |
| /k <sup>w</sup> a   <sup>h</sup> a/ ‘to vomit’                   | <i>kwè</i> ‘vomit’       |
| /kuku, k <sup>w</sup> ak <sup>w</sup> a/ ‘to dig with the hands’ | <i>kōη</i> ‘dig’         |

Even more pairs could be listed if the semantic correspondences were stretched. Additional similar forms could be added if we violate the *k~k* pseudo-correspondence. We are definitely not trying to claim that Hadza and Akye might be related; rather, that lexical similarities occur due to chance quite often. This is especially true when the segment inventories and phonotactics of the languages are similar.

## 8. Discussion

Militarev’s proposal of Hadza as an Afroasiatic language (with or without a particularly close connection to Chadic, Cushitic or Omotic) concerns itself with linguistic data but is not an ecologically or historically-situated proposal. By this, we mean that there are no proposed times or locations of a Hadza-Afroasiatic homeland. Although Militarev cites Ongota data to support Afroasiatic etymologies, he makes no reference to Fleming (2006) and the reconstruction of prehistory proposed therein.

Militarev’s proposal raises many more questions than it answers. If Hadza is Afroasiatic, then does that mean that Proto-Afroasiatic should not be reconstructed as a language of agriculturalists/pastoralists? If Hadza is more closely connected to Chadic than to other branches of Afroasiatic, then did the Hadza people migrate southwards (e.g. in a scenario similar to that proposed by Blench 2013b)? There is no explanation as to why foragers would migrate (or why non-foragers would have resided further south at a time when there is no archaeological trace of their habitation). It is unclear whether Militarev considers Hadza a separate branch of Afroasiatic (similar to Fleming’s (2006) analysis of Onogota), a branch of Chadic, or other. Regardless of which of these scenarios obtains, later contacts with Cushitic should be expected and not treated as independent sources of evidence for Afroasiatic affiliation.

Early loans between Afroasiatic languages and Hadza (pre-dating contact with Proto West Rift / Southern Cushitic) undoubtedly occur. We know that clicks occur in Dahalo, a Cushitic language formerly spoken by hunter-gatherers, even though these consonants have never been reconstructed for Cushitic or Afroasiatic, so the direction of borrowing need not be assumed to be from Afroasiatic into Hadza but the reverse may also have occurred. Hunter-gatherer groups existed throughout Kenya, Tanzania and SW Ethiopia and as people shifted from a pre-existing language of foragers to an Afroasiatic language, they may have retained vocabulary items, or acquired vocabulary items through later, continued contact with foragers. Since obsidian found near Lake Eyasi (where Hadza is spoken) can be traced to Central Kenya (Goldstein 2022), it is not outrageous to suggest that the area in which a Hadza-like language was spoken may once have been much larger, or that Hadza people once traveled more widely, especially following the wildebeest migrations northward. It is also quite likely that some words may be shared by groups that are not directly connected but which were connected through an intermediary language. Certainly, it should be unsurprising to find Wanderwörter shared by languages that today appear to have no contact at all.

<sup>9</sup> Forms cited are those of Kiguna village.



Our alternate analysis is that 1) there are chance resemblances between Hadza and Afroasiatic; 2) there are some loans from Afroasiatic into Hadza, 3) there are later loans from Cushitic into Hadza; 4) there are Hadza (and Sandawe) loans into Cushitic. That borrowing from preexisting groups into newcomer languages occurred is evident in more recent examples of borrowings from Hadza or Sandawe into Cushitic (see Ehret 2013b and Kießling & Mous 2012). Early contact between Hadza and Afroasiatic is not sufficiently explored; even if we entertained the notion that Hadza were Afroasiatic, there would still be a need to distinguish contact from inherited items.

It is hard to critique a paper that does not propose a concrete scenario. For example, #24 'egg' is compared to two different reconstructed Chadic roots, \**ṣay(ṣay)*- and \**ʔi(n)ṣ-*, and the path between these roots and the Hadza is not made explicit. #31 'foot' is compared to forms in Chadic, with metathesis, and to South Cushitic \**fanq<sup>w</sup>*- (therein \**fanḳw-*) – both forms ignoring the initial syllable *ʔaphukwa* or *ʔuphukwa*. There are three roots reconstructed with the meaning 'all' in the Afroasiatic database (AADB) but none match the forms proposed to connect to the Hadza form *wa'i* with that meaning.

In his discussion, Militarev suggests that Hadza may be parallel to Cushitic and Omotic as a sub-branch of South Afrasian but continues to stress the striking similarities with Chadic. The link with Chadic is probably more surprising and indeed tantalizing. It would be interesting to attempt to develop that idea by comparing Hadza to Proto-Chadic. We have the impression that Proto-Chadic is more often than Proto-Cushitic suggested in the proposed Hadza-Afrasian cognate sets. Working with reconstructed languages rather than individual languages in these proposals for cognate sets has the methodological advantage of reducing chance resemblances.

## 9. Conclusion

In this paper we have critiqued many of the lexical comparison sets proposed by Militarev to provide support for a linguistic genetic relationship between Hadza and Afroasiatic. We argue that some of these sets are better explained as being due to contact with other languages and others may be discounted because they do not accord with a Hadza-internal analysis. Of the sets which remain, none show more than two repeated, regular sound correspondences; this falls under the level of proof needed to secure a label of relatedness. Given that Hadza has many consonants, the number of lexical items with any one consonant or consonant type is relatively low compared to languages with fewer consonants; this means that it is relatively difficult for the language to have retained a large number of words with each consonant over a great time depth. Just as it can be difficult to recognize old, phonologically-assimilated loans from Latin, etc. in Basque and Berber (cf. Trask 1996, Kossmann 2013), so too, is it difficult to recognize early loans in Hadza – particularly since the donor languages are not anywhere near as well-attested as Latin. The challenge to *prove* that Hadza might be related to another language is an attractive but daunting one. We hope that we have outlined some of the types of errors that can befall the intrepid linguist willing to take a chance on exploring such little-known ground.

## References

- Aklilu Yilma, Ralph Siebert, Kati Siebert. 2002. Sociolinguistic Survey of the Omotic Languages Sheko and Yem. *SIL Electronic Survey Reports (SILESR)*, 2002-053.
- Alemayehu Abebe. 2002. Omoto Dialect Pilot Survey Report. *SIL Electronic Survey Reports (SILESR)*, 2002-068.

- Bastin, Yvonne, André Coupez, Evariste Mumba, Thilo C. Schadeberg (eds.). 2002. *Bantu Lexical Reconstructions 3*. Tervuren: Royal Museum for Central Africa.
- Blench, Roger. 2013a. Linguistic aspects of Hadza interactions with animals. In: Alena Witzlack-Makarevich, Martina Ernszt (eds.). *Khoisan Languages and Linguistics (Proceedings of the 3rd International Symposium July 6-10, 2008, Riezlern/Kleinwalsertal)*. *Quellen zur Khoisan-Forschung / Research in Khoisan Studies*, 29: 101–110. Köln: Rüdiger Köppe.
- Blench, Roger. 2013b. Links between Cushitic, Omotic, Chadic and the position of Kujarge. In: Marie-Claude Simeone-Senelle, Martine Vanhove (eds.). *Proceedings of the 5th International Conference on Cushitic and Omotic Languages, Paris, 16-18 April 2008 (Cushitic and Omotic Studies, 3)*: 67–80. Köln: Rüdiger Köppe.
- Campbell, Lyle, Terrence Kaufman, Thomas C. Smith-Stark. 1986. Mesoamerica as a linguistic area. *Language* 62(3): 530–570.
- Chebanne, Andy. 2017. Common Khoisan: A reflection on their linguistic and cultural relationships. In: Kwesi Kwaa Prah, Lazarus Miti (eds.). *Deconstructing the African Tower of Babel: Between the Harmonisation and Fragmentation of African Language Orthographies (CASAS Book Series, No. 120)*: 95–116. Cape Town: Centre for Advanced Studies of African Society.
- Coburn, Jeremy, Bonny Sands, Andrew Harvey, Richard Griscom (forthc.). Tonal Patterns of Hadza nouns. In: *The Proceedings of the 10th World Congress on African Linguistics, Leiden, 7-11 June 2021 (Contemporary African Linguistics)*. Berlin: Language Science Press.
- Cyffer, Norbert. 1994. *English-Kanuri dictionary (Westafrikanische Studien: Frankfurter Beiträge zur Sprach- und Kulturgeschichte, Bd 3)*. Köln: Rüdiger Köppe.
- Decker, Ken, Innocent Jonah, John Muniru, Yakubu Danladi, Benard Abraham, David Onoja, Christina Riepe. 2021. A Sociolinguistic Profile of the Akye (Ake) [aik] Language of Nasarawa State, Nigeria. *Journal of Language Survey Reports* 2021(042): 1–51.
- Ehret, Christopher. 2013a. Sub-Saharan Africa: Linguistics. In: Immanuel Ness, Peter Bellwood (eds.). *The Encyclopedia of Global Human Migration. Vol. I: Prehistory*: 96–106. Malden, MA / Oxford: Wiley-Blackwell.
- Ehret, Christopher. 2013b. The Extinct Khoesan languages in East Africa. In: Rainer Vossen (ed.). *The Khoesan Languages*: 465–479. New York: Routledge.
- Elderkin, Edward Derek. 1978. Loans in Hadza: Internal evidence from consonants. *Occasional Paper 3*, Dept. of Foreign Languages, Dar es Salaam. University of Dar es Salaam.
- Elderkin, Edward Derek. 1982. On the classification of Hadza. *Sprache und Geschichte in Afrika* 4: 67–82.
- Elderkin, Edward Derek. 2014. Clicks, prosodies and Khoisan. In: Tom Güldemann & Anne-Maria Fehn (eds.). *Beyond 'Khoisan': Historical Relations in the Kalahari Basin*: 103–122. Amsterdam: John Benjamins.
- Fleming, Harold C. 2006. *Ongota: A decisive language in African prehistory*. (*Aethiopistische Forschungen*, 64). Wiesbaden: Harrassowitz.
- Goodman, Morris. 1971. The strange case of Mbugu. In: Dell Hymes (ed.). *Pidginization and Creolization of Languages*: 243–254. Cambridge University Press.
- Gordon, Matthew J. 1995. The phonological composition of personal pronouns: implications for genetic hypotheses. *Berkeley Linguistics Society* 21: 117–128.
- Griscom, Richard. 2019. *Topics in Asimjeeg Datooga Verbal Morphosyntax*. Ph.D. thesis, University of Oregon.
- Griscom, Richard, Andrew Harvey. 2021. *Hadza: an archive of language and cultural material from the Hadzabe people of Eyasi (Arusha, Manyara, Singida, and Simiyu regions, Tanzania)*. *Endangered Languages Archive*. Available online at: [hdl.handle.net/2196/82e2b99d-5c62-4210-8903-8dd976337c10](https://hdl.handle.net/2196/82e2b99d-5c62-4210-8903-8dd976337c10).
- Güldemann, Tom. 2014. 'Khoisan' linguistic classification today. In: Tom Güldemann & Anne-Maria Fehn (eds.). *Beyond 'Khoisan': Historical Relations in the Kalahari Basin*: 1–41. Amsterdam: John Benjamins.
- Harvey, Andrew 2019. *Ihanzu: an archive of language and cultural material from the Ihanzu people of Mkalama (Singida Region, Tanzania)*. London: SOAS, Endangered Languages Archive. Available online at: [hdl.handle.net/2196/00-0000-0000-0014-1365-F](https://hdl.handle.net/2196/00-0000-0000-0014-1365-F).
- Harvey, Andrew. 2021. *The lack of labiodental in Ihanzu as a result of contact with Hadza*. Talk given at the 10th World Congress of African Linguistics (WOCAL10), Leiden (Online) 10/06/2021. doi: 10.17605/OSF.IO/HNDC5
- Honken, Henry. 2013. Genetic relationship: An overview of the evidence. In: Rainer Vossen (ed.). *The Khoesan Languages*: 13–24. New York: Routledge.
- Hudson, Grover. 1989. *Highland East Cushitic dictionary*. Hamburg: Helmut Buske.
- Kibebe Tsehay Taye. 2015. *Documentation and grammatical description of Chabu*. Addis Ababa: Addis Ababa University doctoral dissertation.
- Kießling, Roland, Maarten Mous. 2003. *The lexical reconstruction of West Rift (Southern Cushitic)*. Köln: Rüdiger Köppe Verlag.

- Kossmann, Maarten. 2013. *The Arabic Influence on Northern Berber*. *Studies in Semitic Languages and Linguistics*. Leiden: Brill.
- Miller, Kirk (ed.), w. Mariamu Anyawire, G. G. Bala, Bonny Sands. 2013. *A Hadza Lexicon*. *Mang'ola, Tanzania*. Ms.
- Militarev, Alexander. 2023. Hadza as Afrasian? *Journal of Language Relationship* 21(2): ???.
- Mous, Maarten, Martha Qorro, Roland Kießling. 2002. *An Iraqw – English Dictionary*. *Cushitic Language Studies*, 15. Köln: Rüdiger Köppe.
- Nurse, Derek, Thomas J. Hinnebusch. 1993. *Swahili and Sabaki, a Linguistic History*. *University of California Publications in Linguistics*, 121. Berkeley: University of California Press.
- Peterson, Daudi (w. Richard Baalow & Jon Cox). 2013. *Hadzabe, By the Light of a Million Fires*. Dar es Salaam: Mkuki na Nyota Press.
- Reinisch, S. Leo. 1902. *Die Somali-Sprache. Bd II: Wörterbuch Somali-Deutsch, Deutsch-Somali*. Vienna: Alfred Hölder.
- Reinisch, S. Leo. 1895. *Wörterbuch der Bedauye-Sprache*. Vienna: Alfred Hölder.
- Rzyski, Christoph, Tiago Tresoldi et al. 2019. *The Database of Cross-Linguistic Colexifications: reproducible analysis of cross-linguistic polysemies*. Available online at: [clics.cld.org](https://clics.cld.org), doi: 10.1038/s41597-019-0341-x
- Sands, Bonny. 1998. *Eastern and Southern African Khoisan: Evaluating Claims of Distant Linguistic Relationships*. *Quellen zur Khoisan-Forschung/Research in Khoisan Studies*, 14. Köln: Rüdiger Köppe.
- Siebert, Ralph, Simon Caudwell. 2002. Sociolinguistic Survey Report of the Melo (Malo) and Mursi Languages of Ethiopia. *SIL Electronic Survey Reports (SILESR)*, 2002-046.
- Starostin, George. 2008. K voprosu o geneticheskoj prinadlezhnosti jazyka hadza [On the issue of the genetic affinity of Hadza]. In: V. F. Vydrin (ed.). *Afrikanskiy sbornik 2007*: 262–278. Saint Petersburg: Nauka.
- Starostin, Georgij S. 2013. *Jazyki Afriki. Opyt postrojenija leksikostatisticheskoy klassifikatsiji. T. I. Metodologija. Kojсанские языки* [Languages of Africa: an attempt at a lexicostatistical classification. Volume 1: Methodology; Khoisan languages]. Moscow: Jazyki slav'anskoj kultury.
- Takács, Gábor. 2011. Omotic lexicon in its Afro-Asiatic Setting I: Omotic \*b- with dentals, sibilants, and velars. In: Luca Busetto, Roberto Sottile, Livia Tonelli, Mauro Tosco (eds.). *"He bitaney lagge": Studies on Language and African Linguistics in Honour of Marcello Lamberti*: 183–200. Milano: Qu.A.S.A.R.
- Trask, Robert Lawrence. 1996. *The History of Basque*. London: Routledge.
- Urban, Matthias. 2012. *Analyzability and Semantic Associations in Referring Expressions: A Study in Comparative Lexicology*. Ph.D. thesis, Leiden University.
- Wedekind, Klaus. 2008. *Gedeo Dictionary, Revision of 1978, Listing by Semantic Categories*. Available online at: [www.kwedekind.de/Eingang1/Biblio1.htm](http://www.kwedekind.de/Eingang1/Biblio1.htm).
- Wedekind, Klaus, Charlotte Wedekind, Abuzeinab Musa. 2007. *A Learner's Grammar of Beja (East Sudan)*. *Grammar, Texts and Vocabulary (Beja-English and English-Beja)*. (Afrikawissenschaftliche Lehrbücher Volume 17). Köln: Rüdiger Köppe.
- Zorc, R. David, Madina M. Osman. 1993. *Somali-English Dictionary with English Index, 3rd edition*. Wheaton MD: Dunwoody Press.

Бонни Сэндс, Маартен Моус, Мауро Тоско, Эндрю Харви. Почему хадза (скорее всего) не афразийский язык: ответ на статью А. Ю. Милитарева «Хадза — афразийский язык?»

В настоящем ответе на статью А. Милитарева обсуждаются проблемы, связанные с использованием лексических данных для обоснования афразийского происхождения языка хадза. Некоторые из них, как, например, некорректный морфологический анализ ряда сложных форм и неопознанные заимствования из соседних языков, легко устранимы при более глубоком изучении внутренней структуры и лингвистических контактов хадза. Другие скорее связаны с выбранной методикой исследования; так, широкий разброс семантики сравниваемых слов увеличивает вероятность случайных сопоставлений, равно как и включение в сравнение слов, носящих звукоподражательный характер. Главной проблемой, не позволяющей признать гипотезу доказательной, тем не менее, следует считать отсутствие регулярных и рекуррентных звуковых соответствий между хадза и афразийскими языками.

Ключевые слова: языковые изоляты; хадза язык; койсанские языки; кушитские языки; афразийские языки; чадские языки; дальнейшее родство языков.

## How we could show that Hadza is Afroasiatic: a response to Militarev’s “Hadza as Afrasian?”<sup>1</sup>

In this brief response to Alexander Militarev’s paper on the Afroasiatic (Afrasian) affiliation of Hadza, I attempt to identify the major theoretical issues with his lexicostatistical analysis of the similarities between Hadza and the various branches of the Afroasiatic family, explaining why this analysis cannot be accepted as conclusive, and suggesting some steps that could be taken in order to weed out insignificant evidence (e.g. more attention toward meticulous step-by-step reconstruction of the proto-wordlists potentially involved in such a comparison).

*Keywords:* language isolates; Hadza language; Khoisan languages; Afroasiatic languages; long-distance relationship; lexicostatistics.

Dr. Alexander Militarev’s paper, in which he presents a (seemingly) impressive amount of linguistic evidence for the Afrasian (Afroasiatic) affiliation of Hadza, one of the most famous and intriguing isolates on the African continent, is of immense interest to myself — not only because I, too, have been seriously involved for more than a decade in figuring out the genetic and areal connections of Hadza on a lexical basis (Starostin 2008, 2013, etc.), but also because the arguments laid out in Militarev’s paper have very wide-reaching historical implications. Essentially, the paper could be interpreted as a specific case study in trying to answer a complicated general question — *is it possible at all, and if yes, how is it possible to convincingly demonstrate the genetic affinity between a modern day linguistic isolate and an entire macro-family (= super-family, phylum, etc.) of languages going back to the very limits (some might argue, even beyond the limits) of the classic comparative method?*

In fact, this particular case study is as perfect as it could ever be. On one side of the equation, we have Hadza, a language that clearly has no living relatives that would not be separated from it by thousands of years (how many thousands — remains to be seen); has been first attested no more than a hundred or so years ago (if we take something like Obst 1912 as the starting point); and, because of the small number of speakers, shows very little, if any, dialectal variety, making internal reconstruction based on dialectal comparison impossible. (Careful and detailed study of the language itself allows for a little bit of internal reconstruction based on morphemic analysis of its lexemes, as is shown in Sands et al. 2023 in this volume, but one should never overestimate the potential of such a reconstruction). In short, we know almost nothing of the linguistic history of Hadza, as compared to, say, an average Indo-European or an average Bantu language, the chronological distance between which and its hypothetical distant relatives (e.g. from other branches of “Nostratic” or “Niger-Congo”) can be easily and significantly shortened through proper historical reconstruction based on comparison with its close relatives.

---

<sup>1</sup> Despite my overall skeptical assessment of Dr. Militarev’s hypothesis, I would nevertheless wish to offer my deepest gratitude to him for his painstaking research on lexical connections between Afroasiatic and its neighboring language, providing plenty of food for thought and material on which to refine and rethink our methodology of evaluating “far-reaching” hypotheses of language relationship.

Additionally, this response could not have been made possible without generous financial support (in 2023) from the Basic Research Program at the National Research University Higher School of Economics.

On the other side of the equation, we have Afroasiatic — one of the few “macro-family” level linguistic taxa that continues to enjoy widespread mainstream acceptance, despite many significant issues, such as the internal constituency of the taxon (e.g., with serious doubts cast upon the inclusion of the Omotic branch, see Theil 2006), its internal classification, and multiple disagreements on the right way to reconstruct Proto-Afroasiatic phonology and lexicon. None of these problems, however, have managed to shatter the historical validity of Afroasiatic as a genetic unit; and no matter how old it is, how many Afroasiatic etymologies can be considered established beyond reasonable doubt, or how many languages it actually contains, there is — at least in theory — absolutely no reason to consider its current borders permanently closed to the acceptance of new branches.

Furthermore, *pace* the concerns expressed in the final section of the response offered by Sands et al. 2023, I cannot see any waterproof historical reasons to regard the Hadza-Afroasiatic scenario as an *a priori* highly unlikely one. Militarev's explanation proposes the situation of a language shift, in which the Hadza people would adopt the (possibly more “prestigious”, under historically obscure circumstances) language of their Afroasiatic neighbors; such an idea not only explains why Hadza, as opposed to early Afroasiatic speakers, remain foragers rather than pastoralists or agriculturalists, but is also relatively consistent with the African situation as a whole, where numerous cases of similar language shifts, from the «Pygmies» to the more geographically close ethnicities linguistically belonging to the South Nilotic group (Rottland 1982), have been detected. While I agree with Sands et al. that a more concise scenario, grounded in ethnographic and archeological reality, would be welcome in this situation, I also concur with Militarev that it makes little sense to spend time and effort on building up such a scenario *before* the Hadza-Afroasiatic connection has been demonstrated on solid linguistic ground. The question, then, is whether the connection has really been demonstrated, or, as Sands et al. claim, the presented evidence is thoroughly insufficient for such a purpose.

When we postulate a linguistic scenario for the affinity between a linguistic isolate and a macro-family, from a purely phylogenetic point of view there are two logical possibilities. One is that the isolate may be a top level division of the macrofamily, i.e. historically the first branch to split from the common protolanguage, or one of the first several branches in the potential case of an original multifurcation. In the case of Afroasiatic, which Militarev currently dates to (approximately) the 11th millennium BC (Militarev & Nikolaev 2020: 200), this would mean approximately 12–13 thousand years of independent development (including phonetic change and gradual cognate loss) for Hadza. Lexicostatistical calculations, be they carried out according to the Swadesh formula ( $c = e^{-0.14t}$ , where  $c$  = percentage of retentions,  $t$  = time elapsed) or the revised Starostin formula ( $c = e^{-0.05ct^2}$ ), would in this case both agree that Hadza could theoretically retain about 20% of its basic vocabulary (= 100-item Swadesh wordlist) in this time interval, although, of course, (at least if we adopt Starostin's method) the actual number of items would be much smaller than 20 because of all the extra borrowings received along the way (and also because at least a few of these items could be retained in Hadza, but not in any other branch of Afroasiatic and therefore become technically unrecoverable as genuine Afroasiatic legacy).

While 20% (give or take a few) could look like an impressive figure by itself, lexicostatistics and glottochronology typically deal with pairwise comparisons; and the same formulae, when applied to pairs of languages, naturally yield much smaller figures. Again, both in Swadesh's and Starostin's formulae a period of divergence between two languages that is equal to approximately 10,000 years translates into a figure of approximately 5–6% of matches between the two ( $c = e^{-0.28t}$  for Swadesh,  $c = e^{-0.1\sqrt{ct^2}}$  for Starostin); for 12,000–13,000 years of separate development we should expect an even smaller number of matches. Glancing at Mili-

tarev's figures, one does indeed often find these kinds of values (around 5–6% of matches with many random Omotic, Cushitic, Chadic, etc. languages), although there are occasional unexpected surges (on which I shall comment later) and equally unexpected drops (e.g. only 3% with South Omotic and a measly 2% with Qwadza).

The problem is — how would a 5-6% amount of matches (remember that these are *percentages*; elimination of borrowings from the 100-item wordlist means that the actual number of matches is even lower) be distinguishable from chance resemblance? Recent experiments conducted, e.g., on the basis of data included into the Global Lexicostatistical Database, show that 5-6% of accidental similarities between two unrelated languages is quite a realistic figure. The only way to increase the significance of such pairwise comparisons, then, would be to demonstrate that they are not pairwise, but N-wise, i.e. that Hadza regularly yields matches with not just *one*, but *several* branches of Afroasiatic at the same time — and that in all such cases, we are genuinely dealing with a solid lexical candidate for the Proto-Afroasiatic level.

Unfortunately, Militarev's paper does not summarize specific numbers of three-way, four-way, etc. matches between Hadza and the different branches or subbranches of Afroasiatic; these have to be calculated by the reader based on the data he provides. However, while this is not an impossible task, it is not quite clear whether it is a useful one, because it is not enough to merely find look-alikes — it is just as important to demonstrate their *reconstructibility*. That is, if a certain item in Hadza is *lexicostatistically* compared to a certain item in one or two Chadic languages and a certain item in one or two Cushitic languages, it is of vital importance to the comparison to be able to show that both in Chadic and Cushitic, the item in question has a solid chance of reflecting the respective Proto-Chadic and Proto-Cushitic item.

To take one specific example of how this principle is undermined, let us look at the word 'eat' (Hadza *seme* ~ *simi*), which is compared by Militarev to two counterparts in Egyptian and Chadic, both of which are marked with = (symbolizing exact lexical match). Upon first glance, the parallel with Egyptian *sɛm* 'eat' and Proto-Chadic *\*sVm-* 'eat' may seem like an impressive three-way consonantal match. However:

(a) the Egyptian word is by no means the original Egyptian basic word with this meaning. It is not attested at all in Old Egyptian, and its first and most common meaning is rather 'to swallow' ("schlucken lassen, hinunterspülen lassen" in Hannig 2006: 2107), from which figurative extensions to both 'drink' and 'eat' are occasionally encountered. Meanwhile, the basic and most common equivalent for 'to eat' throughout the entire history of Egyptian and Coptic is the verb *wnm*, which by no means matches Hadza (or Chadic, for that matter);

(b) the Chadic equivalent is put together from the data of two Western languages (Angas and Sura) and a few more Central languages, clearly insufficient to ensure the Proto-Chadic status of this item, much less in the basic meaning 'eat'; in terms of semantics and distribution, the ideal candidate for the status of Proto-Chadic 'eat' is the lexical root *\*ti/aw/?-* (Stolbova 2016: 317), well represented in all the three major branches of that family in precisely the requested meaning. Again, no match with Hadza.

From this point of view, Hadza *seme* ~ *simi* would be much better compared with, for instance, Austroasiatic, where Ilia Peiros in his database reconstructs a hypothetical *\*sVm* 'eat' based on Proto-Palaungic *\*som* and Proto-Munda *\*ɜVm* (although the root is not attested in any other branches of the family, it is at least quite safely reconstructible in this basic meaning to the top level of Proto-Palaungic and Proto-Munda; cf. also *\*cuum* ~ *\*cəm* 'to eat' in Shorto 2006: 364).

Just a few paragraphs away, we find Hadza *furu-ne* 'to be many, plenty, full' (not even the most default equivalent for the meaning 'full' in Hadza itself) compared with Zenaga *tu-fur-t* 'full' (at least this item has a very slim chance of going back to Proto-Berber, given the isolated status of Zenaga within this family); a single dialectal Hausa form (!) representing the entirety

of Chadic; and a late Egyptian form  $\text{ḥpr}$  that is clearly not the most basic term for this notion in Egyptian (it is  $\text{mḥ}$ , well attested at all stages from the Old Kingdom to Coptic). Below that is the comparison of Hadza  $\text{kwe-}$  'to give' with a small handful of Berber and Chadic forms whose Proto-Berber and Proto-Chadic statuses are not assured in the least, let alone their Proto-Afro-Asiatic antiquity. Meanwhile, from Peiros' Austroasiatic database, compare  $\text{*b(?)i:r}$  'full', with reflexes in Khmer and Vietnamese, and  $\text{*?Vk}$  'to give', with the same monoconsonantal match as in the presented hypothetical etymology (the Austroasiatic comparison is, of course, not to be taken seriously, merely to underscore how generally easy it is to find potential cognates in a significantly large pool of comparative data extracted from one family).

Admittedly, the author himself understands the issue at stake, adding in a footnote that "Hadza-AA matches representing a common AA... or at least a common AA branch root... are, of course, of much better *quality* than Hadza matches with a few isolate and disperse AA terms". The ensuing appeal to relative statistics, however, does not seem like a direct way to solve the outlined problem. Even if it *is* true, as claimed by Militarev, that Hadza consistently yields higher numbers of phonetic and semantic matches with various branches of Afro-Asiatic than Nubian or Kuliak, genetic relationship is not the only possible explanation in this case; at least some such surges may be caused by areal contact, and this is why it is particularly important to analyze them closely in order to understand whether genetic inheritance from a common ancestor is a more logical and economical reason for the matching.

This is where the second phylogenetic possibility comes into focus: namely, a specific surge in matches between Hadza and one specific branch of Afro-Asiatic could hint at Hadza being not one of the top-level branches, but rather a historical offshoot of some subdivision of this macrofamily, such as Chadic or Cushitic. Indeed, were this actually the case, it might have been easier to demonstrate the Afro-Asiatic affinity of the language isolate in question — simply because the absolute chronology of a "Hadza-Chadic" or "Hadza-Cushitic" taxon would be smaller, and thus, we could expect a relatively larger number of cognates with better identifiable phonetic correspondences.

Unfortunately, this possibility manifests itself in arguably the least useful way of all: according to Militarev, Hadza does in fact show a slightly closer affinity with both the Omotic and the Cushitic families of the macrofamily, but since Militarev's own glottochronology has Cushitic and Omotic splitting around the 10th millennium B.C. — barely a thousand years after the primary split of Proto-Afro-Asiatic itself — this has hardly any phylogenetic or etymological significance. Moreover, a slight increase in the number of matches between Hadza and "Cush-Omotic" (a highly dubious taxon in itself, according to my own lexicostatistical calculations) could be easily explained — at least in theory — by areal contact between Hadza and these branches, both of which (especially Cushitic) are Hadza's closest neighbors of all the Afro-Asiatic stock.

An additional observation about "the high percentage of coincidences with individual Chadic languages (Tera, Mubi, etc.), which is not easy to explain" puts the entire comparison in even higher jeopardy. Clearly, if Hadza is genetically related to Chadic, its lexicostatistical matches must be with Proto-Chadic, not with individual Chadic languages. If there are more such matches between Hadza and Tera or Mubi than there are between Hadza and Proto-Chadic, such a circumstance may be explained in three ways:

(a) the respective Tera or Mubi equivalents (for instance, Mubi  $\text{mdé}$  'good' = Hadza  $\text{muta-}$ ; Tera  $\text{kiya}$  'who?' = Hadza  $\text{*?ak}^w$ -) do indeed go all the way back to Proto-Chadic roots with the same basic meanings. In this case, credible etymological scenarios must be proposed to demonstrate this, and explain why it is those roots and not the ones with wider distribution across the Chadic continuum that should be projected back to Proto-Chadic. Until this has been done to general satisfaction, such an explanation must be rejected;

(b) the extra links with Tera, Mubi, etc. represent areal contacts between speakers of Hadza and those of various subbranches of Chadic already after the split of Common Chadic into several distinct lineages. This is quite implausible on geographical and historical grounds, and would be nothing short of a miracle were it convincingly demonstrated;

(c) the extra links with Tera, Mubi, etc., are accidental resemblances. Given that, in absolute numbers rather than percentages, what we are talking about here is, at most, 2-3 cases out of 50, this is much less incredible than it might seem to the author of the hypothesis<sup>2</sup>.

Returning to the extra links with Cushitic and Omotic, Militarev's occasional reasoning against explaining these as results of borrowing feels equally unsatisfactory. Thus, when commenting upon the striking similarity between Hadza *mitl'a* 'bone' and Dahalo *miĉĉ-o* id., he writes: "...lack of other known Dahalo loanwords in Hadza speaks against the idea of 'bone' (which belongs to the most stable part of the core wordlist and is borrowed extremely rarely) to be the only word borrowed into Hadza from Dahalo". However, if the Dahalo word is indeed traceable back to Proto-Cushitic *\*miĉ-*, this means that it is not necessarily Dahalo that might have served as the source of the borrowing, but any other Cushitic branch or language that was still preserving the lateral articulation of the affricate at the time of contact. As for the argument about rare borrowing, consider the situation of the nearby Ethiopian isolate Shabo (whose speakers are, in many ways, sharing the same conditions today as the Hadza) — its basic lexicon is, to a large degree, autochthonous, but the small bunch (about 10% out of the Swadesh 100-item wordlist) of recent borrowings from nearby Ethiosemitic, Omotic, and Surmic languages does include *ema-ka* 'bone' ← Majang (North Surmic) sg. *eme-nan*, pl. *eme-k* 'bone', implying that such a borrowing is not at all implausible in that region (Starostin 2017: 715).

In the end, arguably the only genuinely impressive piece of evidence that could tie Hadza to Afro-Asiatic as a whole is probably the Hadza paradigm of 1<sup>st</sup>/2<sup>nd</sup> person pronouns, which seems to be patterned along the same "N/T" principle as the majority of Afro-Asiatic: the opposition of *ono* 'I' vs. *t<sup>h</sup>e* 'thou' is comparable to Proto-Afro-Asiatic *\*ʔanV* 'I' : *\*(ʔan-)tV* 'thou' — moreover, this seems to be a more or less exclusive isogloss between Hadza and AA, as the other pronominal systems on the African continent all seem to follow significantly different patterns. Considering that such pronominal paradigms are indeed among the most stable and long-lasting «building blocks» of language, this, in itself, would be a fairly strong argument in favor of the Afro-Asiatic roots of Hadza. Unfortunately, accidental paradigmatic matchings of this kind do occasionally happen, and without any additional corroboration this single argument will probably be not enough to validate the hypothesis.

Note also that the pronominal argument, *pace* Militarev, does not extend to the 1st pl. pronoun: Hadza inclusive 'we' *uni-bii*, compared by Militarev with Afro-Asiatic *\*nV(h)-* 'we', is to be morphologically analyzed as *u-ni-bii*, where *-ni-* is not a pronominal morpheme but rather a suffixed marker of inclusivity — cf. *u-bii* 'we (excl.)', as well as the corresponding female forms *o-bee* (excl.), *o-ne-bee* (incl.), clearly showing that the pronominal root here is *\*u ~ \*o*. Accepting Militarev's comparison is only possible if we decide that Proto-Afro-Asiatic *\*nV(h)-* is originally a reinterpreted marker of inclusivity — an extremely implausible decision based on distributional grounds, and one that finds no support on properly Afro-Asiatic grounds. Consequently, this "match" has to be rejected, like so many others.

<sup>2</sup> To illustrate the possibility of chance resemblances, consider the following semantically exact and phonetically plausible "matches" between Ari (South Omotic) and Modern English: Ari *de?* = 'die'; Ari *ʔa:ni* = 'hand'; Ari *ʔi* = 'I'; Ari *na:mi* = 'name'; Ari *kay-* = 'go'; Ari *wo<sup>h</sup>* = 'we'. Note that all of these words, with the exception of 'go', belong to the "ultra-stable" 50-item part of the Swadesh wordlist; and that a few more cases could be easily added by setting up some simple phonetic rules (e.g. Ari *se:n-* = 'stone' if one sets up a perfectly plausible rule of initial cluster simplification in Ari).



Of the non-pronominal comparisons, probably the only impressive match between a fairly widely distributed and reliably reconstructible (in the required basic meaning) Afro-Asiatic root and Hadza is the word for 'blood': Hadza *át<sup>h</sup>a?má-* = AA *\*(?a-)dam-*, which is also likely to be the optimal candidate for 'blood' on the Narrow Afro-Asiatic (Semitic + Berber + Chadic) level. The biconsonantal match is difficult to brush away as a mere accident (though, of course, accidental biconsonantal matches are quite well-known in comparative linguistics), and impossible to explain as the result of contact with Cushitic or Omotic (since it is not attested in precisely those branches). Even so, a hypothesis of Hadza-Afro-Asiatic relationship whose only strong points — by "strong" I mean "etymological arguments resting on significant phonetic, semantic, *and* distributional evidence" — are confined to the 1st-2nd pronominal paradigm and the word for 'blood' would be considered a shaky hypothesis indeed.

My own general methodological stance, on which I have written in detail in many previous publications, is that we do not necessarily require some sort of "bulletproof", 100% rigorous evidence of genetic relationship in order to label some particular hypothesis of the latter as "promising" or "deserving of further investigation"; all hypotheses of genetic relationship can — at least in theory — be ranged along a probability scale. Thus, even the staunchest opponent of the Altaic hypothesis would probably have to admit that it is more *likely* for Turkic languages to be genetically related to Mongolic or Tungusic than to, say, Niger-Congo — and not just for reasons of geographic proximity, but for actual linguistic evidence as well (e.g. the remarkable similarity in pronominal systems). For this reason, I do not rule out the Hadza-Afro-Asiatic connection as impossible or implausible: a small part of the evidence collected by Militarev allows for a genetic explanation. However, this is such a small part that even accepting the Hadza-Afro-Asiatic connection as a "working" hypothesis can hardly be done before it is clearly and explicitly shown that this evidence is statistically more significant than, for instance, the evidence that links Hadza to the various "Khoisan" families, or, for that matter, any other genetic lineage on the African continent.

In conclusion, I have to stress once again — as I have already done previously in another critical assessment of the same author's conception of Afro-Asiatic (Starostin 2021) — that our current state of awareness on lexical data around the world, their internal connections and their degrees of (both accidental or non-random) similarity, strictly prohibits to treat macro-families like Afro-Asiatic the same way we would treat more shallow families like Indo-European. A lexicostatistical comparison between a set of living Afro-Asiatic languages and Hadza is about as useful as a comparison between a set of living Indo-European languages, Finnish, and Turkish to confirm or refute the Nostratic hypothesis (to the best of my knowledge, not even Sergei Starostin, to whose authority the author of the discussed paper constantly refers, ever engaged in such comparisons, sticking instead to calculating matches between more or less reliably reconstructible Proto-Indo-European, Proto-Uralic, Proto-Turkic, etc. etyma). The only truly meaningful procedure in this case requires, first and foremost, a diligent and accurate preparation of lexicostatistical lists for all the reliably reconstructible stages of the branches of Afro-Asiatic, removing all or most of the transparent innovations from comparison as obvious sources of "noise" for any further external comparison — a procedure that, as some of the examples discussed above demonstrate, would certainly take a lot of the Hadza-Afro-Asiatic comparanda proposed by Militarev out of the equation, but at the same time could perhaps strengthen the validity of some of the others. Until such lists (accompanied with all the necessary etymological explanations) have been made available, lexicostatistical demonstration of the Afro-Asiatic affinity of Hadza (or, in fact, any other potentially Afro-Asiatic language) is, almost by definition, impossible.

## References

- Hannig, Rainer. 2006. Ägyptisches Wörterbuch II / Hannig-Lexica 5: Mittleres Reich und Zweite Zwischenzeit. Kulturgeschichte der antiken Welt 112. Mainz: Philipp von Zabern.
- Militarev, Alexander, Sergei Nikolaev. 2020. Proto-Afrasian names of ungulates in light of the Proto-Afrasian homeland issue. *Journal of Language Relationship* 18(3): 199–226.
- Obst, Erich. 1912. Von Mkalama ins Land der Wakindiga. II. Die Sprache der Wakindiga. *Mitteilungen der Geographischen Gesellschaft in Hamburg* 26(1): 29–45.
- Rottland, Franz. 1982. *Die Südnilotischen Sprachen: Beschreibung, Vergleichung und Rekonstruktion*. Berlin: Dietrich Reimer Verlag.
- Sands, Bonny, Maarten Mous, Mauro Tosco, Andrew Harvey. 2023. Why Hadza is (probably) not Afroasiatic: a discussion of Militarev’s “Hadza as Afrasian?” *Journal of Language Relationship* 21(2): ???.
- Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary*. Canberra: Pacific Linguistics.
- Starostin, Georgij. 2008. К вопросу о генетической принадлежности языка хадза [On the issue of the genetic affinity of Hadza]. In: V. F. Vydrin (ed.). *Afrikanskiy sbornik 2007*: 262–278. Saint Petersburg: Nauka. 262–278.
- Starostin, Georgij. 2013. *Jazyki Afriki: opyt postrojenija leksikostatisticheskoy klassifikacii. Tom I: Metodologija. Kojanskije jazyki*. Moscow: Jazyki slav'anskoj kul'tury.
- Starostin, Georgij. 2017. *Jazyki Afriki: opyt postrojenija leksikostatisticheskoy klassifikacii. Tom III: Nilo-saharskije jazyki*. Moscow: Jazyki slav'anskoj kul'tury.
- Starostin, George. 2021. The Difficulties of Reconstructing the Cultural Lexicon for a Macrofamily-Level Proto-Language (Based on the Afrasian Example). *Etnograficheskoe obozrenie* 2021(4): 267–273.
- Stolbova, Olga. 2016. *Chadic Etymological Dictionary*. Moscow: Institute of Oriental Studies, Russian Academy of Sciences.
- Theil, Rolf. 2006. Is Omotic Afroasiatic? Ms. Available online at: <https://www.uio.no/studier/emner/hf/iln/LING2110/v07/> (accessed 06.10.2023).

*Г. С. Старостин.* Как можно было бы показать, что хадза — афразийский язык (ответ на статью А. Ю. Милитарева «Хадза — афразийский язык?»)

В настоящей статье, представляющей собой ответ на гипотезу А. Милитарева об афразийской аффилиации языка-изолята хадза, проводится попытка определить главные теоретико-методологические недостатки лексикостатистического анализа А. Милитарева и показать, почему этот анализ нельзя считать доказательным. В качестве конструктивной альтернативы изложено, как метод, используемый А. Милитаревым, может быть усовершенствован для получения более надежных и исторически достоверных результатов; важнейшим условием такого усовершенствования является переход от использования данных живых языков к методу ономазиологической реконструкции 100-словных списков для промежуточных праязыковых состояний внутри афразийской макросемьи.

*Ключевые слова:* языки-изоляты; хадза язык; койсанские языки; афразийские языки; дальнейшее родство языков; лексикостатистика.

## Классификация восточной ветви диалектов группы Минь и реконструкция правосточноминьского 100-словного списка

Статья посвящена исследованию внутренней классификации восточноминьской подгруппы миньской ветви синитической семьи на основании анализа ее базисной лексики; попутно осуществляется попытка реконструировать праязыковое состояние на материале 100-словного списка Сводеша. В результате анализа в восточноминьской базисной лексике удастся не только выделить явный слой австроазиатской лексики, но и обосновать возможное существование архаичного субстрата в современном ареале распространения восточноминьских диалектов, оставившего следы (как минимум) на лексическом уровне. Новое филогенетическое дерево подтверждает распределение на две традиционно принятые подветви (фуниин и хоугуань), основанное на фонетических критериях, но с небольшими отклонениями. Исследование также раскрывает взаимоотношения восточной группы с диалектом путянь, подтверждая его независимый (как от восточных, так и от южных диалектов) статус ввиду отсутствия общих инноваций.

*Ключевые слова:* диалектология; древнекитайский язык; среднекитайский язык; миньские диалекты; восточноминьские диалекты; диалект путянь; лексикостатистика; базисная лексика.

### Введение

Настоящая работа посвящена проблемам лексикостатистического анализа восточных диалектов группы Минь<sup>1</sup> синитической ветви сино-тибетской семьи, распространенных на территориях провинций Фуцзянь и Чжэцзян (КНР), а также во многих странах Юго-Восточной Азии. Она представляет собой логическое продолжение предыдущей работы автора по миньским диалектам (Лоренц 2020), главной целью которой являлось составление аннотированных списков базисной лексики на материале всех существующих ветвей группы Минь и попытка ономазиологической реконструкции аналогичного списка для праминаьского состояния.

Исследование, с одной стороны, подтвердило многослойность лексического фонда изучаемой языковой группы, в котором, помимо предсказуемых параллелей, генетически унаследованных от классического и позднедревнекитайского языка, были обнаружены лексемы, скорее характерные для раннедревнекитайских диалектов, а также потенциальные заимствования из австроазиатской, австронезийской и тай-кадайской семей. Однако решить все проблемы, возникающие в ходе лексикостатистического анализа и основанной на нем филогенетической классификации миньских диалектов, в исследовании не удалось, в первую очередь — ввиду недостаточного объема материала, отобранного по отдельным ветвям миньской группы. Во многом предварительная реконструкция праминаьского списка, приведенная в статье, остается приблизительной, поскольку не использовала в должной степени «ступенчатый» принцип, предполагающий последовательное составление промежуточных списков для отдельных ветвей и уже затем све-

<sup>1</sup> Подробную классификацию см. в LACD 2012: 177–179.

дение их воедино для реконструкции общеминьского состояния. В отдельных случаях нарушение этого принципа может (по крайней мере теоретически) приводить к неучету ареальных связей (например, взаимное влияние территориально близких диалектов, относящихся к генетически разным ветвям), что, в свою очередь, может пагубно сказаться на отборе «оптимальных кандидатов» для реконструкции общеминьского списка. Реально полноценный анализ праминьского состояния возможен только при наличии отдельных праязыковых списков хотя бы по всем тем миньским ветвям, которые надежно (т. е. без существенных противоречий с общепринятыми классификациями и вне зависимости от применяемого метода построения дерева) выделяются на генеалогическом дереве.

Одной из бесспорных и однозначно (вне зависимости от выбранного метода и параметров) выделяющихся ветвей можно считать восточноминьскую (миньдунскую). Она была отобрана в качестве первого звена для применения «ступенчатого» метода реконструкции постольку, поскольку обнаруживает некоторые уникальные лексические особенности. Так, во всех языках-потомках этой ветви сохраняется архаизм раннедревнекитайского и классического древнекитайского (далее по тексту — РДК и КДК соответственно) периодов 'собака' (犬 \*k<sup>hw</sup>ín), который на момент распада праминьской общности (III–IV вв. н.э.) уже был вытеснен (по крайней мере, в общенациональном литературном языке) инновативной лексемой 狗 КДК \*kó (что исключает возможность его «восстановления» на общевосточноминьском уровне под влиянием литературного языка). Уже сам по себе этот факт может поставить под сомнение существование миньского единства как такового и задуматься о возможности обособления восточноминьского таксона в самостоятельную диалектную группу уже в классическую эпоху (например, где-нибудь в диапазоне IV–III вв. до н.э.).

Во-вторых, на филогенетическом дереве, полученном в ходе первого этапа нашего исследования при глоттохронологическом анализе данных с помощью компьютерной среды StarLing (рис. 1), были отмечены некоторые странности, плохо согласующиеся с устоявшимися взглядами на аффилиацию восточноминьских диалектов. Так, ближайшим родственником восточноминьской ветви на дереве оказался диалект путянь, что чрезвычайно сомнительно; при добавлении же несколько менее надежных данных по диалектам датянь и юси диалект путянь вообще включается глоттохронологическим алгоритмом внутрь восточной ветви (рис. 2). Решение последней проблемы, впрочем, следует отложить до появления более подробных и надежных лексических данных по датянь и юси (очевидно, что на результаты негативно влияет большое количество лакун в 100-словном списке)<sup>2</sup>, так что в данной работе мы сконцентрируемся исключительно на разрешении проблемы взаимоотношений между восточноминьскими диалектами и путянь.

Исследователями было неоднократно отмечено, что т.н. пу-сяньская подгруппа, включающая в себя диалекты путянь и юсянь, представляет собой отдельную «пограничную» ветвь, которая имеет общие черты как с восточными, так и с южными диалектами Минь, но, согласно общепринятому мнению, эти диалекты на уровне фонетики, лексики и грамматики все же оказываются ближе к южной группе и, скорее всего, отделились от южных диалектов, а уже затем попали под влияние восточных (Cai 2006: 237–239). Видно, однако, что как минимум на уровне первичного, «грубого» лексикостатистического анализа эта концепция не находит однозначного подтверждения и, следовательно, нуждается в более тщательной проверке.

<sup>2</sup> Диалектологи обычно выделяют эти диалекты в отдельный «пограничный» узел миньской ветви, так как из-за своего географического расположения они находятся под сильным влиянием одновременно центральной, северной, восточной и пусяньской ветвей, поэтому для точной классификации необходимо проводить очень тщательный анализ на уровне субдиалектов (Hou 2002: 240–241; Liu & He 1998: 383).

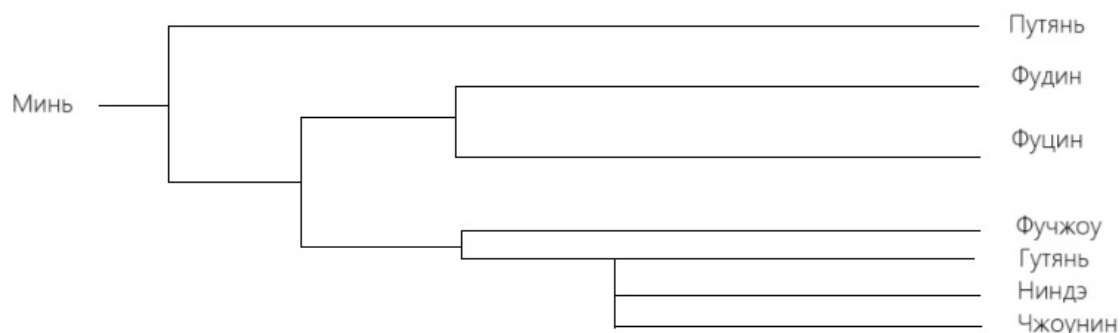


Рис. 1. Классификация восточноминьской ветви с включением диалекта путянь.

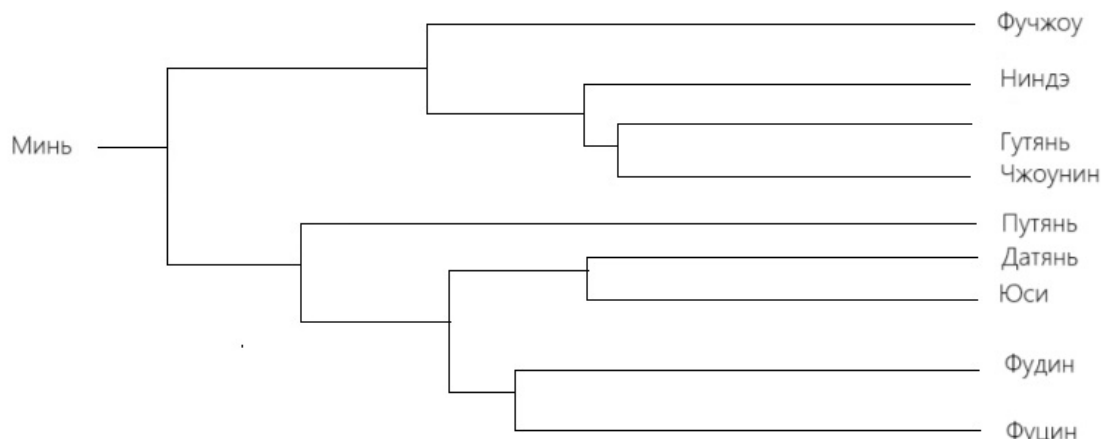


Рис. 2. Классификация восточноминьской ветви с включением диалектов путянь, датянь и юси.

Для того, чтобы максимально надежно определить место восточноминьского таксона внутри (или, ввиду таких уникальных архаизмов, как 'собака', *вне*) миньской общности, а также четко определить границы восточноминьского узла и уточнить внутреннюю классификацию входящих в него диалектов, необходимо составить обновленные и уточненные списки базисной лексики соответствующих диалектов, распространенных на территории провинции Фуцзянь<sup>3</sup>, а также провести ономаσιологическую реконструкцию аналогичного списка для правосточноминьского состояния (напомним, что в нашей предыдущей работе такая реконструкция — в предварительном порядке — предлагалась лишь для общеминьского состояния, без детального учета данных на промежуточных уровнях). Представляется, что после того, как соответствующий анализ позволит отсечь от рассмотрения всевозможные ареальные инновации, на основании сопоставления полученного прасписка с диалектом путянь и с другими миньскими диалектами можно будет сделать более уверенные выводы об их генетических и ареальных связях.

### Предыстория вопроса

Опубликованные на настоящий момент статьи и монографии по диахроническим аспектам восточноминьской ветви, как правило, сфокусированы на вопросах сравнительной фонологии, не уделяя существенного внимания лексической реконструкции.

Прежде всего необходимо отметить краткий обзор наиболее важных общих лексических, фонетических, тональных особенностей пяти восточноминьских диалектов

<sup>3</sup> В работу, к сожалению, не удалось включить данные по восточноминьским диалектам провинции Чжэцзян из-за отсутствия подробных источников.

(фучжоу, фуцин, чжэян, фуань, ниндэ) Дж. Нормана (Norman 1977) на основании данных, собранных автором самостоятельно в ходе работы с информантами. В работе также представлен список инициалей и финалей для каждого из этих диалектов и перечень 422 однословных; однако автор концентрируется исключительно на описании данных диалектов, без детального изучения фонетики или лексики.

Первой крупной работой, имеющей отношение к фонологической истории именно восточноминьского региона, является монография А. Хироюки (Hiroyuki 2018), однако она полностью посвящена фонологической реконструкции т.н. диалекта «пра-ниндэ» на основании новейших, ранее не опубликованных данных по диалектам, расположенным в соответствующем округе (их главная особенность — сохранение терминалей *-m/-p*, *-n/-t*, *-ŋ/k*, *-ʔ*). Исследование проводилось на материале словников, собранных самим автором, без подробных лексических комментариев, однако оснащенных небольшим количеством контекстов. Автор реконструирует не только систему инициалей и рифм, но и тональную систему восстанавливаемого им языка; при этом, однако, в работе нет четкого обоснования той идеи, что анализируемые в ней диалекты действительно составляют единый «эксклюзивный» кластер внутри восточноминьской ветви.

Другая работа того же автора (Hiroyuki 2020) посвящена исследованию фонологии и лексики четырех восточноминьских диалектов фуцин, шоунин, фуань и пиннань, чтобы уточнить внутреннюю классификацию изучаемой ветви; в ней же представлена фонологическая реконструкция финалей восточноминьского праязыка. В отдельных случаях А. Хироюки приводит примеры лексической реконструкции с привлечением большого количества сравнительных данных, однако системная реконструкция лексических списков для него не является первостепенной задачей.

### Источники и предварительная классификация

По сравнению с работой Лоренц 2020, в которой использовались данные по диалектам фучжоу, ниндэ, гутянь, чжоунин, фуцин и фуцин восточной ветви, для задач данной работы список исследуемых диалектов был значительно расширен. Также в ней исправлены неточности и отдельные ошибки по данным тех диалектов, которые уже были проанализированы ранее. К сожалению, настоящая статья все равно не может учитывать данные всех восточноминьских диалектов, поскольку основными источниками для составления списков служат словники, а из-за большей концентрации внимания к изучению фонетики и грамматики, чем лексики, корректно и детально составленные словники по многим диалектам просто отсутствуют. Ввиду этого в новую лексикостатистическую базу по восточноминьским диалектам из ранее не использованных вошли базисно-лексические данные только по диалектам фуань, шоунин, пиннань, хубэй, сяньцунь и цзюду.

Всего для целей настоящего исследования удалось, таким образом, собрать полноценные списки по 14 диалектам восточноминьской ветви (таблица 1). В качестве основных источников использовались в большей степени словники и в меньшей — словари. Сведения по диалекту путянь в основном базируются на источнике Nakajima 1979, но с некоторыми уточнениями, почерпнутыми из электронного словаря Hinghwa Dialect Digital Dictionary Project (Hinghua yuji) (<https://hinghua.cn/>)<sup>4</sup>.

<sup>4</sup> Полностью в данной статье список базисной лексики по диалекту путянь не приводится, поскольку использовать его для реконструкции восточноминьского состояния было бы некорректно; к тому же он уже был полностью опубликован в работе Лоренц 2020.

| Диалект      | Ветвь    | Сокращенное обозначение | Источники данных                              |
|--------------|----------|-------------------------|---|
| Гутянь хуа   | Фуниин   | GTN                     | Chen & Li 1991; Lin 2002; Liu & He 1998       |
| Ниндэ хуа    | Хоугуань | NND                     | Chen & Li 1991; Lin 2002; Liu & He 1998       |
| Пиннань хуа  | Фуниин   | PIN                     | Hiroyuki 2020; Liu & He 1998                  |
| Сяньцунь хуа | Хоугуань | XNC                     | Hiroyuki 2018                                 |
| Фуань хуа    | Фуниин   | FUA                     | Lin 2002; Hiroyuki 2020; Liu & He 1998        |
| Фудин хуа    | Хоугуань | FDN                     | Chen & Li 1991; Liu & He 1998                 |
| Фучжоу хуа   | Фуниин   | FCH                     | Nakajima 1979; Li & Liang 1994; Liu & He 1998 |
| Фуцин хуа    | Фуниин   | FQN                     | Feng 1993; Hiroyuki 2020                      |
| Хубэй хуа    | Хоугуань | HUB                     | Hiroyuki 2018                                 |
| Цзюду хуа    | Хоугуань | JDU                     | Hiroyuki 2018                                 |
| Чанлэ хуа    | Фуниин   | CNL                     | Lin 2002, Liu & He 1998                       |
| Чжоунин хуа  | Хоугуань | ZHN                     | Chen & Li 1991; Liu & He 1998                 |
| Шоунин хуа   | Хоугуань | SHON                    | Lin 2002; Hiroyuki 2020; Liu & He 1998        |
| Юнтай хуа    | Фуниин   | YNT                     | Lin 2002; Liu & He 1998                       |

Таблица 1. Перечень восточноминьских диалектов, учтенных в настоящей работе.

Наиболее широко распространенной внутренней классификацией миньдунских диалектов считается схема, предлагаемая в «Лингвистическом атласе китайских диалектов» (LACD 2012). В атласе предлагаются три фонетических критерия для внутреннего деления миньдунской группы на две подветви — *фуниин* и *хоугуань*:

— все слова ветви *хоугуань*, в которых присутствует медиаль или слогаобразующий гласный *y*, в диалектах ветви *фуниин* соответствуют словам с гласным *i*, например, фучжоу 橋 *kyo*<sub>2</sub>, 獻 *hyɕ*<sub>4</sub>, но фуань 橋 *kiu*<sub>2</sub>, 獻 *hiŋ*<sub>4</sub> и т. д.;

— в ветви *хоугуань* финали могут иметь разную фонетическую реализацию в зависимости от тона, однако эта особенность не прослеживается в ветви *фуниин*. Так, например, в диалекте фучжоу финаль *-iŋ* в восходящем тоне (上聲) произносится без изменений (等 *tiŋ*<sub>3</sub>), но меняется при следующих тонах: «темный» падающий (陰去) — 鎮 *teiŋ*<sub>4</sub>, «светлый» падающий (陽去) — 陳 *teiŋ*<sub>5</sub>, «темный» входящий (陰入) — 的 *tei*<sub>6</sub>, «светлый» входящий (陽入) — 笛 *ti*<sub>7</sub><sup>5</sup>;

— часть финалей третьего дэна (по классификации среднекитайской фонетики) в диалектах *хоугуань* имеют гласные *i*, *y* там, где диалекты *фуниин* обнаруживают гласные *e*, *a*: 厘 фучжоу *li*<sub>2</sub>, гутянь *lie*<sub>2</sub>, ниндэ *le*<sub>2</sub>, фуань *le*<sub>2</sub>, 琴 фучжоу *k<sup>h</sup>iŋ*<sub>2</sub>, гутянь *k<sup>h</sup>iŋ*<sub>2</sub>, ниндэ *k<sup>h</sup>em*<sub>2</sub>, фуань *k<sup>h</sup>eiŋ*<sub>2</sub> (LACD 2012: 113–114).

Тем не менее, при отборе праязыкового кандидата в нашей ономаσιологической реконструкции мы предпочитаем не опираться априорно на данную классификацию как на установленный факт; скорее, наоборот, необходимо предварительно убедиться в том, работает ли она на лексическом материале. Таким образом, в рамках реконструкции мы не будем стремиться к тому, чтобы в соответствии со «ступенчатым» принципом восстанавливать отдельно формы «пра-фуниин» и «пра-хоугуань».

Следует добавить, что некоторые субдиалекты, локализованные возле г. Фуцин, Фудин, Фуань, согласно данному атласу относятся к *пу-сяньской* ветви, однако при этом

<sup>5</sup> Противопоставление «светлых» и «темных» тонов в классической китайской фонологии коррелирует с признаком звонкости («светлые» тона) и глухости («темные» тона) у соответствующего начального согласного (инициали).

диалекты, расположенные непосредственно в самих этих городах, все равно относятся к миньдунской ветви.

При работе с источниками по новым восточноминьским диалектам, не учтенным в работе Лоренц 2020, несколько обострилась проблема разграничения между разговорными и литературными вариантами отдельных слов; особенно часто она проявляется на материале диалектов чанлэ, юнтай, фуань, шоунин и пиннань. Нередко в источниках смешиваются литературные и разговорные нормы без каких-либо указаний на то, замещает ли в живом, разговорном, стилистически нейтральном языке «новое» литературное чтение «старое» разговорное (сохраняющееся, например, как стилистически маркированный архаизм). Как правило, литературные чтения в диалектах, появляющиеся под влиянием престижного государственного языка, не вытесняют разговорные; однако не исключены ситуации, не отраженные напрямую в наших источниках, в которых современные носители уже предпочитают использовать литературное чтение вместо разговорного — подобная проблема может быть решена только при работе с информантами.

Наглядным примером смешения разговорных и литературных чтений может служить лексема 'соль', для которой основные источники по ряду диалектов указывают только литературные чтения (CNL *siɛŋ*<sub>2</sub>, PIN *siŋ*<sub>2</sub>, FUA *siŋ*<sub>2</sub>, SHON *siɛŋ*<sub>2</sub>); однако при подключении сведений из дополнительного источника Liu & He 1998 обнаруживаются и разговорные чтения: CNL *siɛŋ*<sub>5</sub>, FUA *siŋ*<sub>5</sub>, SHON *siɛŋ*<sub>5</sub>, PIN *siŋ*<sub>5</sub>. Не располагая однозначной информацией относительно того, как реально взаимодействуют эти варианты на уровне повседневного общения, мы вынуждены прибегать к одному из двух формалистических решений: (а) либо слепо следовать за основным источником, предполагая, что базовым эквивалентом в соответствующих диалектах служат литературные чтения (и, таким образом, помечать их в лексикостатистической базе как заимствования, выкидывая из подсчетов); (б) либо последовательно обобщать данные из обоих типов источников (там, где они есть), формально помечая оба варианта как взаимозаменяемые синонимы (реально, таким образом, подсчет процентов совпадений будет вестись по разговорным чтениям, т.к. литературные все равно будут маркироваться как заимствования). Чтобы иметь возможность использовать большее количество данных, мы будем в целом придерживаться стратегии (б), за исключением тех случаев, когда «исконно-разговорное» чтение эксплицитно маркируется как стилистически окрашенное.

Разумеется, для целей ономаσιологической реконструкции эта проблема нерелевантна, т.к. при выборе правосточноминьского эквивалента предпочтение в любом случае будет отдаваться «разговорным» (т. е. исконно восточноминьским), а не «литературным» (заимствованным) чтениям. Здесь остается только проблема корректного исторического разграничения между вариантами, которая решается, как правило, на фонетическом уровне (разные слои лексики отражают разные (под)системы фонетических соответствий).

Методология ономаσιологической реконструкции, примененная в данной работе, в целом остается такой же, как и в предыдущей статье Лоренц 2020, поэтому смысла подробно описывать ее здесь нет (см. общие принципы реконструкции прасписков, описанные в работах Kassian et al. 2010, Starostin 2019).

Все древнекитайские и среднекитайские фонетические транскрипции и реконструкции, используемые в настоящей статье, приводятся согласно системе С. А. Старостина (Старостин 1989). Как и в работе Лоренц 2020 по реконструкции общеминьского состояния, в статье также использованы диахронические формы, сконструированные



Г. С. Старостиным (по системе, предложенной С. А. Старостиным) от раннедревнекитайского (РДК) к классическому древнекитайскому (КДК) и далее к позднедревнекитайскому (ПДК) состояниям (Starostin 2019). Правосточноминьские формы, отмеченные звездочкой, реконструированы лично автором статьи в соответствии с сегментными и супraseгментными регулярными соответствиями, характерными для восточноминьских диалектов; подробная таблица фонетических соответствий между инициалами, финалями и тонами учитываемых в настоящем исследовании диалектов приводится в онлайн приложении к статье, доступном на сайте журнала «Вопросы языкового родства» ([jolr.ru/jlr21/lorentz.zip](http://jolr.ru/jlr21/lorentz.zip)).

Слова, помеченные в списках символом решетки (#), взяты из определенных контекстов или устойчивых сочетаний, внутри которых имеют место тональные сандхи; в таких случаях этимологически ожидаемый тон может отличаться от реально зафиксированного в источнике.

В тех случаях, когда для того или иного диалекта лексема обнаружена только в дополнительном источнике (см. список таких источников в таблице 1), он специально отмечается сокращенно в скобках: Hiroyuki 2020 — (H), Lin 2002 — (LN), Liu & He 1998 — (LH), Li & Liang 1994 — (LL).

При расстановке индексов когнации мы объединяем базовый однослог и композиты, в состав которых входит данный однослог, в одну группу, при этом префигурируемые вариативные компоненты отделяются символом (=), а суффилируемые — символом (-). В рамках стандартного лексикостатистического обсчета данных, основанного на принципе этимологического совпадения корневой морфемы, однослоги и двуслоги (многослоги) засчитываются как лексикостатистические совпадения, так как вариативные компоненты в двусловах, как правило, выполняют лишь функцию снятия лексической омонимии, а в отдельных (связанных) контекстах могут и просто опускаться. Тем не менее, поскольку информация о морфемном составе той или иной сложной основы все же может быть в отдельных случаях релевантной для классификации, при публикации списков мы все же стратифицируем наблюдаемые варианты, выделяя основы, расширенные за счет вариативных компонентов, в отдельные подгруппы (например, ‘ухо’: а — {耳} FCH  $\eta ai_6$ , a.1 — {耳囟} FQN  $\eta i_4-(k)ian_2$ ).

Большинство лексем, претендующих на статус правосточноминьского (далее — ПВМ) эквивалента того или иного значения из 100-словного списка Сводеша, оснащены этимологическими и фонетическими комментариями, обосновывающими соответствующий выбор. Исключение — случаи, когда все или подавляющее большинство восточноминьских рефлексов отражают один и тот же древнекитайский этимон.

### Сравнительный список базисной лексики восточноминьских диалектов

1) «all/все»: {都} FCH  $tu_{u1}$ , HUB  $tu_{u1}$ , XNC  $tu_{u1}$ , JDU  $tu_{u1}$ , FQN  $tu_{u1}$  (H), SHON  $tu_{u1}$  (H), FUA  $tu_{u1}$  (H), PIN  $tu_{u1}$  || ПВМ {都}  $*tu_{u1}$

Основной КДК корень 皆  $*kr\bar{a}j$  со значением ‘все’ очевидно был вытеснен инновативным корнем ПДК 都  $*t\bar{a}$  уже на общеминьском уровне; он же, по-видимому, сохраняется и на ПВМ уровне (Starostin 2019: 171–172; Лоренц 2020: 131).

2) «ashes/пепел»: а) {火灰~火焮} FCH  $hui_2=u_1$  (LL), GTN  $xuo_{i4}=u_1$ , NND  $x\o_{y3}=xu_{u1}$ , ZHN  $huai_3=u_1$ , FDN  $xuo_{i3}=xu_{u3}$ , CHL  $xui_2=u_1$ , FQN  $uo_{i4}=u_1$ , YNT  $uo_{i3}=u_1$ , FUA  $hui_4=u_1$ , SHON  $xuo_{i4}=u_1$ ; а.1) {灰} HUB  $xu_{u1}$ , JDU  $hu_{u1}$  || ПВМ {火焮}  $*hui_3=hu_{u1}$

Старый КДК корень 灰  $*\eta\bar{a}j$  в форме однослога не зафиксирован ни в одном источнике по ВМ, кроме словаря Накадзимы, где дается форма FCH  $hui_1$  {灰}<sup>6</sup> (Nakajima 1979: 10; Starostin 2019: 165).

По данным словаря фучжоуского диалекта 灰  $hui_1$  передает значение ‘известь’, а композит {火灰}  $hui_2=ui_1$  — ‘пепел’ (Li & Liang 1994: 145–146). Важно добавить, что в диалекте FQN 灰  $huoi_1$  также передает значение ‘известь’ (Feng 1993: 170). Аналогичная информация для этой же лексемы указывается А. Хироюки для XNC, FUA (Hiroyuki 2018: 320; Hiroyuki 2020: 645).

В словнике Линь Ханьшэн по диалектам FCH, CNL, FQN, YNT, FUA, NND, SHON, ZHN, FDN слова ‘пепел’ нет, зато обнаруживается бином 火焮, который для всех диалектов маркируется со значением 草木灰 ‘зола для удобрений’, при этом второй иероглифический знак заменен (Lin 2002: 40). Иероглифический знак 焮 (СК  $x\bar{a}w$ ), используемый для записи второго слога, в литературных памятниках отсутствует; впервые он упоминается в среднекитайском словаре рифм Цзюнь (集韻) как региональное слово в узких диалектах. Однако иероглиф 灰 продолжает использоваться для передачи инновативной лексем.

Реконструкцию первого слога см. ниже (‘fire/огонь’).

Мы предполагаем, что КДК этимон был заменен на ПВМ уровне с полисемичным значением ‘зола/пепел’ (поскольку КДК терминали с конечным  $*-j$  обычно реализуются в ВМ с конечным  $-i$ , т. е. зафиксированные формы не могут отражать КДК 灰  $*\eta\bar{a}j$ ). На ПВМ уровень в статусе основного эквивалента выносятся бином {火焮}  $*hui_3=hu_1$ , ввиду того, что в большинстве языков-потомков зафиксирована именно биномиальная форма. Усечение инициали у второго слога в большинстве диалектов является закономерной особенностью, вызванной фонетическими изменениями в миньдунских диалектах на стыке слов (см. этимоны ‘огонь’, ‘облако’ ‘хороший’, в которых инициаль  $h-$  в односложных лексемах хорошо сохраняется).

### 3) «bark/кора»: FCH $ts^h i u_5=p^h u i_2$ {樹皮} || ПВМ {樹皮} $*ts^h i u_5=p^h u i_2$

Слово ‘кора’ фиксируется только в диалекте FCH, однако, исходя из данных внешнего сравнения, можно предположить, что и на ПВМ уровне, как в других миньских диалектах, также был представлен бином 樹皮 (Лоренц 2020: 131), букв. ‘дерева-кожа’. О фонетических особенностях этих корней см. ниже комментарии к этимонам ‘дерево’ и ‘кожа’.

### 4) «belly/живот»: {腹肚~腹老} FCH $pu_6=lo_3$ , GTN $pu_7=lu_3$ , NND $pok_7=lu_3$ , HUB $pu_6=lu_3$ , XNC $po_7=lo_3$ , JDU $pu_7=lu_3$ , ZHN $pu_7=lu_3$ , FDN $pu_7=lu_3$ , FQN $pu_7=lo_2$ , CHL $pu_5=lo_3$ , YNT $pu_7=lo_3$ , FUA $puk_6=lu_3$ , SHON $pu_7=lo_3$ , PIN $puk_6=tu_3$ || ПВМ {腹肚} $*puk_6=tu_3 \sim *puk_6=lu_3$

По всей миньдунской ветви видно сложение двух лексем — архаичной 腹 КДК  $*puk$  и инновативной 肚 КДК  $*d\bar{a}$  с общим значением ‘живот’. Следует добавить, что 肚 появляется только в поздних миньских литературных памятниках, предположительно с исходным значением ‘желудок’ (Starostin 2019: 167). Непосредственно значение ‘живот’ в литературных памятниках появляется только с эпохи Тан. В диалекте FQN у лексем 肚 также сохраняется значение ‘желудок’, но не в автономной форме 腸肚  $t\bar{c}\eta_2=to_5$  (Hiroyuki 2020: 707).

Данные показывают, что терминаль  $-k$  в лексеме 腹  $*puk$  в большинстве диалектов не сохранилась, а трансформировалась в гортанную смычку, что, вероятно, вызвано сандхи на стыке слов. Однако следует принимать во внимание, что в трех фонетически архаичных диалектах NND, FUA, PIN сохраняется терминаль  $-k$ , что доказывает присутствие этой тер-

<sup>6</sup> Согласно данным Macklay & Baldwin 1898: 334, разговорным вариантом для слова ‘пепел’ в фучжоуском диалекте является именно  $hu_1$ , а  $hui_1$  — литературным. Возможно, вариант, представленный в источнике Накадзимы, ошибочен.

минали на праязыковом уровне. Найденный в диалектах HUB, XNC, JDU однослог 腹 также фиксируется с терминалью *-k* (*pu<sub>7</sub>k*), что подтверждает нашу гипотезу (Hiroyuki 2018: 316).

Данные также показывают любопытную ассимиляцию на стыке двух морфем, в результате которой инициаль *t-* в морфеме 肚 переходит в *l-*, что представляет собой специфически восточноминьскую черту. Акитани Хироюки в своей работе по диалекту праиндэ описывает некоторые изменения в потоке речи инициали второго слога: так инициали *t-*, *th-*, *θ-* во втором слоге фонетически реализуются как *l-*, если после первого слога идет гортанная смычка (Hiroyuki 2018: 27); по-видимому, этот закон актуален и для большинства остальных восточноминьских диалектов. Тем не менее, в списках слов по миньдунским диалектам, составленных Линь (Lin 2002: 59), обнаружены примеры отсутствия такого фонетического изменения на стыке слов, например: GTN *pu<sub>7</sub>?=ty<sub>3</sub>* {腹肚}, FDN *pu<sub>6</sub>?=tu<sub>3</sub>* {腹肚}. Малоупотребительный однослог 肚 в диалектах HUB *tu<sub>3</sub>*, XNC *to<sub>6</sub>* также сохраняет инициаль *t-* (Hiroyuki 2018: 472).

Из сопоставления всех этих данных можно сделать вывод, что, хотя ассимиляция произошла не повсеместно, на ПВМ этапе переход *\*t-* → *\*l-* в подобного рода контекстах уже начался хотя бы на уровне отдельных говоров. По этой причине мы предлагаем восстанавливать на ПВМ уровне диалектное варьирование *\*pu<sub>6</sub>=tu<sub>3</sub>* ~ *\*pu<sub>6</sub>=lu<sub>3</sub>*.

5) «big/большой»: {大} FCH *tuai<sub>6</sub>*, GTN *tuai<sub>6</sub>*, NND *tu<sub>3</sub>*, HUB *tu<sub>6</sub>*, XNC *tu<sub>3</sub>*, JDU *tu<sub>6</sub>*, ZHN *tu<sub>6</sub>*, FDN *tua<sub>6</sub>*, FQN *tua<sub>5</sub>*, CHL *tuai<sub>6</sub>*, YNT *tuai<sub>6</sub>*, FUA *to<sub>6</sub>*, SHON *tua<sub>6</sub>*, PIN *tuai<sub>6</sub>* || ПВМ {大} *\*tuai<sub>6</sub>*

6) «bird/птица»: а) {鳥~爪} FCH *tseu<sub>2</sub>*, NND *tsiu<sub>3</sub>*, FDN *tseu<sub>3</sub>*, FQN *tseu<sub>2</sub>*, CNL *tseu<sub>3</sub>*, FUA<sup>7</sup> *tso<sub>3</sub>*, SHON *tseu<sub>3</sub>*, PIN *tseu<sub>3</sub>*; а.1) {爪~隻} GTN *tsiu<sub>3</sub>-ian<sub>3</sub>*, ZHN *tsi<sub>4</sub>-en<sub>3</sub>*, YNT *tsiu<sub>3</sub>-ian<sub>3</sub>*, HUB *tfiu<sub>3</sub>-ian<sub>3</sub>*, XNC *tfiu<sub>3</sub>-ian<sub>4</sub>~tfiu<sub>3</sub>-uan<sub>4</sub>*, JDU *tfiu<sub>3</sub>-uan<sub>3</sub>* || ПВМ {鳥} *\*tseu<sub>3</sub>*

По всей видимости, старый КДК этимон 鳥 *\*tʰw*, несмотря на нерегулярное развитие инициали *\*t-* → *ts-* вместо сохранения дентального согласного перед старым долгим гласным, стабильно продолжает использоваться в миньдунских диалектах. Подобная ситуация наблюдается и в остальных диалектах миньской группы (см. Лоренц 2020: 132). Мы склоняемся к реальности этимологической связи между КДК этимоном и ВМ лексемой по ряду причин. Во-первых, тон и финаль данной лексемы регулярно отражаются во всех ВМ диалектах. Во-вторых, нерегулярные изменения инициали у слова 'птица' зафиксированы по всему синитическому ареалу (ср. абсолютно нерегулярную носовую инициаль в совр. мандаринском *niǎo*). В-третьих, никаких убедительных альтернативных этимологий для миньской лексемы 'птица' на данный момент не обнаружено.

По-видимому, ввиду неочевидности этимологической связи между 鳥 СК *tiew* и миньскими формами, 'птица' в миньской орфографии часто записывается знаками 爪 (КДК *\*crú*) 'когти' и 隻 (КДК *\*tek*) 'один из пары', подобранными по принципу фонетической (и семантической) близости (Лоренц 2020: 133).

В диалектах GTN, ZHN, YNT базовым эквивалентом значения 'птица' оказывается двусложный вариант с диминутивным суффиксом 囡 (СК *kén*) 'сын, ребенок', который в литературных китайских памятниках отсутствует и появляется только в среднекитайских словарях. Норман и Мэй считают, что эта лексема была заимствована из австроазиатских языков, где она достаточно широко распространена: ср. кхмерский *koun*, монский *kon*, бру *kəp*, чонг *kheen* и др. (Norman & Mei 2000: 486). Употребление этого суффикса также замечено в южноминьских диалектах (Лоренц 2020: 132–133). Тем не менее, судя по ограниченной дистрибуции данного варианта, он представляет собой ареальную инновацию, и на ПВМ уровень следует выносить односложную форму *\*tseu<sub>3</sub>*, продолжа-

<sup>7</sup> Хироюки указывает для диалекта фуань форму 雀囡 *tʃi<sub>6</sub>-jiw<sub>3</sub>* ~ *tʃi<sub>4</sub>-jiw<sub>3</sub>* (Hiroyuki 2020: 674).

ющую КДК 鳥 \*tíw с нерегулярным развитием инициали (точнее, нерегулярным сокращением долготы гласного, обусловившего палатализацию).

7) «**bite/кусать**»: {咬~齧} FCH ka<sub>6</sub>, HUB ka<sub>5</sub>, XNC ka<sub>5</sub>, JDU ka<sub>5</sub>, FQN ka<sub>5</sub>, SHON ka<sub>5</sub> (H), FUA ka<sub>5</sub> (H), PIN ka<sub>5</sub> || ПВМ {咬} \*ka<sub>5</sub>

Мы наблюдаем здесь лексическую замену уже на общеминьском уровне, так как ни один из эквивалентов базового значения ‘кусать’, представленных в КДК текстах (齧 \*ŋ<sup>h</sup>ēt, 噬 \*dač), не представлен в изучаемых диалектах (Starostin 2019: 171; Лоренц 2020: 133). Напрямую связывать зафиксированные формы с более поздней (ханьской) инновацией 咬 (КДК \*ŋ<sup>h</sup>rǎw) трудно по двум причинам: (а) дополнительных примеров развития КДК инициали \*ŋ<sup>(h)</sup>- в инициаль k- в миньских диалектах не обнаружено; (б) лексема 咬 относится к тональной категории 上聲 (восходящий тон), в то время как наши данные отражают тональную категорию 去聲 (падающий тон) по всей миньдунской ветви.

По мнению А. Шюсслера, миньская основа может иметь тайское происхождение, ср. сиамский k<sup>h</sup>iau<sub>3</sub>, по-ай čeeu<sub>3</sub>, лао k<sup>h</sup>iau<sub>3</sub>, диой kiaou и др. (Schuessler 2007: 560; по данным Li 1977: 199, 201 реконструируется \*giau.C).

8) «**black/чёрный**»: {烏} FCH u<sub>1</sub>, GTN u<sub>1</sub>, NND u<sub>1</sub>, HUB u<sub>1</sub>, XNC u<sub>1</sub>, JDU u<sub>1</sub>, ZHN u<sub>1</sub>, FDN uo<sub>1</sub>, FQN u<sub>1</sub>, CNL u<sub>1</sub>, YNT u<sub>1</sub>, FUA ou<sub>1</sub>, SHON u<sub>1</sub>, PIN u<sub>1</sub> || ПВМ {烏} \*u<sub>1</sub>

Базовый КДК этимон 黑 \*ŋǎk, очевидно, заменился уже в ПМ на инновативный корень 烏 \*ʔā (с исходным значением ‘ворона’) и с производными значениями ‘темный, черный’; следов старой КДК основы не обнаружено ни в одном диалекте (Starostin 2019: 162; Лоренц 2020: 133).

9) «**blood/кровь**»: {血} FCH haik<sub>6</sub>, GTN xei<sub>7</sub>, NND xet<sub>7</sub>, HUB xet<sub>7</sub>, XNC hek<sub>7</sub>, JDU het<sub>7</sub>, ZHN xet<sub>7</sub>, FDN xe<sub>7</sub>, FQN he<sub>7</sub>, CNL haik<sub>7</sub>, YNT hei<sub>7</sub>, FUA heik<sub>7</sub>, SHON hek<sub>7</sub>, PIN xei<sub>7</sub> || ПВМ {血} \*het<sub>7</sub>

10) «**bone/кость**»: {骨} FCH kouk<sub>6</sub>, FQN ko<sub>7</sub>, SHON kǎ<sub>7</sub> (H), FUA kǎ<sub>7</sub> (H), PIN kǎ<sub>7</sub>, HUB kǎ<sub>7</sub>, XNC kǎ<sub>7</sub>, JDU kǎ<sub>7</sub> || ПВМ {骨} \*kǎ<sub>7</sub>

Непосредственно слово ‘кость’ в ряде источников отсутствует, при этом корневая морфема ‘кость’ очевидно входит в состав слова ‘ребро’: 肋排骨 FCH lɛ<sub>2</sub> pɛ<sub>2</sub> vu<sub>6</sub>, CNL lɛ<sub>2</sub> pɛ<sub>2</sub> ɔuk<sub>6</sub>, FQN lɔy<sub>7</sub> teu<sub>4</sub> o<sub>6</sub>, YNT lɔy<sub>7</sub> pɛ<sub>5</sub> ou<sub>7</sub>, GTN leik<sub>4</sub> pɛ<sub>4</sub> ouk<sub>6</sub>, FDN le<sub>7</sub> pɛ<sub>2</sub> ko<sub>7</sub>; 排骨 FUA pɛ<sub>2</sub> ɔuk<sub>6</sub>, NND pɛ<sub>2</sub> ɔk<sub>7</sub>, SHON pɛ<sub>2</sub> kǎ<sub>7</sub> (Lin 2002: 57).

Таким образом, отчетливо видно, что по крайней мере как корневая морфема старый этимон 骨 КДК \*kwǎt сохраняется во всех ВМ диалектах. Информации о возможном замещении его в базовом значении ‘кость’ какими-то другими основами в имеющихся у нас данных также нет.

11) «**breast/грудь**»: а) {心肝頭} FQN siŋ<sub>1</sub>=ŋaŋ<sub>1</sub>=t<sup>h</sup>au<sub>2</sub>, YNT siŋ<sub>1</sub>=ŋaŋ<sub>1</sub>=t<sup>h</sup>au<sub>2</sub>, GTN siŋ<sub>1</sub>=ŋaŋ<sub>1</sub>=nau<sub>2</sub> (LN), FUA sim<sub>1</sub>=maŋ<sub>1</sub>=nau<sub>2</sub>, SHON siŋ<sub>1</sub>=ŋaŋ<sub>1</sub>=t<sup>h</sup>au<sub>2</sub>, ZHN θin<sub>1</sub>=ŋaŋ<sub>1</sub>=nau<sub>2</sub> (LN), FDN θin<sub>1</sub>=kan<sub>1</sub>=nau<sub>2</sub> (LN), а.1) {心肝} FCH θi<sub>1</sub>-ŋaŋ<sub>1</sub>, CNL siŋ<sub>1</sub>-ŋaŋ<sub>1</sub>, NND θim<sub>1</sub>-maŋ<sub>1</sub> || ПВМ {心肝頭} \*siŋ<sub>1</sub>=ŋaŋ<sub>1</sub>=nau<sub>2</sub>

Рефлексы КДК основы 胸 \*ŋoŋ в ВМ диалектах не обнаружены. Вместо нее наблюдается метонимическая инновативная форма 心肝頭, букв. ‘голова сердца и печени’, или ее сокращенный вариант 心肝, букв. ‘сердце и печень’, который, вероятно, является вторичным сокращением трехсложной формы (Starostin 2019: 166; Лоренц 2020: 134). В литературных китайских памятниках это сочетание появляется только начиная с эпохи Сун, либо в буквальном значении ‘сердце и печень’, либо в переносном ‘чуткость, истинные чувства’ (HYDC VII: 376).

КДК корень 心 \*səm ‘сердце’ регулярно реализуется в диалектах миньдун как siŋ<sub>1</sub> с «темным» ровным тоном (陰平). На стыке первой и второй морфем, восходящей к КДК

корню 肝 *\*kān* ‘печень’, происходит ассимиляция, в результате которой глухой веллярный *k-* в инициали второго корня переходит в назальный *m-*, *n-*, *ŋ-* под влиянием назальной терминали первого корня (Hiroyuki 2018: 29). (В соответствующей односложной лексеме инициаль *k-* сохраняется; см. ниже разбор лексемы ‘печень’). Аналогичная ситуация с ассимиляцией наблюдается и у последней морфемы ‘голова’ КДК 頭 *\*dʰō*, где в большинстве представленных примеров инициаль *tʰ-* развивается в *n-* под влиянием назальной терминали второго компонента.

Поскольку в большинстве диалектов представлена трехсложная, а не двусложная основа, мы предполагаем, что двусложная возникает уже на уровне отдельных диалектов как результат разговорного сокращения и предлагаем вывести на прауровень форму {心肝頭} *\*siŋ<sub>1</sub>=ŋaŋ<sub>1</sub>=nau<sub>2</sub>* (ассимилировано ← *\*siŋ<sub>1</sub>=kaŋ<sub>1</sub>=thau<sub>2</sub>*).

12) «burn/жечь»: {燒} FCH θiu<sub>1</sub>, GTN siu<sub>1</sub>, NND siu<sub>1</sub>, HUB θiəu<sub>1</sub>, XNC θiu<sub>1</sub>, JDU siu<sub>1</sub>, ZHN siu<sub>1</sub>, FDN sieu<sub>1</sub>, FQN θiu<sub>1</sub>, FUA θiu<sub>1</sub> (H), SHON siu<sub>1</sub> (H), PIN seu<sub>1</sub> || ПВМ {燒} *\*siu<sub>1</sub>*

Мы констатируем факт лексической замены КДК этимона 焚 *\*bən* → 燒 КДК *\*ŋew* уже на общеминьском уровне по причине отсутствия этимона 焚 *\*bən* во всех ветвях Минь (Starostin 2019: 167; Лоренц 2020: 134).

13) «claw (nail)/ноготь»: а) {指甲} FCH tsien<sub>1</sub>=ŋak<sub>3</sub>, GTN tsien<sub>3</sub>=kaʔ<sub>6</sub>, NND tsien<sub>3</sub>=ka<sub>6</sub> (LN), FDN tsien<sub>3</sub>=kaʔ<sub>6</sub>, FUA tsien<sub>3</sub>=kaʔ<sub>6</sub> (H), SHON tseiŋ<sub>1</sub>=kaʔ<sub>6</sub> (H), PIN tsien<sub>3</sub>=ŋak<sub>6</sub>, HUB tʃin<sub>6</sub>=ŋap<sub>7</sub>, XNC tʃin<sub>3</sub>=ŋap<sub>7</sub>, JDU tʃin<sub>6</sub>=ŋap<sub>7</sub>, а.1) FQN {手掌甲} tsʰiu<sub>3</sub>=tsien<sub>3</sub>=ŋaʔ<sub>6</sub> || ПВМ {指甲} *\*tsien<sub>3</sub>=kap<sub>6</sub>*

Этимон 爪 *\*crūʔ*, передающий искомое значение в КДК, не встречается в миньской группе; на ПВМ уровне, как и в других миньских ветвях, вместо него представлена инновативная лексема КДК 甲 *\*krāp* ‘твердая оболочка, панцирь, щит’ (Starostin 2019: 168; Лоренц 2020: 134–135). Данные показывают, что употребляется она исключительно в составе биномов, причем в большинстве случаев расширяющим элементом является морфема *\*tsien<sub>3</sub>*, которую — с некоторыми оговорками — допустимо возводить к КДК основе 指 *\*kij* ‘палец’.

Последняя имеет в ВМ диалектах и «обычный» рефлекс с нулевой терминалью, ср. собственно этимон ‘палец’: FCH 趾指 tshien<sub>3</sub> zai<sub>3</sub>, CHL tsʰien<sub>3</sub> zai<sub>3</sub>, YNT tsien<sub>3</sub> zai<sub>3</sub>, GTN tsien<sub>4</sub> zai<sub>5</sub>, FQN 手趾指 tsʰiu<sub>2</sub> zien<sub>2</sub> tsai<sub>3</sub> (Lin 2002: 59) ← ПВМ *\*tsai<sub>3</sub>*. Из-за столь значительного расхождения с морфемой в составе сложной основы ‘ноготь’ последнюю при иероглифической записи диалектных форм часто передают через иероглиф 掌 *\*tāŋ* ‘ладонь’; однако это скорее чистая условность, за которой не стоит этимологическая реальность, поскольку КДК финаль *\*-áŋ* не дает в ВМ диалектах рефлекс *-iŋ*; к тому же маловероятным представляется семантическое развитие ‘покрыв/панцирь ладони’ → ‘ноготь’ (в отличие от ‘панциря пальца’).

Учитывая, что инициаль *\*ts-* и восходящий тон, присущие первому слогу (в диалектах FCH и SHON «темный» ровный тон 陰平 мог проявиться в результате тональных сандхи), полностью совпадают с основным ВМ ‘пальцем’, можно допустить на ПВМ уровне развитие *\*tsien<sub>3</sub>=kap<sub>6</sub>* ← *\*tsij<sub>3</sub>=kap<sub>6</sub>* с нерегулярной (контекстно обусловленной?) назализацией терминали. Что касается второго компонента, то здесь надежно восстанавливается терминаль *\*-p* по данным диалектов HUB, JDU, XNC. Уже после распада ВМ общности в ряде диалектов (FCH, FQN, PIN и др.) запускается процесс ассимиляции *\*tsien<sub>3</sub>=kap<sub>6</sub>* → *\*tsien<sub>3</sub>=ŋap<sub>6</sub>*.

14) «cloud/облако»: а) {雲} GTN xun<sub>2</sub>, NND xon<sub>2</sub>, HUB xon<sub>2</sub>, XNC hon<sub>2</sub>, JDU hon<sub>2</sub>, ZHN xon<sub>1</sub>, FDN xun<sub>2</sub>, FQN hun<sub>1</sub>, SHON xun<sub>2</sub> (H), FUA houŋ<sub>2</sub> (H), PIN xɔun<sub>2</sub> (H); а.1) {雲彩} FCH un<sub>2</sub>-tsʰai<sub>4</sub> || ПВМ {雲} *\*hun<sub>2</sub>*

**15) «cold/холодный»:** а) {寒} FCH kan<sub>2</sub>, GTN kan<sub>2</sub>, NND kan<sub>2</sub>, HUB kan<sub>2</sub>, XNC kan<sub>2</sub>, JDU kan<sub>2</sub>, ZHN kan<sub>2</sub>, ZHN kan<sub>2</sub>, FDN kan<sub>2</sub>, SHON kan<sub>2</sub> (H), FUA kan<sub>2</sub> (H), PIN kan<sub>2</sub> (noen<sub>0</sub>), б) {靚~清} FCH ts<sup>h</sup>eiŋ<sub>4</sub>, GTN ts<sup>h</sup>eiŋ<sub>4</sub>, NND ts<sup>h</sup>eiŋ<sub>4</sub>, HUB t<sup>h</sup>eiŋ<sub>4</sub>, XNC ɸin<sub>4</sub>, JDU t<sup>h</sup>eiŋ<sub>4</sub>, ZHN ts<sup>h</sup>en<sub>4</sub>, FDN ts<sup>h</sup>eiŋ<sub>4</sub>, FQN ts<sup>h</sup>eiŋ<sub>5</sub>, PIN ts<sup>h</sup>eiŋ<sub>4</sub> || ПВМ {寒} \*kan<sub>2</sub> и {靚} \*ts<sup>h</sup>eiŋ<sub>4</sub>

Во всех диалектах, кроме FQN, представлен не только архаичный корень 寒 ҚДК \*gān, но и инновационный корень 靚 ҚДК \*shēŋh, четкую семантическую разницу между которыми установить не удастся. Важно отметить, что подобные синонимичные пары встречаются и в остальных ветвях диалектной группы (Лоренц 2020: 135).

Лексема 靚 \*shēŋh, согласно сведениям из словаря рифм «Цзи юнь» 集韻, была употребительной в южных регионах Чу и У, но в литературных памятниках не встречается. (Впрочем, согласно мнению Дж. Нормана, зафиксированное в литературном памятнике V в. н.э. 世說新語 «Шишо синьюй» редкое слово СК 洵 *châiŋ* ‘холодный’, представляет собой графическую разнопись 靚, см. Norman 1983: 207–208). Вероятно, эта основа представляет собой ареальную инновацию, но на ПВМ уровень она выносится вполне уверенно как старый синоним более архаичного 寒 \*kan<sub>2</sub>.

**16) «come/приходить»:** {來} FCH li<sub>2</sub>, GTN li<sub>2</sub>#, NND le<sub>2</sub>#, HUB lei<sub>2</sub>, XNC lei<sub>2</sub>, JDU lei<sub>2</sub>, ZHN le<sub>2</sub>#, FDN li<sub>2</sub>#, FQN li<sub>1</sub>, FUA lei<sub>2</sub> (H), SHON li<sub>2</sub> (H), PIN le<sub>2</sub> || ПВМ {來} \*li<sub>2</sub>

**17) «die/умирать»:** {死} FCH θi<sub>3</sub>, GTN si<sub>3</sub>, NND si<sub>3</sub>, HUB θi<sub>6</sub>, XNC θei<sub>3</sub>, JDU si<sub>6</sub>, ZHN si<sub>3</sub>, FDN si<sub>3</sub>, FQN si<sub>3</sub>, FUA θi<sub>3</sub>, SHON si<sub>3</sub>, PIN si<sub>3</sub> || ПВМ {死} \*si<sub>3</sub>

**18) «dog/собака»:** {犬} FCH k<sup>h</sup>ien<sub>1</sub>, GTN k<sup>h</sup>eiŋ<sub>4</sub>#, NND k<sup>h</sup>eiŋ<sub>3</sub>#, HUB k<sup>h</sup>en<sub>3</sub>, XNC k<sup>h</sup>en<sub>3</sub>, JDU k<sup>h</sup>en<sub>3</sub>, ZHN k<sup>h</sup>en<sub>3</sub>#, FDN k<sup>h</sup>eiŋ<sub>5</sub>#, FQN k<sup>h</sup>eiŋ<sub>3</sub>, CNL k<sup>h</sup>eiŋ<sub>3</sub>#, YNT k<sup>h</sup>eiŋ<sub>3</sub>#, FUA k<sup>h</sup>eiŋ<sub>4</sub>#, SHON k<sup>h</sup>eiŋ<sub>3</sub>#, PIN k<sup>h</sup>eiŋ<sub>3</sub> || ПВМ {犬} \*k<sup>h</sup>eiŋ<sub>3</sub>

Родовой термин для передачи значения ‘собака’ 犬 ҚДК \*k<sup>hw</sup>it<sub>1</sub>, характерный только для ҚДК и РДК, употребляется во всех диалектах миньдун и является исключительной особенностью данной ветви, так как в остальных ветвях используется инновативная лексема 狗 ҚДК \*k<sup>o</sup>, возможно, исходно заимствованная из южных языков для обозначения видов собак, применимых в гастрономии (Starostin 2019: 169–170; Лоренц 2020: 136). Следует добавить, что на предполагаемый момент распада праминьской общности (III–IV в. н.э.) слово 狗 \*k<sup>o</sup> в литературном китайском языке уже вытеснило из употребления старую основу 犬 \*k<sup>hw</sup>it<sub>1</sub>.

**19) «drink/пить»:** а) {食} FCH sie<sub>7</sub>, GTN sia<sub>7</sub>, NND sia<sub>7</sub>, ZHN sie<sub>7</sub>, HUB θie<sub>7</sub><sup>9</sup>, XNC ɸie<sub>7</sub>, JDU ɸie<sub>7</sub>, FDN sia<sub>7</sub>, FQN sia<sub>7</sub>, CNL sie<sub>7</sub>, YNT sie<sub>5</sub>, FUA sie<sub>7</sub>, SHON sia<sub>7</sub>, PIN se<sub>7</sub>?; б) {啜} FCH ts<sup>h</sup>yok<sub>6</sub> || ПВМ {食} \*sia<sub>7</sub>

ҚДК этимон 飲 \*ʔit ‘пить’ в требуемом значении не проявляется в ВМ диалектах и, очевидно, был рано вытеснен полисемичным корнем 食 ҚДК \*lak со значениями ‘есть’, ‘пить’, ‘курить’. Подобное единение значений ‘есть’ и ‘пить’ является общеминьской инновацией и, вероятно, произошло под влиянием соседних чжуан-тайских языков, где эти значения выражаются одной и той же лексемой (Starostin 2019: 170; Лоренц 2020: 136). При этом, однако, старая лексема 飲 во всех миньских диалектах кроме центральной вет-

<sup>8</sup> Данная лексема впервые появляется в письменных текстах в V–III вв. до н. э. и замещает старую основу 犬 уже в начале эпохи Хань (III в. до н.э. — II в. н.э.; Starostin 2019: 169–170).

<sup>9</sup> Акитани Хироюки (Hiroyuki 2018, 2020) вместо 7 тонов выделяет во всех восточноминьских диалектах 8 из-за расщепления «светлого» падающего тона (陽去) на 陽去a и 陽去b в результате тональных сандхи. Из соображений экономности мы в данной работе оставляем систему из 7 тонов.

ви развивает именное значение 米湯 ‘рисовый отвар, рисовый суп’, например: FCH  $a\eta_3$ , GTN  $a\eta_3$ , сямэньский  $am_3$ , цзяньоуский  $ai\eta_3$  и др. (Hiroyuki 2018: 538; Cheng & Li 1991: 25).

Накадзима в качестве основного глагола ‘пить’ для фучжоуского диалекта указывает лексему 啜  $ts^h\eta_6$  с дополнительным значением ‘отхлебнуть; потягивать’, восходящую к ханьскому корню  $*t^w\alpha t$  (записывается также как 𪗇) ‘пить, есть, отхлебнуть, потягивать’ (Nakajima 1979: 141; Schuessler 2007: 198). Ли и Лян указывают этот корень в качестве синонимичной основы исключительно для значения ‘пить’, таже отмечая дополнительное значение ‘пить большими глотками’ (Li & Liang 1994: 55, 405). Эта основа также широко представлена в южной, центральной, цюньвэньской, шао-цзянской ветвях (Лоренц 2020: 136); тем не менее, на ВМ уровне она выглядит как очевидная семантическая инновация.

**20) «dry/сухой»:** {焦~ 焦~ 涸} FCH  $ta_1$ , GTN  $ta_1$ , NND  $ta_1$ , HUB  $ta_1$ , XNC  $ta_1$ , JDU  $ta_1$ , ZHN  $ta_1$ , FDN  $ta_1$ , FQN  $ta_1$ , CNL  $ta_1$ , FUA  $ta_1$ , YNT  $ta_1$ , SHON  $ta_1$  || ПВМ {焦}  $*ta_1$

КДК этимон 乾  $*g^h\alpha n$ , по всей видимости, был вытеснен инновативным корнем  $*ta_1$  уже на ПМ уровне, так как он не представлен ни в одном диалекте изучаемой диалектной группы (Starostin 2019: 162; Лоренц 2020: 137). При записи данной лексемы чаще всего используются иероглифические знаки 焦 или 焦, которые обычно передают семантику ‘горелый’, при этом непосредственно слово ‘горелый’, восходящее к другому корню, также записывается этим знаком: HUB  $tʃi\alpha_1$ , JDU  $tʃiv\alpha_1$  (Hiroyuki 2018: 306).

А. Шюсслер предполагает, что миньская лексема ‘сухой’ является заимствованием и этимологически связана с правьет-мыонгским корнем  $*traw$  с тем же значением (Schuessler 2007: 308); убедительных этимологических гипотез, в рамках которых слово могло бы объясняться как исконно китайское, нет.

**21) «ear/ухо»:** а) {耳} FCH  $\eta i_6$ , GTN  $\eta i_6$  (LH), NND  $\eta e i_6$  (LH), HUB  $\eta e i_3$ , XNC  $\eta e i_5$ , JDU  $\eta e i_3$ , ZHN  $\eta e_6$  (LH), FDN  $\eta i_5$ , CNL  $\eta e i_6$ , а.1) {耳} FQN  $\eta i_4$ -(k) $\eta i\alpha_2$ , YNT  $\eta i_1$ - $\eta i\alpha_3$ , PIN  $\eta e_5$ - $\eta i\alpha_3$ , а.2) {耳} FUA  $\eta i_6$ - $l\alpha_3$ , SHON  $\eta i_6$ - $l\alpha_3$  || ПВМ {耳}  $*\eta i_6$

КДК этимон 耳  $*n^h\acute{\alpha}$  уверенно сохраняется во всей ВМ ветви, но в некоторых случаях расширяется до статуса бинома. Так, в диалекте FQN вторым компонентом является  $(k)\eta i\alpha_2 \leftarrow$  СК 囡  $*k\acute{\epsilon}n$  ‘ребенок’ (см. комментарий к ‘птице’). В двух других диалектах FUA и SHON употребляется морфема СК 梛  $l\acute{\alpha}w$  ‘корзина’, которая в литературных памятниках встречается только начиная с эпохи Тан.

Следует добавить, что у данного слова во всех диалектах есть литературное чтение  $\eta i_3$  с восходящим тоном (上聲) (Liu & He 1998: 489–490). Исходя из этих данных, мы предполагаем, что литературные чтения в диалектах GTN, NND, ZHN по источнику Chen & Li 1991 указаны вместо разговорных, и исправляем по источнику Liu & He 1998. Весьма вероятно также, что и по диалектам HUB, JDU указаны литературные чтения вместо разговорных.

**22) «earth/земля»:** {塗~ 塗~ 土} FCH  $t^h\alpha_2$ , GTN  $t^h\alpha_2$ , NND  $t^h\alpha_2$ , HUB  $t^h\alpha_2$ , XNC  $t^h\alpha_2$ , JDU  $t^h\alpha_2$ , ZHN  $t^h\alpha_2$ , FDN  $t^h\alpha_2$ , FQN  $t^h\alpha_2$ , CHL  $t^h\alpha_2$ , YNT  $t^h\alpha_2$ , FUA  $t^h\alpha_2$ , SHON  $t^h\alpha_2$ , PIN  $t^h\alpha_2$  || ПВМ {塗}  $*t^h\alpha_2$

Здесь налицо лексическая замена: КДК 土  $*t^h\acute{\alpha} \rightarrow$  塗 (КДК  $*L^h\bar{\alpha}$ ) (а) ‘глина’; (б) ‘(невысокая?) дорога’ как на ПМ уровне, так и на общеминьском уровне см. (Starostin 2019: 172; Лоренц 2020: 137–138), так как во всех примерах присутствует глухая придыхательная инициаль в сочетании со вторым тоном, что указывает на наличие звонкой придыхательной инициали на ПМ уровне. В южных и восточных диалектах эта лексема обычно передает значение ‘глиняная земля’.

В качестве основного иероглифического знака для записи этого слова используется КДК 塗  $*L\bar{\alpha}$  ‘грязь, глина; дорога’, а также иногда встречается знак КДК 土  $*t^h\acute{\alpha}$  ‘земля’.

**23) «eat/есть»:** {食} FCH sieʔ<sub>7</sub>, GTN siaʔ<sub>7</sub>, NND siaʔ<sub>7</sub>, HUB θieʔ<sub>7</sub>, XNC ɕieʔ<sub>7</sub>, JDU ɕieʔ<sub>7</sub>, ZHN sieʔ<sub>7</sub>, FDN siaʔ<sub>7</sub>, FQN sia<sub>7</sub>, CNL sieʔ<sub>7</sub>, YNT sieʔ<sub>5</sub>, FUA siek<sub>7</sub>, SHON siaʔ<sub>7</sub>, PIN seeʔ<sub>7</sub> || ПВМ {食} \*siak<sub>7</sub>  
См. лексический разбор выше ('пить').

**24) «egg/яйцо»:** {卵} FCH louŋ<sub>5</sub>, GTN louŋ<sub>5</sub>, NND louŋ<sub>5</sub>, HUB louŋ<sub>5</sub>, XNC louŋ<sub>5</sub>, JDU louŋ<sub>5</sub>, ZHN louŋ<sub>5</sub>, FDN louŋ<sub>5</sub>, FQN louŋ<sub>5</sub>, CHL louŋ<sub>5</sub>, YNT louŋ<sub>5</sub>, FUA louŋ<sub>5</sub>, SHON louŋ<sub>5</sub>, PIN louŋ<sub>5</sub> || ПВМ {卵} \*lon<sub>5</sub>

**25) «eye/глаз»:** {目珠} FCH møyk<sub>7</sub>-tsiu<sub>1</sub>, GTN meiʔ<sub>6</sub>-tsiu<sub>1</sub>, NND muʔ<sub>7</sub>-tsiu<sub>1</sub>, HUB mit<sub>7</sub>-tʃiu<sub>1</sub>, XNC muk<sub>7</sub>-tʃiu<sub>1</sub>, JDU mit<sub>7</sub>-tʃiu<sub>1</sub> ~ muk<sub>7</sub>-tʃiu<sub>1</sub>, ZHN miʔ<sub>7</sub>-tsiu<sub>1</sub>, FDN muʔ<sub>7</sub>-tsiu<sub>1</sub>, FQN møyʔ<sub>7</sub>-tsiu<sub>3</sub>, CNL meiʔ<sub>7</sub>-tsiu<sub>1</sub>, YNT meiʔ<sub>7</sub>-tsiu<sub>1</sub>, FUA mik<sub>7</sub>-ʒiu<sub>1</sub>, SHON muʔ<sub>1</sub>-tsiu<sub>1</sub>, PIN muk<sub>7</sub>-tsiu<sub>1</sub> || ПВМ {目珠} \*muk<sub>7</sub>-tsiu<sub>1</sub>

КДК этимон 目 \*m<sup>h</sup>uk 'глаз' стабильно сохраняется во всех миньдунских диалектах, но в расширенной до биннома форме (с помощью морфемы КДК 珠 \*to 'жемчужина').

Необычной особенностью представленной основы является нерегулярное развитие КДК гласного \*и в гласные верхнего подъема -i, -e в большинстве ВМ диалектов. А. Хироюки отмечает, что подобное нерегулярное развитие наблюдается и в других словах, содержащих морфему 'глаз'. Он распределяет все рефлексy на 4 группы и приходит к выводу, что формы, которые имеют финали -iʔ, -ui, -i, -ik, не восходят к праминьскому; при этом, однако, для того, чтобы понять их происхождение, данных явно недостаточно (Hiroyuki 2020: 811–816). Согласие с таким выводом предполагает неясное и неэкономное решение, требующее либо заимствований из неизвестных субстратов, либо постулирование многочисленных лишних сущностей с семантикой 'глаз' на ПВМ уровне. Мы все же предпочитаем считать все перечисленные формы исконно родственными, в предварительном порядке объясняя нерегулярные варьирования в финалях ассимилятивными процессами. Стоит отметить, что в ВМ диалектах сама лексема КДК 珠 \*to 'жемчужина' имеет две формы: 1) с финалью io ~ и в автономном употреблении (Hiroyuki 2018: 512), 2) с финалью iu в слове 'глаз'.

**26) «fat/жир»:** а) {肥} FCH pui<sub>2</sub>, GTN pui<sub>2</sub>, NND pui<sub>2</sub>, HUB poi<sub>2</sub>, XNC poi<sub>2</sub>, JDU poi<sub>2</sub>, ZHN pui<sub>2</sub>, FDN pui<sub>2</sub>, FQN poi<sub>2</sub>, CNL pui<sub>2</sub>, YNT poi<sub>2</sub>, FUA pøi<sub>2</sub>, SHON pui<sub>2</sub>, PIN pui<sub>2</sub>, б) {膏} FCH ko<sub>1</sub> || ПВМ {肥} \*pui<sub>2</sub>

Видно, что КДК этимон 膏 \*kāw из всех диалектов отражается только в FCH; это заставляет предположить, что он был вытеснен на ПВМ уровне лексемой КДК 肥 \*baj с аналогичным значением (Starostin 2019: 164). Хироюки отмечает, что эта лексема может употребляться как 'животный жир', так и 'толстый' по отношению к человеку (Hiroyuki 2018: 358).

Мы предполагаем, что в фучжоуском диалекте этимон 膏 ko<sub>1</sub> был восстановлен под влиянием литературного языка в качестве синонима к базовой лексеме 肥 \*pui<sub>2</sub>.

**27) «feather/перо»:** а) {毛} NND mo<sub>2</sub># (LN), HUB mo<sub>2</sub>, XNC mo<sub>2</sub>, JDU mo<sub>2</sub>, ZHN mo<sub>2</sub># (LN), FDN mo<sub>2</sub># (LN), FQN mo<sub>2</sub># (LN), CNL mo<sub>2</sub># (LN), YNT mo<sub>2</sub># (LN), FUA mo<sub>2</sub># (LN), SHON mo<sub>2</sub># (LN), PIN mo<sub>2</sub>; а.1) FCH y<sub>3</sub>-mou<sub>2</sub> {羽毛} || ПВМ {毛} \*mo<sub>2</sub>

Искомое слово по основным источникам ВМ лексики в автономном употреблении зафиксировано в ограниченном числе диалектов FCH, HUB, XNC, JDU, PIN. Данные для остальных диалектов взяты из многосложной лексемы 鷄毛拂 'метелка из перьев петуха' (Lin 2002: 41); в полностью автономном употреблении в этих диалектах слово не зафиксировано.

КДК этимон 羽 \*w(r)á самостоятельно не используется, но сочетается с полисемичным корнем КДК 毛 \*m<sup>h</sup>āw 'шерсть, перо, волосы на теле' в рамках биннома 羽毛 и исключительно в диалекте FCH (Лоренц 2020: 138). В остальных диалектах при передачи данного значения употребляется только однослог КДК 毛 \*m<sup>h</sup>āw. Как отмечает А. Хироюки,



лексема КДК 毛 *\*m<sup>h</sup>āw* сохраняет свою полисемичность и в миньских диалектах (Hiroyuki 2018: 306, 508). Следует добавить, что она также отдельно употребляется в *цюньвэньской* и *южной* ветвях.

**28) «fire/огонь»:** {火} FCH hui<sub>3</sub>, GTN хuoi<sub>4</sub>#, NND хəy<sub>3</sub>#, HUB hu<sub>1</sub>i<sub>3</sub>, XNC həy<sub>3</sub>, JDU həy<sub>3</sub>, ZHN huai<sub>3</sub>#, FDN хuoi<sub>3</sub>#, FQN hui<sub>3</sub>#, CHL хui<sub>2</sub>#, YNT uoi<sub>3</sub>#, FUA hui<sub>4</sub>#, SHON хuoi<sub>4</sub>#, PIN хoi<sub>3</sub> || ПВМ {火} \*hui<sub>3</sub>

**29) «fish/рыба»:** {魚} FCH ηy<sub>2</sub>, GTN ηy<sub>2</sub>, NND ηəy<sub>2</sub>, HUB ηli<sub>2</sub>, XNC ηə<sub>2</sub>, JDU ηəy<sub>2</sub>, ZHN ηəu<sub>2</sub>, FDN ηi<sub>2</sub>, FQN ηi<sub>2</sub>, CNL ηy<sub>2</sub>, YNT ηy<sub>2</sub>, FUA ηəi<sub>2</sub>, SHON ηy<sub>2</sub>, PIN ηə<sub>2</sub> || ПВМ {魚} \*ηi<sub>2</sub>

**30) «fly/летать»:** {飛} FCH pui<sub>1</sub>, GTN puoi<sub>1</sub>, NND pəy<sub>1</sub>, HUB puoi<sub>1</sub>, XNC pui<sub>1</sub>, JDU pəy<sub>1</sub>, ZHN pui<sub>1</sub>, FDN puei<sub>1</sub>, FQN puoi<sub>1</sub>, CNL pui<sub>1</sub>, YNT pui<sub>1</sub>, FUA puoi<sub>1</sub>, SHON puoi<sub>1</sub>, PIN poi<sub>1</sub> || ПВМ {飛} \*pui<sub>1</sub>

**31) «foot/нога»:** {跣~骹} FCH k<sup>h</sup>a<sub>1</sub>, GTN k<sup>h</sup>a<sub>1</sub>, NND k<sup>h</sup>a<sub>1</sub>, HUB k<sup>h</sup>a<sub>1</sub>, XNC k<sup>h</sup>a<sub>1</sub>, JDU k<sup>h</sup>a<sub>1</sub>, ZHN k<sup>h</sup>a<sub>1</sub>, FDN k<sup>h</sup>a<sub>1</sub>, FQN k<sup>h</sup>a<sub>1</sub>, CNL k<sup>h</sup>a<sub>1</sub>, YNT k<sup>h</sup>a<sub>1</sub>, FUA k<sup>h</sup>a<sub>1</sub>, SHON k<sup>h</sup>a<sub>1</sub> || ПВМ {跣} \*k<sup>h</sup>a<sub>1</sub>

КДК этимон 足 *\*cok* был вытеснен инновативным корнем 骹 КДК *\*k<sup>h</sup>rāw* уже на общеминьском уровне (Starostin 2019: 171; Лоренц 2020: 139). Основа 骹 толкуется в словаре «Шовэнь цзецзы» как ‘голень’, несколько раз встречаясь в текстах классического периода.

**32) «full/полный»:** а) {滿~瞞} FCH muan<sub>3</sub>, SHON muan<sub>3</sub> (H), FUA muan<sub>3</sub> (H), PIN muan<sub>3</sub>, FQN muan<sub>3</sub>, HUB man<sub>3</sub>, XNC man<sub>3</sub>, JDU man<sub>3</sub>; б) {溘} FCH tien<sub>5</sub>, HUB tiem<sub>5</sub>, XNC tim<sub>5</sub>, JDU tim<sub>5</sub> || ПВМ {滿} \*muan<sub>3</sub>

Широкоупотребимый КДК корень 盈 *\*leŋ*, очевидно, был вытеснен на общеминьском уровне, так как не отмечен ни в одной из ветвей группы Минь (Starostin 2019: 170; Лоренц 2020: 139).

В данных по ВМ встречаются два синонима: 滿 (КДК *\*mān*) и 溘 (СК *tien*). Инновативный корень 滿 *\*mān* начинает встречаться в литературных памятниках начиная с ханьского периода. Альтернативная лексема СК 溘 *tien*, очевидно, не имеет ни прямой связи с иероглифическим знаком, которым записывается, ни параллелей в КДК текстах; иероглиф подобран исключительно из-за фонетической схожести с лексемой. Мы предполагаем, что данное слово может быть связано с чжуан-тайскими формами, см. тайск. *tem*, лунчжоу *tim*, по-ай *lim* и др. (*\*tlim*.А согласно Li 1977: 118). Следует добавить, что этот корень также зафиксирован в *южной* и *цюньвэньской* ветвях (Лоренц 2020: 139).

На ПВМ уровень из двух кандидатов уверенно выводится только 滿 *\*muan<sub>3</sub>*, поскольку, в отличие от СК 溘 *tiem<sub>5</sub>*, он имеет широкую дистрибуцию и однозначно опознается как исконно синитическая лексема.

**33) «give/давать»:** а) {乞} FCH k<sup>h</sup>əy<sub>6</sub>, GTN k<sup>h</sup>ei<sub>7</sub>#, NND k<sup>h</sup>i<sub>7</sub>#, HUB k<sup>h</sup>i<sub>6</sub>#, XNC k<sup>h</sup>it<sub>6</sub>#, JDU k<sup>h</sup>i<sub>6</sub>#, ZHN k<sup>h</sup>e<sub>7</sub>#, FDN k<sup>h</sup>i<sub>7</sub>#, FQN k<sup>h</sup>ə<sub>6</sub>, CHL k<sup>h</sup>əy<sub>6</sub>, YNT k<sup>h</sup>y<sub>7</sub>#, SHON k<sup>h</sup>y<sub>7</sub>#, б) {錢} FUA tsin<sub>2</sub>#, с) PIN k<sup>h</sup>uai<sub>4</sub># || ПВМ {乞} \*k<sup>h</sup>yt<sub>6</sub>

Базовый для КДК текстов этимон 與 *\*lá* не сохраняется ни в одном из современных миньских диалектов (Starostin 2019: 168; Лоренц 2020: 139-140). Обладающий наиболее широкой дистрибуцией корень 乞 (КДК *\*k<sup>h</sup>ət*; на КДК уровне значение — ‘просить’), по всей видимости, унаследован от КДК эпохи (HYDC I: 760). В классических текстах, начиная с эпохи Хань, встречается производная от него форма *\*k<sup>h</sup>ət-s* (→ КДК *\*k<sup>h</sup>ət*) ‘давать’ (Schuessler 2007: 422), однако миньские формы с этим же значением восходят непосредственно к *\*k<sup>h</sup>ət*, что заставляет предположить, что на ПМ уровне значение производного глагола ‘давать’ перешло на основной глагол ‘просить’, и ПВМ наследует эту особенность. Ср. потенциальные сино-тибетские когнаты, приводимые А. Шюсслером (Schuessler 2007: 422): лимбу *kett* ‘передавать, доставлять’, канаури *ket*, тхебор *k<sup>h</sup>et* ‘давать’ — если эти

слова действительно этимологически родственны китайским, мы, возможно, имеем здесь дело с аналогичным (независимым) семантическим сдвигом.

Этимология локальных инноваций b) и c) остается неясной.

**34) «good/хороший»:** {好} FCH hо<sub>3</sub>, GTN х<sub>3</sub>, NND х<sub>3</sub>, HUB х<sub>5</sub>, XNC h<sub>5</sub>, JDU h<sub>5</sub>, ZHN х<sub>3</sub>, FDN х<sub>3</sub>, FQN hо<sub>3</sub>, SHON х<sub>3</sub> (H), FUA h<sub>3</sub> (H), PIN х<sub>3</sub> || ПВМ {好} \*hо<sub>3</sub>

**35) «green/зеленый»:** a) {綠} FCH luok<sub>7</sub>, HUB luo<sub>7</sub>, XNC lu<sub>7</sub>, JDU lu<sub>7</sub>, FQN luo<sub>7</sub>, SHON lyø<sub>7</sub> (H), FUA lu<sub>7</sub> (H), PIN lo<sub>1</sub>; b) {青} FCH ts<sup>h</sup>an<sub>1</sub>, HUB ts<sup>h</sup>an<sub>1</sub>, XNC t<sup>h</sup>an<sub>1</sub>, JDU ts<sup>h</sup>an<sub>1</sub>, FUA ts<sup>h</sup>an<sub>1</sub> (H), SHON ts<sup>h</sup>an<sub>1</sub> (H), PIN ts<sup>h</sup>an<sub>1</sub> || ПВМ {綠} \*luok<sub>7</sub> и {青} \*ts<sup>h</sup>an<sub>1</sub>

В миньдунской ветви используются два корня со значением ‘зеленый’: более архаичная лексема КДК 青 \*s<sup>h</sup>ēŋ, обозначающая широкий спектр холодных оттенков от зеленого, синего, голубого до серого и черного, и более инновативная лексема КДК 綠 \*rok, передающая исключительно значение ‘зеленый’, но при этом в классических текстах применяемая, как правило, для описания искусственно окрашенных предметов.

Ли Жулун и Лян Юйчжан приписывают лексеме 綠 luok<sub>7</sub> в фучжоуском диалекте основное значение ‘позеленевшее лицо’, а непосредственно значение ‘зеленый’ указывает только для 青 ts<sup>h</sup>an<sub>1</sub>, однако остальные источники эту информацию не подтверждают (Li & Liang 1994: 37, 193). При этом диффузная семантика у слова 青 \*ts<sup>h</sup>an<sub>1</sub> в миньдунских диалектах также сохраняется, поэтому лексема может использоваться для передачи других оттенков холодного спектра, например, для черного (Li & Liang 1994: 37). Любопытно, что в лексеме ‘маш, зеленая фасоль’ 綠豆 используется именно лексема КДК 綠 \*rok (Hiroyuki 2018: 72, 164; Nakajima 1979: 33). Тем не менее, поскольку четкой и убедительной информации о семантических различиях между этими двумя основами источники, как правило, не предоставляют, приходится выводить обе основы (\*luok<sub>7</sub> и \*ts<sup>h</sup>an<sub>1</sub>) на прауровень в качестве «технических» синонимов.

**36) «hair/волосы»:** a) {頭髮 ~ 頭發} FCH t<sup>h</sup>au<sub>2</sub>=uok<sub>6</sub>, GTN t<sup>h</sup>au<sub>2</sub>=muo<sub>6</sub>, NND t<sup>h</sup>au<sub>2</sub>=u<sub>6</sub>, HUB t<sup>h</sup>au<sub>2</sub>=uot<sub>6</sub>, XNC t<sup>h</sup>a<sub>2</sub>=ut<sub>6</sub>, JDU t<sup>h</sup>a<sub>2</sub>=ut<sub>6</sub>, ZHN t<sup>h</sup>au<sub>2</sub>=u<sub>6</sub>, FDN t<sup>h</sup>au<sub>2</sub>=puo<sub>6</sub>, FQN t<sup>h</sup>au<sub>2</sub>=huo<sub>6</sub>, FUA t<sup>h</sup>a<sub>2</sub>=wu<sub>6</sub> (H), SHON t<sup>h</sup>au<sub>2</sub>=puo<sub>6</sub> (H), PIN t<sup>h</sup>an<sub>2</sub>=mok<sub>6</sub>; b) {頭毛} FQN t<sup>h</sup>au<sub>2</sub>=mo<sub>2</sub> || ПВМ {頭髮} \*t<sup>h</sup>au<sub>2</sub>=put<sub>6</sub>

КДК этимон 髮 \*pat ‘волосы’ стабильно сохраняется во всех миньдунских диалектах, но только в составе композита 頭髮, где первый компонент — КДК 頭 \*d<sup>h</sup>ō ‘голова’. Иногда вместо знака 髮 \*pat используется омонимичный знак 發 \*pat (‘начинать, выходить’); причины такой замены остаются неясными.

В диалекте фуцин, кроме вышеуказанного слова, также употребляется его полный синоним {頭毛} t<sup>h</sup>au<sub>2</sub>=mo<sub>2</sub> (букв. ‘шерсть головы’), который широко распространен в других ветвях группы Минь. Возможно, он появился под ареальным влиянием соседнего диалекта путянь, в котором используется исключительно вариант ‘шерсть головы’ (Лоренц 2020: 140).

На прауровень надежно выводится бином 頭髮 \*t<sup>h</sup>au<sub>2</sub>=put<sub>6</sub>, на следующих основаниях: 1) первый компонент ‘голова’ стабильно сохраняется во всех диалектах, без отклонений в фонетических соответствиях; 2) второй компонент ‘волосы’ в большинстве диалектов хотя и утрачивает инициаль на стыке слов, как это обычно происходит в миньдунских диалектах, но сохранение инициали в диалектах FCH, FDN, FQN доказывает ее выводимость на ПВМ уровень; 3) терминаль \*-t восстанавливается по диалектным данным А. Хиroyuki, в которых она регулярно сохраняется (см. аналогичные примеры ‘кровь’, ‘кость’).

Особо выделяются варианты в диалектах GTN и PIN, где произошел нерегулярный переход инициали второго слога p → m.

37) «**hand/рука**»: {手} FCH tshiu<sub>3</sub>, GTN tshiu<sub>3</sub>, NND tshiu<sub>3</sub>, HUB tshiu<sub>3</sub>, XNC tshou<sub>3</sub>, JDU çiu<sub>3</sub>, ZHN tshiu<sub>3</sub>, FDN tshiu<sub>3</sub>, FQN tshiu<sub>3</sub>, CNL tshiu<sub>3</sub>, YNT tshiu<sub>3</sub>, FUA tshieu<sub>3</sub>, SHON tshiu<sub>3</sub>, PIN tshiu<sub>3</sub> || ПБМ {手}\*tshiu<sub>3</sub>

38) «**head/голова**»: {頭} FCH thau<sub>2</sub>, GTN thau<sub>2</sub>, NND thau<sub>2</sub>, HUB thau<sub>2</sub>, XNC thau<sub>2</sub>, JDU thau<sub>2</sub>, ZHN thau<sub>2</sub>, FDN thau<sub>2</sub>, FQN thau<sub>2</sub> FUA thau<sub>2</sub>, SHON thau<sub>2</sub>, PIN thau<sub>2</sub> || ПБМ {頭} \*thau<sub>2</sub>

39) «**hear/слышать**»: {聽~聽} FCH thian<sub>1</sub>, FQN thian<sub>1</sub>, HUB thien<sub>1</sub>, XNC thian<sub>1</sub>, JDU thian<sub>1</sub>, FUA thien<sub>1</sub> (H), SHON thian<sub>1</sub> (H), PIN thian<sub>1</sub> || ПБМ {聽} \*thian<sub>1</sub>

Старый КДК этимон 聞 \*mən не сохранился в восточноминьских диалектах и был вытеснен лексемой КДК 聽 \*lēŋ еще на ПМ уровне, так как в остальных ветвях группы Минь он тоже не зафиксирован (Starostin 2019: 169; Лоренц 2020: 141).

40) «**heart/сердце**»: {心} FCH θin<sub>1</sub>, GTN sin<sub>1</sub>#, NND θim<sub>1</sub>#, HUB θim<sub>1</sub>, XNC θim<sub>1</sub>, JDU sim<sub>1</sub>, ZHN θin<sub>1</sub>#, FDN θin<sub>1</sub>#, FQN sin<sub>1</sub>, CNL sin<sub>1</sub>#, YNT sin<sub>1</sub>#, FUA sim<sub>1</sub>#<sup>10</sup>, SHON sin<sub>1</sub>#, PIN sin<sub>1</sub> || ПБМ {心} \*sim<sub>1</sub>

41) «**horn/рог**»: {角} FCH koyk<sub>6</sub>, GTN køy<sub>6</sub>, NND kœk<sub>6</sub>, HUB kœk<sub>6</sub>, XNC kœk<sub>6</sub>, JDU kœk<sub>6</sub>, ZHN kœuk<sub>6</sub>, FDN ke<sub>6</sub>, FQN kœ<sub>6</sub> (H), FUA kœ<sub>6</sub> (H), SHON kœ<sub>6</sub> (H), PIN kœ<sub>6</sub> || ПБМ {角} \*koyk<sub>6</sub> {角}

42) «**I/я**»: {我} FCH ŋuai<sub>3</sub>, GTN ŋuai<sub>3</sub>, NND ua<sub>5</sub>, HUB ua<sub>5</sub>, XNC ua<sub>5</sub>, JDU ua<sub>5</sub>, ZHN ua<sub>3</sub>, FDN ua<sub>3</sub>, FQN ŋua<sub>3</sub>, CHL ŋui<sub>3</sub>, YNT ŋuo<sub>3</sub>, FUA ŋo<sub>3</sub>, SHON ŋua<sub>3</sub>, PIN uai<sub>3</sub> || ПБМ {我} \*ŋuai<sub>3</sub>

43) «**kill/убивать**»: {劊 ~ 治 ~ 夷} FCH thai<sub>2</sub><sup>11</sup>, GTN thai<sub>2</sub>, NND thai<sub>2</sub>, HUB thai<sub>2</sub>, XNC thai<sub>2</sub>, JDU thai<sub>2</sub>, ZHN thai<sub>2</sub>, FDN thai<sub>2</sub>, FQN thai<sub>2</sub>, CHL thai<sub>2</sub>, YNT thai<sub>2</sub>, FUA thai<sub>2</sub>, SHON thai<sub>2</sub> || ПБМ {劊} \*thai<sub>2</sub>

ВМ основа \*thai<sub>2</sub> унаследована от общеминьского состояния; в ПМ эта основа вытеснила КДК этимон 殺 \*srāt (Starostin 2019: 163; Лоренц 2020: 141–142).

Было предложено две возможных этимологии инновативного корня. Одна из них выдвинута Дж. Норманом, который считает миньский корень производным от КДК лексемы 治 \*L<sup>h</sup>əh ‘управлять, регулировать’ с производными значениями ‘чистить, очищать, убивать, зарезать’ (Norman 1979: 179–181). Эта гипотеза сомнительна ввиду необычности семантической деривации ‘регулировать’ → ‘зарезать, убивать’ и нерегулярности развития КДК финали \*-əh (с падающим тоном) → -ai (с ровным тоном!) в миньских диалектах. Иероглифические знаки, использующиеся для записи данного слова, вероятно, были подобраны исходя из фонетического сходства.

Более убедительной выглядит другая этимология, выдвинутая Дэн Сяохуа, по которой миньская лексема ‘убивать’ была заимствована из тай-кадайских языков в общеминьский период, см. ‘умереть’: чжуанск. *tāi<sub>1</sub>*, дай *tāi<sub>1</sub>*, линьгао *dai<sub>1</sub>*, дунск. *tāi<sub>1</sub>*, мулао *tai<sub>1</sub>*, суй *tai<sub>1</sub>*, ли *tāu<sub>1</sub>*. Он также приводит возможные параллели в языках мяо-яо: цяньдун мяо *ta<sub>5</sub>* ‘убить’, *tua<sub>6</sub>* ‘умереть’, чуань-цянь дяньмяо *tua<sub>5</sub>* ‘убить’, *tua<sub>6</sub>* ‘умереть’, меник (яо) *tai<sub>5</sub>* ‘убить’, *tai<sub>6</sub>* ‘умереть’, бяоминь яо *tai<sub>5</sub>* ‘убить’, *tai<sub>4</sub>* ‘смерть, умереть’, и в австронезийских языках: амис *pataj* ‘убить’, бунун *matað* ‘умереть’, пайвань *matsay* ‘умереть’, цатский *tai<sub>42</sub>* ‘умереть’ и др. (Deng 1994: 39), хотя непосредственная связь их с миньскими формами сомнительна (скорее они могут отражать древние ареальные связи в австрическом регионе или даже восходить к эпохе общеевстрического единства).

Внешняя этимология Дэн Сяохуа выглядит убедительнее, чем внутренняя Дж. Нормана, однако странным остается факт заимствования непереходного глагола ‘умирать’

<sup>10</sup> Акитани Хироюки указывает форму *θeiŋ<sub>1</sub>* для диалекта фуань (Hiroyuki 2020: 625).

<sup>11</sup> М. Накадзима указывает для данного слова исключительно значение ‘разделять’, при этом все прочие источники упоминают значение ‘убивать’ (Nakajima 1979: 23).

как переходно-каузативного ‘убивать’ при отсутствии соответствующих форм в тайских источниках. Тем не менее, очевидно, что других кандидатов на статус пра-ВМ этимона ‘убивать’ нет.

**44) «кнее/колени»:** а) {骸腹頭 ~ 骸肚頭} GTN  $k^{h}a_1-u_4-lau_2$ , ZHN  $k^{h}a_1-lu_3-lau_2$ , FQN  $k^{h}a_3-tu_3-t^{h}au_1$ ; а.1) {骸肚頭} FCH  $k^{h}a_1-βu_4-t^{h}au_2$  (LL), FDN  $k^{h}a_1-pu_6-lau_2$ ; а.2) NND  $k^{h}a_1-ɔt_6-t^{h}au_2$  {骸骨頭}, XNC  $k^{h}a_1-ut_6-t^{h}au_2$  {骸窟頭}; а.2.1) JDU {窟頭}  $k^{h}ot_6-t^{h}au_2$ ; а.3) HUB {骸頭}  $k^{h}a_1-lau_1$ ; а.4) {骸頭節} PIN  $k^{h}a_1-lau_2$ <sup>12</sup>-ZEIK<sub>6</sub> || ПВМ {骸肚頭}  $*k^{h}a_1-lu_3-lau_2 \sim *k^{h}a_1-lu_3-t^{h}au_2$

ҚДК этимон 膝 *\*sit* ‘колени’, следы которого отсутствуют в современных миньских диалектах, был вытеснен трехморфемным композитом еще на общеминьском уровне (Starostin 2019: 165; Лоренц 2020: 142). Инновативная лексема 骸 (ҚДК *\*k<sup>h</sup>āw*) ‘голень’ в большинстве миньдунских диалектов составляет основу метафоры, буквально означающей ‘голова живота голени’. Однако следует отметить, что широко распространенным компонентом с семантикой ‘живот’ является 肚 ҚДК *\*dā* (при этом ассимиляция произошла не равномерно), и только в диалектах FCH и FDN используется лексема 腹 ҚДК *\*rik*, иероглифические знаки при этом варьируются, очевидно, из-за семантической близости (см. разбор лексемы ‘живот’).

Трехсложный композит с буквальным значением ‘голова кости голени’ (в котором, таким образом, морфема ‘живот’ заменяется на морфему ‘кость’) был обнаружен только в 3-х диалектах (см. а.2), при этом в диалекте XNC иероглиф 骨 ‘кость’ заменяется на знак 窟 (ҚДК *\*k<sup>h</sup>ūt* ‘пещера, отверстие’), так как оба слова являются почти полными омофонами (см. Hiroyuki 2018: 593 и см. разбор лексемы ‘кость’). В диалекте HUB (а.3) представлен просто бином ‘голова ноги’. Уникальный компонент ҚДК 節 *\*cīt* ‘коленце бамбука, узел, звено’ представлен в диалекте PIN, где все слово должно буквально переводиться как ‘звено головы голени’.

Отдельно следует подчеркнуть ситуацию с ассимиляцией инициали в компоненте ‘голова’ (ҚДК 頭 *\*d<sup>h</sup>ō*), которая отличается от случая, описанного ранее: ассимиляция *\*T- → l-* может происходить, если предыдущий слог относится к ‘темному’ тону (Hiroyuki 2020: 61, 189). Впрочем, как видно из наших данных, это правило не универсально: так, в диалектах FCH, NND, XNC соответствующей ассимиляции не наблюдается. Исходя из всего вышесказанного, мы предлагаем восстанавливать трехкомпонентный вариант ПВМ {骸肚頭}  $*k^{h}a_1-lu_3-lau_2 \sim *k^{h}a_1-lu_3-t^{h}au_2$  с диалектным варьированием на прауровне.

**45) «know/знать»:** а) {八傳} NND  $pət_7-lyŋ_2$ , FUA  $pok_6-liŋ_2$ , HUB  $pyt_6-tyoŋ_2$ , XNC  $pit_6-tyoŋ_2 \sim pyt_6-tyoŋ_2$ , JDU  $put_6-tioŋ_2$ ; а.1) {會八 ~ 解八} GTN  $ε_6=pei_7$ , CNL  $ε_3=rai_6$ , YNT  $e_7=pei_6$ , PIN  $ε_5=peik_6$  ( $tŋ_2$ ); а.2) {八傳} ZHN  $pi_7-iaŋ_2$ ; а.3) FCH  $paik_6$ <sup>13</sup>; б) {曉識 ~ 曉得} FDN  $xieu_3-si_7$ ; б.1) {曉來} FQN  $hie_u_2-li_3$ ; б.2) {曉得} SHON  $xieu_3-li_6$  || ПВМ  $*pət_6$

ҚДК основа 知 *\*tre*, очевидно, перестала быть базисной на ПВМ уровне, так как не зафиксирована ни в одном из языков-потомков данной ветви.

В миньдун отчетливо прослеживаются два корня, которые образуют биномы с разными компонентами. Наиболее широко распространен корень а), не имеющий связи с китайскими литературными эквивалентами. Дж. Норман и Ц. Мэй полагают, что эта лексема была заимствована из австроазиатских языков, приводя в качестве сравнения вьетнамск. *biēt* ‘узнавать, знать’, ‘понимать’ (Norman & Mei 2000: 487). (Иероглифический

<sup>12</sup> Для второго слога указан регистровый тон 11, который, вероятно соотносится с тоном 2 («светлый ровный» 陽平; Hiroyuki 2020: 660).

<sup>13</sup> Накадзима указывает для данного слова исключительно значение ‘узнавать’, однако согласно данным из источников Li 1995, Lin 2002, слово имеет и статическое значение ‘знать’.

знак 八 (КДК \**prēt* ‘восемь’), которым записывается данная основа, подобран исключительно по критерию фонетической близости).

Наиболее частотными расширяющими компонентами для заимствованного корня в ВМ диалектах являются морфемы а.1) \**ε<sub>5</sub>*, которая родственна корню СК 解 *γáj* ‘мочь, уметь’, зафиксированному с эпохи Тан, и а) \**tioŋ<sub>2</sub>*, восходящая, по всей видимости, к лексеме 傳 КДК \**d<sup>h</sup>rwan* ‘передавать’ (Schuessler 2007: 288).

Другая, менее широко распространенная основа (b) 曉 (КДК \**ɲéw*) ‘знать’ берет начало в чуском диалекте КДК со значением ‘знать, понимать’ и также во всех случаях встречается только в составе биномов (Schuessler 2007: 536). Так, в диалекте FDN расширителем является корень 識 КДК \**tək* ‘знать’. В остальных диалектах второй корень распознать не удается.

**46) «leaf/лист»:** а) {葉~箬} FCH nuok<sub>7</sub>, GTN nyøʔ<sub>7</sub>, NND nøk<sub>7</sub>, HUB niɿʔ<sub>7</sub>, XNC nyʔ<sub>7</sub>, JDU nyʔ<sub>7</sub>, ZHN nyʔ<sub>7</sub>, FDN nie<sub>7</sub>, CNL luoʔ<sub>7</sub>, YNT nuoʔ<sub>7</sub>, FUA nik<sub>7</sub> (H), SHON nyøʔ<sub>7</sub> (H), PIN nyoʔ<sub>7</sub> (H); а.1) FQN ts<sup>h</sup>iu<sub>5</sub>=nyo<sub>1</sub> {樹箬} || ПВМ {箬} \*niok<sub>7</sub>

КДК этимон 葉<sup>14</sup> \**l<sup>h</sup>ap* был заменен инновативной лексемой 箬 КДК \**neuk* ‘кора бамбука, листья бамбука’ (восходит, по-видимому, к чускому диалекту, см. Schuessler 2007: 448) на общеминьском уровне, так как не обнаружен ни в одном из диалектов изучаемой диалектной группы в качестве основного слова (Starostin 2019: 163; Лоренц 2020: 143–144).

**47) «lie/лежать»:** а) {倒} FCH tɔ<sub>3</sub>, GTN tɔ<sub>3</sub>, NND tɔ<sub>3</sub>, ZHN tɔ<sub>3</sub>, FDN tɔ<sub>3</sub>, FQN tɔ<sub>3</sub>, FUA tɔ<sub>3</sub>; а.1) {倒下} SHON tɔ<sub>3</sub>-a<sub>5</sub> (H); а.2) {筊倒} PIN ɔuŋ<sub>5</sub>=tɔ<sub>3</sub>, XNC œ<sub>3</sub>=tɔ<sub>3</sub>; b) {暈} HUB k<sup>h</sup>ɔn<sub>4</sub>; c) JDU œuŋ<sub>3</sub> || ПВМ {倒} \*tɔ<sub>3</sub>

В КДК языке глагол ‘лежать’ плохо фиксируется и практически не отделяется от динамических глаголов ‘ложиться, идти спать’ (см. Starostin 2019: 172–173), однако в миньских диалектах довольно отчетливо фиксируется корень 倒 КДК \**táw* ‘падать, переворачиваться’ (Лоренц 2020: 144).

Расширяющие морфемы в диалектах PIN и XNC и JDU лексема œuŋ<sub>3</sub> со значением ‘валиться, ложиться на, полагаться на’ очевидно родственны друг другу, однако их этимология остается неизвестной; иероглифический знак 筊 (КДК \**gāŋ* ‘палка для белья’) вновь подобран из-за фонетической схожести основ (Hiroyuki 2018: 178, 291). Эта же основа встречается в диалекте GTN с тем же значением (Li 2014: 53).

Любопытно, что в диалекте HUB для передачи искомого значения используется лексема 暈 (КДК \**k<sup>h</sup>wān*) ‘спать’; таким образом, здесь представлена полисемия ‘спать/лежать’.

**48) «liver/печень»:** {肝} FCH kaŋ<sub>1</sub>, GTN kaŋ<sub>1</sub>, NND kaŋ<sub>1</sub>, HUB kaŋ<sub>1</sub>, XNC kaŋ<sub>1</sub>, JDU kaŋ<sub>1</sub>, ZHN kaŋ<sub>1</sub>, FDN kaŋ<sub>1</sub>, FQN kaŋ<sub>1</sub>, CHL kaŋ<sub>1</sub>, YNT kaŋ<sub>1</sub>, FUA kaŋ<sub>1</sub>, SHON kaŋ<sub>1</sub>, PIN kaŋ<sub>1</sub> || ПВМ {肝} \*kaŋ<sub>1</sub>

**49) «long/длинный»:** {長} FCH touŋ<sub>2</sub>, GTN tɔuŋ<sub>2</sub>, NND tɔŋ<sub>2</sub>, HUB tɔŋ<sub>2</sub>, XNC tɔŋ<sub>2</sub>, JDU tɔŋ<sub>2</sub>, ZHN tɔŋ<sub>2</sub>, FDN tɔŋ<sub>2</sub>, FQN tɔŋ<sub>2</sub>, CNL touŋ<sub>2</sub>, YNT touŋ<sub>2</sub>, FUA tɔuŋ<sub>2</sub>, SHON tɔŋ<sub>2</sub>, PIN tɔuŋ<sub>2</sub> || ПВМ {長} \*tɔŋ<sub>2</sub>

**50) «louse/вошь»:** {虱母} FCH θaik<sub>6</sub>-mo<sub>4</sub>, GTN seiʔ<sub>6</sub>-mɔ<sub>3</sub>, NND set<sub>6</sub>-mɔ<sub>3</sub>, HUB θet<sub>6</sub>-mo<sub>3</sub>, XNC θet<sub>6</sub>-mɔ<sub>3</sub>, JDU sɛm<sub>6</sub>-mɔ<sub>3</sub>, ZHN set<sub>6</sub>-mɔ<sub>3</sub>, FDN saʔ<sub>6</sub>-mo<sub>3</sub>, FQN seʔ<sub>6</sub>-mo<sub>3</sub>, CNL seik<sub>6</sub>-mo<sub>3</sub>, YNT seiʔ<sub>6</sub>-mo<sub>3</sub>, FUA seik<sub>6</sub>-mɔ<sub>3</sub>, SHON seʔ<sub>6</sub>-mɔ<sub>3</sub>, PIN seik<sub>6</sub>-mɔ<sub>3</sub> || ПВМ {虱母} \*set<sub>6</sub>-mɔ<sub>3</sub>

КДК этимон 虱 \**srit* сохраняется во всех диалектах ВМ ветви, но только в расширенной до бинома форме (с помощью КДК морфемы 母 \**má* ‘мать’).

<sup>14</sup> Данное слово в миньдунских диалектах употребляется исключительно для передачи фамилии (Hiroyuki 2018: 548).

**51) «man/мужчина»:** а) {丈夫儂} FCH tɔŋ<sub>5</sub>=muo<sub>1</sub>-nøŋ<sub>2</sub>, GTN tɔŋ<sub>5</sub>=muo<sub>1</sub>-nøŋ<sub>2</sub>, XNC tɔŋ<sub>2</sub>=mu<sub>1</sub>-(nøeŋ<sub>2</sub>) ~ tœŋ<sub>2</sub>=mu<sub>1</sub>-(nøeŋ<sub>2</sub>), ZHN toŋ<sub>5</sub>=ŋu<sub>1</sub>-nøeŋ<sub>2</sub>, FDN toŋ<sub>5</sub>=muo<sub>1</sub>-neŋ<sub>2</sub>, FQN tɔŋ<sub>6</sub>=muo<sub>1</sub>-nøŋ<sub>2</sub>, YNT tœŋ<sub>6</sub>=muo<sub>1</sub>-nøŋ<sub>5</sub>, FUA tɔŋ<sub>6</sub>=mu<sub>1</sub>-nøeŋ<sub>2</sub>, SHON toŋ<sub>1</sub>=muo<sub>1</sub>-neŋ<sub>2</sub>, PIN tɔŋ<sub>5</sub>=mo<sub>1</sub>-nøeŋ<sub>2</sub>; а.1) {丈夫} NND tɔŋ<sub>5</sub>=mu<sub>1</sub>, HUB tɔŋ<sub>5</sub>=muo<sub>2</sub><sup>15</sup>, JDU tɔŋ<sub>2</sub>=mu<sub>1</sub>; б) FCH naŋ<sub>3</sub>-nøŋ<sub>2</sub> {男人}; б.1) {男界} CNL naŋ<sub>2</sub>-ŋai<sub>4</sub> || ПВМ {丈夫儂} \*tɔŋ<sub>5</sub>=muo<sub>1</sub>-nøeŋ<sub>2</sub>

Старый КДК этимон 男 \*nām ‘мужчина’ наблюдается только в двух соседних диалектах FCH, CNL, где он в принципе мог восстановиться под влиянием литературного языка. В обоих случаях он входит в состав композита: в FCH<sup>16</sup> вторым компонентом оказывается КДК 儂 \*nīŋ ‘человек’ (подробнее см. ‘person/человек’), а в CNL — КДК 界 \*krēc ‘группа, класс’ (исходное значение — ‘граница’, ‘разделительная линия’). При этом по всему ВМ ареалу распространена основа 丈夫 (КДК \*drāŋ pa; исходное значение — ‘муж; мужчина’), обычно также расширенная за счет морфемы 儂 \*nīŋ ‘человек’.

В фонетическом плане важно отметить ассимиляцию на стыке двух слогов как в трехсложном композите 丈夫儂, где инициаль второго слога превращается в губной носовой *m-*, так и в бинеме 男界, где она трансформируется в назальный *n-* или *ŋ-*.

**52) «many/много»:** а) {多 ~ 齊 ~ 齊 ~ 徧 ~ 儕} FCH sa<sub>6</sub>, GTN se<sub>6</sub>, NND se<sub>6</sub>, HUB θe<sub>6</sub>, JDU θe<sub>6</sub>, HUB se<sub>6</sub>, ZHN se<sub>6</sub>, FDN se<sub>6</sub>, FQN se<sub>6</sub>, CNL se<sub>6</sub>, FUA se<sub>6</sub>, SHON se<sub>6</sub>, PIN se<sub>6</sub>; б) FUA {過} kuai<sub>1</sub> || ПВМ \*se<sub>6</sub>

В большинстве миньдунских диалектов КДК этимон 多 \*tāj был вытеснен инновативной лексемой неясного происхождения. По фонетическим причинам ВМ корень невозможно возвести к ханьскому корню 儕 (КДК \*zrāj) ‘множество’, на что намекает соответствующая иероглифическая запись: 1) ВМ лексемы относятся к тональной группе «темный входящий» (陰入), в то время как слово 儕 относится к тональной категории «светлый ровный» (陽平); 2) инициаль *s-* (~ *θ-*) в восточной ветви не может восходить к КДК инициали \**ʒ(r)-* (Starostin 2019: 163). Этимология корня *kuai<sub>1</sub>* из диалекта FUA также остается неясной (по фонетическим причинам свести ее к КДК лексеме 過 \*kwājh ‘превосходить’ не удается; сам знак подобран вторично из-за фонетического сходства основ).

**53) «meat/мясо»:** {肉} FCH nyk<sub>7</sub>, GTN nyʔ<sub>7</sub>, NND nyʔ<sub>7</sub>, HUB nyuk<sub>7</sub>, XNC nɔk<sub>7</sub>, JDU nɔk<sub>7</sub>, ZHN nɔuʔ<sub>7</sub>, FDN nu<sub>7</sub>, FQN nyʔ<sub>7</sub>, SHON nuʔ<sub>7</sub> (H), FUA nɔʔ<sub>7</sub> (H), PIN nyk<sub>7</sub> || ПВМ {肉} \*nik<sub>7</sub>

**54) «moon/луна»:** а) {月} GTN ŋuoʔ<sub>7</sub>, NND ŋɔt<sub>7</sub>, ZHN ŋɔt<sub>7</sub>, FDN ŋuo<sub>7</sub>, CNL ŋuok<sub>7</sub>, YNT ŋuoʔ<sub>7</sub>, FUA ŋuk<sub>7</sub>, SHON ŋyøʔ<sub>7</sub>, PIN ŋɔk<sub>7</sub>; а.1) {月光 ~ 月亮} FCH ŋuok<sub>7</sub>-luoŋ<sub>5</sub>, FQN ŋuoʔ<sub>7</sub>-kuoŋ<sub>5</sub>; а.2) {月奶} HUB ŋɔt<sub>7</sub>-ne<sub>3</sub>, XNC ŋɔt<sub>7</sub>-ne<sub>3</sub>, JDU ŋɔt<sub>7</sub>-ne<sub>3</sub> || ПВМ {月} \*ŋuot<sub>7</sub>

КДК 月 \*ŋwat стабильно сохраняется во всех языках-потомках миньдунской ветви, часто расширяясь до двуслога либо за счет присоединения морфемы 光 (КДК \*kwājh) ‘свет, блеск’ (иногда записывается сходным по семантике знаком 亮 (КДК \*raŋh) ‘свет’), либо за счет этимологически неясной морфемы (записывается иероглифическим знаком 奶 *nǎ* (СК *nǎ*) ‘грудь’, подобранным по принципу фонетического сходства).

**55) «mountain/гора»:** {山} FCH θaŋ<sub>1</sub>, GTN saŋ<sub>1</sub>, NND saŋ<sub>1</sub>, HUB θaŋ<sub>1</sub>, XNC θaŋ<sub>1</sub>, JDU saŋ<sub>1</sub>, ZHN saŋ<sub>1</sub>, FDN saŋ<sub>1</sub>, FQN saŋ<sub>1</sub> (θaŋ<sub>1</sub>), CNL saŋ<sub>1</sub>#, YNT saŋ<sub>1</sub>#, FUA saŋ<sub>1</sub>#, SHON saŋ<sub>1</sub>#, PIN saŋ<sub>1</sub> || ПВМ {山} \*saŋ<sub>1</sub>

**56) «mouth/рот»:** {喙~嘴} FCH ts<sup>h</sup>ui<sub>4</sub>, GTN ts<sup>h</sup>y<sub>4</sub>, NND ts<sup>h</sup>oi<sub>4</sub>, HUB ts<sup>h</sup>ui<sub>4</sub>, XNC t<sup>h</sup>ui<sub>4</sub>, JDU ts<sup>h</sup>oi<sub>4</sub>, ZHN ts<sup>h</sup>oi<sub>4</sub>, FDN ts<sup>h</sup>ui<sub>4</sub>, FQN ts<sup>h</sup>ui<sub>4</sub>, CNL ts<sup>h</sup>uoi<sub>4</sub>, YNT ts<sup>h</sup>uoi<sub>4</sub>, FUA ts<sup>h</sup>øi<sub>4</sub>, SHON ts<sup>h</sup>y<sub>4</sub>, PIN ts<sup>h</sup>ui<sub>4</sub> || ПВМ {喙} \*ts<sup>h</sup>ui<sub>4</sub>

<sup>15</sup> А. Хироюки указывает регистровый тон 34, который, вероятно, соотносится с 2-м тоном («светлый ровный» 陽平; Hiroyuki 2018: 324).

<sup>16</sup> В фучжоуском диалекте также возможен лексический вариант *naŋ<sub>2</sub>-ŋai<sub>4</sub>*, как и в диалекте CNL; очевидно, это связано с ареальным влиянием (Lin 2002: 50; Liu & He 1998: 512).

Основной КДК этимон 口 *\*k<sup>h</sup>ō*, который в литературном китайском языке можно считать базовым термином вплоть до начала цинской эпохи, не прослеживается ни в одном из миньских диалектов, поэтому мы фиксируем лексическую замену уже на ПМ уровне (Starostin 2019: 169; Лоренц 2020: 146).

Данные показывают, что инновативной лексемой, вытеснившей старую основу, стала лексема КДК 喙 *\*x<sup>h</sup>was*, ранее обозначавшая ‘клюв’; таким образом, мы наблюдаем типичную для синитического ареала семантическую деривацию ‘клюв’ → ‘рот’. Нередко для записи этого слова из-за семантической схожести используется иероглифический знак 嘴 (КДК *\*cwáj*) ‘клюв’, который также восходит к ханьскому термину, но с иной этимологией.

**57) «name/имя»:** а) {名} GTN mian<sub>2</sub>#, NND mian<sub>2</sub>#, HUB mien<sub>2</sub>, XNC mien<sub>2</sub>, JDU mien<sub>2</sub>, ZHN mian<sub>2</sub>#, FDN mian<sub>2</sub>#, FQN mian<sub>2</sub>, SHON mian<sub>2</sub>, FUA mien<sub>2</sub>, PIN mian<sub>2</sub>; а.1) FCH mian<sub>2</sub>-tsei<sub>5</sub> {名字} || ПБМ {名} *\*mian<sub>2</sub>*

**58) «neck/шея»:** а) {脰項 ~ 脰管 ~ 脰引} GTN tau<sub>5</sub>-uŋ<sub>4</sub>, NND tau<sub>5</sub>-oŋ<sub>5</sub>, HUB ta<sub>5</sub>-un<sub>5</sub>, XNC tau<sub>5</sub>-oŋ<sub>3</sub>, JDU tau<sub>5</sub>-yn<sub>3</sub>, ZHN tau<sub>5</sub>-yŋ<sub>3</sub>, FDN tau<sub>5</sub>-uŋ<sub>3</sub>, FUA tau<sub>1</sub>-uŋ<sub>3</sub>, PIN ta<sub>5</sub>-uŋ<sub>3</sub>; а.1) {脰脰 ~ 脰骨} FCH ta<sub>2</sub>-au<sub>3</sub>, CNL tau<sub>2</sub>-ɔuk<sub>6</sub>, YNT tau<sub>5</sub>-uʔ<sub>6</sub>, FQN tau<sub>5</sub>-uɔʔ<sub>6</sub>; а.2) {脰脖株} FQN tau<sub>5</sub>-(p)βuoʔ<sub>2</sub>-ty<sub>1</sub>; а.3) {脰領} SHON tau<sub>1</sub>-lian<sub>3</sub> || ПБМ {脰項} *\*tau<sub>5</sub>-uŋ<sub>3</sub>*

КДК этимон 領 *\*r<sup>h</sup>éŋ* ‘шея’ сам по себе не зафиксирован ни в одной из ветвей группы Минь в качестве основного корня, но присутствует как расширитель в диалекте SHON (Starostin 2019: 170; Лоренц 2020: 147). Инновативная лексема 脰 *\*d<sup>h</sup>ōh* ‘шея’ восходит к цисскому слову эпохи Чуныцю-Чжаньго (VIII–III вв. до н.э.) и, предположительно, в КДК период имеет более узкое значение ‘шейный позвонок’, т.к. ее употребление прослеживается в контекстах, связанных с немедленной смертью (Schuessler 2007: 216; Starostin 2019: 170–171).

Как видно из наших данных, корень всегда употребляется только в составе композитов; в качестве расширителей чаще всего выступают (а) КДК 項 *\*grōŋ* ‘шея (тыльная часть)’ и (а.1) КДК 骨 *\*kwōt* ‘кость’ (см. подробный разбор этимона ‘bone/кость’). Во всех биномах наблюдается усечение инициали второго слога, характерное для миньских диалектов (Hiroyuki 2018: 36).

**59) «new/новый»:** {新} FCH θin<sub>1</sub>, GTN sin<sub>1</sub>, NND sin<sub>1</sub>, HUB θin<sub>1</sub>, XNC θin<sub>1</sub>, JDU sin<sub>1</sub>, ZHN sin<sub>1</sub>, FDN sin<sub>1</sub>, FQN sin<sub>1</sub>, FUA θein<sub>1</sub>, SHON sin<sub>1</sub>, PIN sin<sub>1</sub> || ПБМ {新} *\*sin<sub>1</sub>*

**60) «night/ночь»:** а) {冥晡 ~ 瞑晡} FCH maŋ<sub>1</sub>-muo<sub>1</sub>, GTN maŋ<sub>2</sub>-muo<sub>1</sub>, NND maŋ<sub>2</sub>-mu<sub>1</sub>, HUB maŋ<sub>2</sub>-ŋuo<sub>1</sub>, XNC maŋ<sub>2</sub>-ŋu<sub>1</sub>, JDU maŋ<sub>2</sub>-ŋu<sub>1</sub>, ZHN maŋ<sub>2</sub>-mu<sub>1</sub>, FQN maŋ<sub>2</sub>-muo<sub>1</sub>, FUA maŋ<sub>2</sub>-mu<sub>1</sub>, PIN maŋ<sub>2</sub>-mo<sub>1</sub>; а.1) {曩晡頭} CNL laŋ<sub>2</sub>-muo<sub>3</sub>-lau<sub>2</sub>, YNT maŋ<sub>2</sub>-muo<sub>1</sub>-lau<sub>2</sub>; а.2) {瞑過} FDN maŋ<sub>2</sub>-kuo<sub>7</sub>; а.3) {瞑間} SHON maŋ<sub>2</sub>-kaŋ<sub>1</sub> || ПБМ {冥晡} *\*maŋ<sub>2</sub>-mu<sub>1</sub>*

Широко распространенный в классический период этимон 夜 *\*liah* вытеснен инновацией еще на общеминьском уровне, так как ни один из языков-потомков не сохранил старый корень (Starostin 2019: 163; Лоренц 2020: 147). Во всей миньдунской ветви используется корень 冥 КДК *\*mēŋ* ‘быть темным’, который в этом значении фиксируется уже в классических памятниках; семантическое развитие ‘темный’ → ‘ночь’ в миньдунских диалектах типологически естественно для самых разных ареалов. Иероглифический знак, используемый для записи этого корня, вариативен и часто записывается с ключом ‘глаз’ (瞑).

В автономной форме этот корень в ВМ диалектах не используется и обычно встречается в двухсложных (реже — в трехсложных) композитах. Чаще всего расширителем оказывается лексема 晡 КДК *\*pā* ‘вечер’, которая начинает встречаться в литературных

памятниках с эпохи Хань (HYDC V: 729); инициаль второго слога в этом контексте ассимилируется предшествующей носовой терминали.

**61) «nose/нос»:** а) {鼻} FCH p<sup>h</sup>ei<sub>4</sub>, GTN p<sup>h</sup>i<sub>4</sub>, NND p<sup>h</sup>ei<sub>4</sub>, HUB p<sup>h</sup>i<sub>5</sub>, XNC p<sup>h</sup>i<sub>5</sub>, JDU p<sup>h</sup>ei<sub>5</sub>, ZHN p<sup>h</sup>ei<sub>4</sub>, FDN p<sup>h</sup>i<sub>4</sub>, FQN p<sup>h</sup>e<sub>4</sub>, CNL p<sup>h</sup>ei<sub>4</sub>, YNT p<sup>h</sup>e<sub>4</sub>; а.1) {鼻公 ~ 鼻空} FUA p<sup>h</sup>i<sub>3</sub>-œŋ<sub>1</sub>; а.2) {鼻頭} SHON p<sup>h</sup>i<sub>4</sub>-lau<sub>2</sub>; а.3) {鼻古} PIN p<sup>h</sup>i<sub>4</sub>-u<sub>3</sub> || ПВМ {鼻} \*p<sup>h</sup>i<sub>4</sub>

**62) «not/не»:** {不 ~ 無 ~ 唔} FCH me<sub>1</sub>, GTN ŋ<sub>6</sub>#, NND ŋ<sub>6</sub>#, HUB n<sub>3</sub>, XNC n<sub>1</sub>, JDU n<sub>3</sub>, ZHN ŋ<sub>6</sub>#, FDN ŋ<sub>6</sub>#, FQN n<sub>1</sub>, CNL iŋ<sub>1</sub>#, YNT iŋ<sub>1</sub>#, FUA ŋ<sub>6</sub>#, SHON ŋ<sub>6</sub>#, PIN mɔ<sub>2</sub> || ПВМ \*m<sub>2</sub> (~ \*ŋ<sub>2</sub>)

Устойчивый литературный эквивалент 不 \*pə был вытеснен уже на ПМ уровне континуантом 無 (КДК \*ma) с исходно запретительной семантикой ‘не надо, не’ (Starostin 2019: 163; Лоренц 2020: 148)<sup>17</sup>. Фонетические рефлекс отрицательной частицы варьируются на стыке слов; из-за этого возникают фантомные представления об использовании в разговорной речи этимологически разных лексем, которые начинают маркироваться разными графическими знаками. Непредсказуемые тональные вариации в приведенных примерах также связаны с тональными сандхи, вызванными связанным характером морфемы. На ПВМ уровень допустимо выносить (в качестве дополнительно распределенных) оба варианта — как исходный с лабиальной инициальной /m/, так и ассимилированный с инициальной /ŋ/.

Стоит добавить, что отрицание может стягиваться с последующим словом, заменяя инициаль второго слога: PIN 聽 t<sup>h</sup>ian<sub>1</sub> — 不聽 paŋ<sub>553</sub> ‘слышать’ — ‘не слышать’, 爬 pa<sub>2</sub> — 不爬 ma<sub>453</sub> ‘подниматься’ — ‘не подниматься’, 坐 sɔi<sub>5</sub> — 不做 nɔi<sub>512</sub> ‘сидеть’ — ‘не сидеть’ (Hiroyuki 2020: 392–394; Hiroyuki 2018: 654; Liu & He 1998: 75).

Для записи чаще всего используются 3 знака: 1) 無 (КДК \*ma), за которым и в литературном языке закреплены значения ‘не надо, не; не иметь’; 2) 不 — видоизмененный графический вариант простого иероглифа 不 ‘не’; 3) 唔 — подобран из соображений фонетической близости, впервые используется в литературных памятниках начиная с эпохи Мин для передачи звукоподражания (HYDC III: 356). Ср. с таким же развитием когнаты из других диалектных групп: У (шанхайск. m<sub>2</sub>), Юэ (гуанчжоу m<sub>2</sub>), хакка m<sub>2</sub> и др. (цит. по Schuessler 2007: 518–519).

**63) «one/один»:** а) {蜀} FCH θuok<sub>7</sub>, GTN syø<sub>7</sub>, NND sø<sub>7</sub>, JDU ɕy<sub>7</sub>, ZHN sɔ<sub>7</sub>, FDN suo<sub>7</sub>, FQN θyo<sub>7</sub> (H), CNL so<sub>4</sub>#, YNT so<sub>7</sub>5#, PIN sək<sub>7</sub>, б) XNC ɕit<sub>7</sub>, FUA sik<sub>7</sub>#, SHON si<sub>7</sub> || ПВМ {蜀} \*suok<sub>7</sub>

Исконное числительное КДК — \*ʔit ‘один’, по-видимому, заместило инновативной лексемой во всем миньском ареале уже на общеминьском уровне, сохраняясь лишь в составе отдельных устойчивых выражений (Starostin 2019: 163; Лоренц 2020: 149).

Этимологическое происхождение этого слова неясно; Нгай предполагает возможную связь с ДК корнем 獨 \*dōk ‘одинокый; уникальный’ (Ngai 2015: 213), однако развитие \*d- → \*s- перед исконными долгими гласными для миньских диалектов нерегулярно. При условии принятия спорадического сокращения долготы (\*dōk → \*dok → ПМ \*ʔok → ВМ \*suok), что в принципе логично для столь частотного числительного, этимология приемлема. (Иероглифический знак 蜀, транскрибирующий данное слово, как раз имеет СК чтение с гласным третьего дэна ʔöik, восходящее к ДК \*dok).

Форма, представленная в диалектах XNC, FUA, SHON, вряд ли может отражать базовый КДК этимон — \*ʔit ‘один’, так как надежных примеров на возможность развития \*ʔ- → \*s- не обнаружено; впрочем, других идей относительно ее вероятного происхождения также нет.

<sup>17</sup> В некоторых ВМ диалектах этот же корень передает предикативное значение ‘не иметь’ (Hiroyuki 2020: 696).



**64) «person/человек»:** а) {儂} FCH nøy<sub>2</sub>, GTN nøy<sub>2</sub>, NND nøy<sub>2</sub>, HUB nøy<sub>2</sub>, XNC nøy<sub>2</sub>, JDU nøy<sub>2</sub>, ZHN nøy<sub>2</sub>, FDN nøy<sub>2</sub>, FQN nøy<sub>2</sub>, CNL løy<sub>2</sub>, YNT nøy<sub>2</sub>, FUA nøy<sub>2</sub>, SHON nøy<sub>2</sub>, PIN nøy<sub>2</sub>; б) {陣} FQN teŋ<sub>5</sub> (H), FUA teŋ<sub>5</sub> (H), PIN teŋ<sub>5</sub> || ПВМ {儂} \*nøy<sub>2</sub>

КДК этимон 人 \*nin в миньской ветви вытесняется инновацией 農 (КДК \*niŋ), исходно = ‘земледелец, крестьянин’ (Starostin 2019: 163; Лоренц 2020: 149–150); для записи миньского слова обычно используется знак, расширенный за счет ключа ‘человек’ (儂). В литературном языке расширенное значение ‘человек’ у этого слова не встречается, но, как отмечает Чжоу Фагао, в качестве местоимения 1-го лица эта лексема встречается в стихотворении эпохи Южных династий (420–589 гг.) «Песня Цзы-е» (子夜歌; цит. по Norman 1983: 208). Норман и Мэй отмечают аналогичное использование данной лексемы (‘крестьянин’ → ‘человек’) также среди диалектов группы У (Norman 1983: 208; Mei 2015: 123).

Для трех диалектов, описанных в исследовании Хироюки, также приводится полный синоним б) неизвестного происхождения (Hiroyuki 2020: 649).

**65) «rain/дождь»:** {雨} FCH hu<sub>6</sub> (LH), GTN hu<sub>5</sub>, NND hu<sub>3</sub>, HUB hu<sub>6</sub>, XNC hu<sub>6</sub>, JDU hu<sub>6</sub>, ZHN hu<sub>5</sub>, FDN hu<sub>6</sub> (LH), FQN hu<sub>6</sub>, CNL hu<sub>6</sub>, YNT hu<sub>6</sub>, FUA hu<sub>6</sub>, SHON u<sub>6</sub> (LH), PIN hu<sub>6</sub> (LH) || ПВМ {雨} \*hu<sub>6</sub>

КДК этимон 雨 \*w<sup>h</sup>á ‘дождь’ стабильно сохраняется в современных диалектах, однако при этом часто наблюдается свободное варьирование литературного и разговорного чтений. В источнике Liu & He 1998: 481–482 оба варианта четко разграничиваются, причем «литературные» формы, как правило, имеют вид i<sub>3</sub> или y<sub>3</sub> с восходящим тоном (上聲), но в других источниках ситуация более хаотична. Так, для диалекта FCH Накадзима указывает литературное чтение; для диалекта FDN Чэнь и Ли также дают только литературное чтение; то же самое для диалектов PIN, SHON и FQN у Хироюки (Nakajima 1979: 4; Chen & Li 1991: 12; Hiroyuki 2020: 616). Возможно, это связано с тем, что «литературное» чтение в миньдунских диалектах характерно для некоторых устойчивых сочетаний, таких, как, например, ‘идет дождь’: 遇雨 ~ 桐雨 FCH touŋ<sub>2</sub> ŋy<sub>3</sub>, CNL touŋ<sub>2</sub> ŋy<sub>3</sub>, FQN touŋ<sub>2</sub> ŋy<sub>3</sub>, YNT touŋ<sub>1</sub> y<sub>3</sub>, GTN touŋ<sub>4</sub> ŋy<sub>2</sub> и т. д. (Lin 2002: 33).

**66) «red/красный»:** {紅} FCH øy<sub>2</sub>, GTN øy<sub>2</sub>, NND øy<sub>2</sub>, HUB øy<sub>2</sub>, XNC øy<sub>2</sub>, JDU øy<sub>2</sub>, ZHN øy<sub>2</sub>, FDN øy<sub>2</sub>, FQN øy<sub>2</sub>, CNL øy<sub>2</sub>, YNT øy<sub>2</sub>, FUA øy<sub>2</sub>, SHON øy<sub>2</sub>, PIN øy<sub>2</sub> || ПВМ {紅} \*øy<sub>2</sub>

Основной КДК этимон 赤 \*k<sup>h</sup>iak ‘красный’ в современных миньских диалектах не встречается, т.к. уже на ПМ уровне он был вытеснен инновацией 紅 (КДК \*g<sup>h</sup>ōŋ), редко встречающейся в КДК текстах, но активно используемой начиная с периода Хань (Starostin 2019: 169; Лоренц 2020: 150).

**67) «road/дорога»:** а) {壻} FCH tuo<sub>5</sub>, GTN tuo<sub>5</sub>, NND tu<sub>5</sub>, HUB tuo<sub>6</sub>, XNC tu<sub>6</sub>, JDU tu<sub>6</sub>, ZHN tu<sub>5</sub>, FDN tuo<sub>5</sub>, FQN tuo<sub>5</sub>, CNL tu<sub>5</sub>#, YNT tuo<sub>5</sub>#, SHON ty<sub>5</sub>#; б) {路} FCH lou<sub>6</sub> || ПВМ {壻} \*tuo<sub>5</sub>

По всему ВМ ареалу в значении ‘дорога’ используется диалектная инновация 壻<sup>18</sup> (СК dò), по-видимому, морфологически связанная с КДК основой 途 \*L<sup>h</sup>ā ‘дорога’, но отличающаяся от нее наличием падающего тона.

В диалекте FCH, помимо этого, в качестве синонима также используется литературное заимствование 路 lou<sub>6</sub>, восходящее к КДК 路 \*rāh ‘дорога, путь’.

**68) «root/корень»:** а) {根} HUB kyŋ<sub>1</sub>, XNC kyŋ<sub>1</sub>, JDU kyŋ<sub>1</sub>, FQN kyŋ<sub>1</sub>, FUA køŋ<sub>1</sub> (H), SHON kyŋ<sub>1</sub>(H), PIN kyŋ<sub>1</sub>; а.1) FCH tsu<sub>4</sub>=kyŋ<sub>1</sub> {樹根}, GTN ts<sup>h</sup>iu<sub>4</sub>=kyŋ<sub>1</sub> || ПВМ {根} \*kyŋ<sub>1</sub>

КДК этимон 本 \*p<sup>h</sup>ān не прослеживается ни в одном из диалектов группы Минь, поскольку был вытеснен инновативным корнем КДК 根 \*kān еще на общеминьском уровне

<sup>18</sup> Иероглифический знак 壻 за пределами среднекитайских словарей не встречается.

(Starostin 2019: 166–167; Лоренц 2020: 150). Следует отметить нерегулярное развитие финали, т.к. КДК *\*-ən* обычно отражается в ВМ диалектах как *-ouŋ ~ -uŋ ~ -oŋ*. ПВМ финаль *\*-uŋ* результат развития из КДК *\*-oŋ, \*-ən, \*-əŋ*.

**69) «round/круглый»:** {圓} FCH *ieŋ<sub>2</sub><sup>19</sup>*, FQN *ieŋ<sub>1</sub>*, HUB *ieŋ<sub>2</sub>*, JDU *in<sub>2</sub>*, XNC *in<sub>2</sub>* || ПВМ {圓} *\*ieŋ<sub>2</sub>*

**70) «sand/песок»:** {沙} FCH *θai<sub>1</sub>*, GTN *sai<sub>1</sub>*, HUB *θuo<sub>1</sub>*, XNC *θuɔ<sub>1</sub>*, JDU *suo<sub>1</sub>*, FQN *θua<sub>1</sub>*, FUA *θo<sub>1</sub>* (H), SHON *sua<sub>1</sub>* (H), PIN *sai<sub>1</sub>* || ПВМ {沙} *\*sai<sub>1</sub>*

**71) «say/сказать»:** а) {講話} FCH *kouŋ<sub>3</sub>-ua<sub>6</sub>*, GTN *kouŋ<sub>3</sub>-ua<sub>5</sub>*, NND *kɔŋ<sub>3</sub>-uɔ<sub>5</sub>*, HUB *kɔŋ<sub>3</sub>-ŋuo<sub>5</sub>*, XNC *kɔŋ<sub>3</sub>-ŋuɔ<sub>5</sub>*, JDU *kɔŋ<sub>3</sub>-ŋuo<sub>5</sub>*, ZHN *kɔŋ<sub>3</sub>-uɔ<sub>5</sub>*, FDN *kouŋ<sub>3</sub>-ua<sub>5</sub>*, CNL *kouŋ<sub>3</sub>-ua<sub>6</sub>*, YNT *kouŋ<sub>3</sub>-ua<sub>6</sub>*, FUA *kouŋ<sub>3</sub>-wo<sub>6</sub>*, SHON *kouŋ<sub>3</sub>-ua<sub>6</sub>*; а.1) {講} FQN *kouŋ<sub>3</sub>*, PIN *kouŋ<sub>3</sub>* || ПВМ {講話} *\*kɔŋ<sub>3</sub>-ua<sub>5</sub>*

В миньских диалектах, как и во всех синитических языках, значения ‘сказать’ и ‘говорить’ разграничить довольно трудно; ввиду нехватки данных (особенно контекстов) у нас нет твердой уверенности в том, что все представленные эквиваленты отражают именно однократное значение ‘сказать’.

В любом случае очевидно, что ни один миньский диалект не сохраняет в значении ‘сказать’ КДК этимон 曰 *\*wat* (Starostin 2019: 172; Лоренц 2020: 151). Главной инновацией внутри ВМ ареала оказывается бином 講話, первый слог которого восходит к КДК 講 *\*króŋ* ‘объяснять, дискутировать’, а второй — к КДК 話 *\*g<sup>hw</sup>rāc* ‘речь; говорить’. Иногда встречается также простой односложный вариант 講 *\*kɔŋ<sub>3</sub>*.

**72) «see/видеть»:** а) {見} FCH *kieŋ<sub>4</sub>* (LN), SHON *kieŋ<sub>5</sub>* (H), PIN *keiŋ<sub>5</sub>*, HUB *ken<sub>5</sub>*, XNC *kin<sub>5</sub>*, JDU *kin<sub>5</sub>*; а.1) {看見} FQN *k<sup>h</sup>aŋ<sub>4</sub>=kieŋ<sub>4</sub>* (*ŋieŋ<sub>4</sub>*), FUA *aŋ<sub>3</sub>=ŋieŋ<sub>4</sub>* (H), PIN *aŋ<sub>3</sub>=ŋeiŋ<sub>4</sub>*; б) {映着} SHON *ɔŋ<sub>5</sub>-tyøʔ<sub>7</sub>* (H) || ПВМ {見} *\*kieŋ<sub>5</sub>*

В словарях и словниках по диалектам Минь, как правило, плохо разграничиваются значения ‘видеть’ и ‘смотреть’; мы стараемся приводить лексемы, соответствующие мандаринскому глаголу 見 *jìàn* ‘видеть’, но в таком виде они зафиксированы лишь в ограниченном количестве диалектов.

В автономном употреблении рефлекс старой лексемы 見 *\*kēnh* наблюдаются в шести диалектах; при этом Хироюки отмечает, что в диалектах HUB, XNC, JDU она может быть расширена до бинома с помощью основы КДК 看 *\*k<sup>h</sup>ānh* ‘смотреть’: HUB *aŋ<sub>4</sub>=ŋeŋ<sub>4</sub>*, XNC *aŋ<sub>1</sub>=ŋin<sub>4</sub>*, JDU *k<sup>h</sup>aŋ<sub>1</sub>=ŋin<sub>4</sub>* (Hiroyuki 2018: 355). В остальных диалектах фиксируется только этот двусложный вариант, причем абсолютно везде на стыке морфем имеет место ассимиляция.

**73) «seed/семя»:** а) {籽 ~ 子} GTN *tsi<sub>3</sub>*, NND *tsi<sub>3</sub>*, ZHN *tsi<sub>3</sub>*, FDN *tsi<sub>3</sub>*; а.1) FCH *tsyŋ<sub>1</sub>-tsi<sub>2</sub>* {種子} || ПВМ {子} *\*tsi<sub>3</sub>*

В восточноминьской ветви КДК этимон 種 *\*tóŋ* ‘семя’ прослеживается только в диалекте FCH, где он расширен за счет дополнительного корня 子 (КДК *\*cá*) ‘сын, ребенок, потомство’. Во всех остальных диалектах значение ‘семя’, напротив, передается именно последним корнем, записываясь иногда как 籽 с дополнительным ключом ‘рис’ (в литературном языке этот иероглиф встречается только в позднесредневековых памятниках). Семантический переход ‘сын, ребенок’ → ‘семя’ в целом достаточно характерен для синитического ареала.

Судя по дистрибуции рефлексов, на ПВМ уровне основным эквивалентом значения ‘семя’ уже был однослог 子 *\*tsi<sub>3</sub>*, хотя на этот переход могло повлиять его первоначаль-

<sup>19</sup> Ли и Лян указывают для однослога 圓 *ieŋ<sub>2</sub>* значение ‘собираться вместе’; само значение ‘круглый’ передается трехслогом 圓珠珠 *ieŋ<sub>2</sub>-ʃou<sub>1</sub>-ʃou<sub>1</sub>* (Li & Liang 1994: 156).

ное использование в составе двуслога 種子; с другой стороны, сохранение его конкретно в фучжоуском диалекте может быть результатом архаизации (влияния литературного языка).

**74) «sit/сидеть»:** {坐} FCH θuoi<sub>5</sub>, GTN soi<sub>5</sub>, NND соу<sub>5</sub>, HUB θoi<sub>5</sub>, XNC θoi<sub>5</sub>, JDU soi<sub>5</sub>, ZHN соi<sub>5</sub>, FDN soi<sub>5</sub>, FQN soi<sub>4</sub>, FUA со<sub>5</sub> (H), SHON соi<sub>5</sub> (H), PIN соi<sub>5</sub> || ПВМ {坐} \*soi<sub>5</sub>

**75) «skin/кожа»** {皮} FCH p<sup>h</sup>ui<sub>2</sub>, GTN p<sup>h</sup>ui<sub>2</sub>, NND p<sup>h</sup>ui<sub>2</sub>, HUB p<sup>h</sup>ui<sub>2</sub>, XNC p<sup>h</sup>ui<sub>2</sub>, JDU p<sup>h</sup>ui<sub>2</sub>, ZHN p<sup>h</sup>ui<sub>2</sub>, FDN p<sup>h</sup>uei<sub>2</sub>, FQN p<sup>h</sup>ui<sub>2</sub>#, CHL p<sup>h</sup>ui<sub>2</sub>, YNT p<sup>h</sup>ui<sub>2</sub>, FUA p<sup>h</sup>ui<sub>2</sub>, SHON p<sup>h</sup>ui<sub>2</sub>, PIN p<sup>h</sup>oi<sub>2</sub> || ПВМ {皮} \*p<sup>h</sup>ui<sub>2</sub>

КДК этимон 膚 \*pra ‘кожа (человека)’ был заменен инновативной лексемой 皮 (КДК \*b<sup>h</sup>aj) ‘кожа (животного)’ еще на общеминьском уровне, так как не зафиксирован ни в одной из ветвей; практически во всех источниках слово 皮 используется исключительно для описания именно человеческой кожи, а не шкуры животного (Starostin 2019: 165; Лоренц 2020: 152).

**76) «sleep/спать»:** {睏} FCH k<sup>h</sup>uon<sub>5</sub>, GTN k<sup>h</sup>uŋ<sub>5</sub><sup>20</sup>, NND k<sup>h</sup>on<sub>5</sub>, HUB k<sup>h</sup>on<sub>5</sub>, XNC k<sup>h</sup>on<sub>5</sub>, JDU k<sup>h</sup>on<sub>5</sub>, ZHN k<sup>h</sup>on<sub>5</sub> (LH), FDN k<sup>h</sup>uŋ<sub>5</sub> (LH), FQN k<sup>h</sup>on<sub>5</sub>, CNL k<sup>h</sup>ouŋ<sub>4</sub>, YNT k<sup>h</sup>ouŋ<sub>4</sub>, FUA k<sup>h</sup>uŋ<sub>4</sub><sup>21</sup>, SHON k<sup>h</sup>uŋ<sub>4</sub>, PIN k<sup>h</sup>on<sub>5</sub> || ПВМ {睏} \*k<sup>h</sup>on<sub>5</sub>

Здесь налицо лексическая замена на общеминьском уровне: КДК 臥 \*ŋ<sup>h</sup>wājh<sup>22</sup> ‘лежать, спать’ → 睏 \*k<sup>h</sup>wānh, т.к. старый этимон не сохраняется ни в одном из современных языков-потомков (Starostin 2019: 171; Лоренц 2020: 152). Инновация 睏 \*k<sup>h</sup>wānh, вероятно, представляет собой результат семантического сдвига от КДК 困 \*k<sup>h</sup>wānh ‘устать, утомиться’; первые употребления данного слова фиксируются только в памятниках конца циньской эпохи и отмечены как диалектные формы (HYDC VII: 1222).

**77) «small/маленький»:** а) {嫩} FCH nouŋ<sub>5</sub>, GTN н<sub>5</sub>у<sub>5</sub>#, NND н<sub>5</sub>н<sub>5</sub>#, HUB н<sub>5</sub>н<sub>5</sub>, XNC н<sub>5</sub>н<sub>5</sub>, JDU н<sub>5</sub>н<sub>5</sub>, ZHN н<sub>5</sub>н<sub>5</sub>#, FDN н<sub>5</sub>н<sub>5</sub> (LN), FQN н<sub>5</sub>н<sub>5</sub> (H), CNL н<sub>5</sub>у<sub>5</sub>, YNT н<sub>5</sub>у<sub>5</sub>, FUA н<sub>5</sub>у<sub>5</sub>; б) {細} FDN se<sub>5</sub>#, FQN se<sub>4</sub>, SHON se<sub>4</sub>, PIN se<sub>4</sub> || ПВМ {嫩} \*n<sub>5</sub>н<sub>5</sub>

Старый КДК этимон 小 \*sew, по всей видимости, был вытеснен инновацией 嫩 ‘молодой, нежный’ (КДК \*nwānh) уже на ПВМ уровне, так как он совсем не прослеживается в миньдунской ветви (Starostin 2019: 163-164; Лоренц 2020: 153); таким образом, отмечается семантический сдвиг ‘молодой, нежный’ → ‘маленький’.

Помимо этого, в целом ряде ВМ диалектов также распространен корень, восходящий к КДК 細 \*sēh ‘мелкий, крошечный; тонкий’, но в диалектах FDN, FQN является синонимом без каких-либо сведений о разнице между двумя лексемами.

**78) «smoke/дым»:** а) {煙} FQN iŋ<sub>1</sub>, FUA eiŋ<sub>1</sub> (H), SHON ieŋ<sub>1</sub> (H), PIN iŋ<sub>1</sub>; а.1) FCH hui<sub>3</sub>=ieŋ<sub>1</sub> {火煙} || ПВМ {煙} \*iŋ<sub>1</sub>

Архаичный КДК этимон 熏 \*hun в ВМ диалектах встречается только в узкоспециализированном значении ‘дым от трубки’ (Lin 2002: 96; Li & Liang 1994: 138); однако базовый эквивалент 煙 \*iŋ<sub>1</sub> ‘дым’ также унаследован от КДК состояния (Starostin 2019: 167-168).

<sup>20</sup> В диалекте гутянь лексема может также употребляться в расширенной до бинома форме {睏眠} k<sup>h</sup>ouŋ<sub>5</sub>-miŋ<sub>1</sub>, где второй компонент 眠 = КДК \*mīn ‘закрывать глаза, сон, спать’ (Liu & He 1998: 564; Li 2014: 105).

<sup>21</sup> Хироюки указывает для диалекта фуань форму ŋ<sub>5</sub>у<sub>5</sub> со «светлым» падающим тоном (陽去; Hiroyuki 2020: 632).

<sup>22</sup> Следует отметить употребление лексемы 臥 nguoi<sub>7</sub> в значении ‘спать’ в фучжоуском диалекте по данным словаря XIX в., однако, вероятно, уже как стилистически маркированный архаизм, т.к. в качестве базового разговорного эквивалента указывается основа 睏 k<sup>h</sup>uon<sub>5</sub> (Macklay & Baldwin 1898: 548).

**79) «stand/стоять»:** {倚 ~ 企} FCH k<sup>h</sup>ie<sub>5</sub>, GTN k<sup>h</sup>ie<sub>5</sub>, NND k<sup>h</sup>ie<sub>3</sub>, HUB k<sup>h</sup>ie<sub>5</sub>, XNC k<sup>h</sup>ie<sub>5</sub>, JDU k<sup>h</sup>ie<sub>5</sub>, ZHN k<sup>h</sup>ie<sub>5</sub>, FDN k<sup>h</sup>ia<sub>5</sub>, FQN k<sup>h</sup>ia<sub>4</sub>, CNL k<sup>h</sup>ie<sub>5</sub>, YNT k<sup>h</sup>ie<sub>5</sub>, FUA k<sup>h</sup>e<sub>5</sub>, SHON k<sup>h</sup>ia<sub>5</sub>, PIN k<sup>h</sup>e<sub>5</sub> || ПВМ {倚} \*k<sup>h</sup>ie<sub>5</sub>

КДК этимон 立 \*rəp ‘стоять’ вытесняется инновацией 企 \*k<sup>h</sup>é (исходно — ‘стоять на цыпочках’) уже на ПМ уровне, т.к. старый этимон не зафиксирован ни в одной из ветвей (Starostin 2019: 169; Лоренц 2020: 153). Лексема чаще всего записывается редким иероглифическим знаком 倚, впервые зафиксированным в значении ‘стоять’ только в словаре «Гуаньюнь» (1007 г.).

**80) «star/звезда»:** а) {天星} GTN t<sup>h</sup>ien<sub>1</sub>=niŋ<sub>1</sub>, NND t<sup>h</sup>en<sub>1</sub>=niŋ<sub>1</sub>, HUB t<sup>h</sup>ien<sub>1</sub>=niŋ<sub>1</sub>, XNC t<sup>h</sup>iŋ<sub>1</sub>=niŋ<sub>1</sub>, JDU t<sup>h</sup>en<sub>1</sub>=niŋ<sub>1</sub>, ZHN t<sup>h</sup>ien<sub>1</sub>=niŋ<sub>1</sub>, FDN t<sup>h</sup>ien<sub>1</sub>=siŋ<sub>1</sub>, FQN (t<sup>h</sup>ien<sub>1</sub>)=siŋ<sub>1</sub>, CNL t<sup>h</sup>ien<sub>1</sub>=niŋ<sub>1</sub>, FUA t<sup>h</sup>iŋ<sub>1</sub>=neiŋ<sub>1</sub>, PIN t<sup>h</sup>ien<sub>1</sub>=niŋ<sub>1</sub>; а.1) {星} FCH θiŋ<sub>1</sub>, YNT siŋ<sub>1</sub>, SHON siŋ<sub>1</sub> || ПВМ {天星} \*t<sup>h</sup>ien<sub>1</sub>=niŋ<sub>1</sub>

КДК этимон 星 \*s<sup>h</sup>ēŋ ‘звезда’ стабильно сохраняется во всем восточном ареале, чаще всего в расширенной до бинорма форме, второй компонент которой КДК 天 \*t<sup>h</sup>iŋ ‘небо’. На стыке слов инициаль второго корня s- закономерно ассимилируется в n- (Hiroyuki 2020: 323).

**81) «stone/камень»:** а) {石} HUB θiɿ<sub>7</sub>, XNC θy<sub>7</sub>, JDU sy<sub>7</sub>, FUA t<sup>h</sup>i<sub>7</sub> (H), SHON syø<sub>7</sub> (H), PIN sø<sub>1</sub>; а.1) {石頭} FCH θuok<sub>7</sub><sup>23</sup>-t<sup>h</sup>au<sub>2</sub>, FQN syo<sub>1</sub>-t<sup>h</sup>au<sub>2</sub> (lau<sub>2</sub>) || ПВМ {石} \*siok<sub>7</sub>

КДК этимон 石 \*diak очевидно сохраняется в ВМ ветви, хотя во многих словарных источниках вхождение ‘камень’ отсутствует, что приводит к неизбежным лакунам. Иногда основа расширяется до бинорма с помощью частотного суффикса 頭 (КДК \*d<sup>h</sup>ō, исходно ‘голова’), точно так же, как это происходит в литературном языке.

В диалекте FUA отмечено нерегулярное развитие инициали (ПДК \*ʒ- → t<sup>h</sup>-); примеров, подтверждающих возможность такого развития, немного (ПДК 勺 \*ʒak ‘половник, чашка’ → SHON t<sup>h</sup>ia<sub>7</sub> ‘черпак, ковш’, PIN t<sup>h</sup>ev<sub>1</sub> ‘черпак, ковш’; ср. также ПДК 石卵 \*ʒek-l<sup>h</sup>wán ‘булыжник’ → ZHN t<sup>h</sup>yk<sub>7</sub>-lɿn<sub>5</sub> ‘булыжник’ (Hiroyuki 2020:180; 407; Lin 2002: 39), но слово все же не хотелось бы считать этимологически отличным ввиду регулярного развития остальных компонентов (финаль, тон).

**82) «sun/солнце»:** {日頭} FCH lik<sub>7</sub><sup>24</sup>-t<sup>h</sup>au<sub>2</sub>, GTN ni<sub>7</sub>-t<sup>h</sup>au<sub>2</sub>, NND nik<sub>7</sub>-t<sup>h</sup>au<sub>2</sub>, HUB nit<sub>7</sub>-t<sup>h</sup>au<sub>2</sub>, XNC ni<sub>0</sub>-au<sub>2</sub>, JDU nit<sub>7</sub>-t<sup>h</sup>au<sub>2</sub>, ZHN ne<sub>7</sub>-t<sup>h</sup>au<sub>2</sub>, FDN ni<sub>7</sub>-t<sup>h</sup>au<sub>2</sub>, FQN ni<sub>7</sub>-t<sup>h</sup>au<sub>1</sub>, CNL ni<sub>7</sub>-t<sup>h</sup>au<sub>2</sub>, YNT ni<sub>7</sub>-t<sup>h</sup>au<sub>2</sub>, FUA ni<sub>7</sub>-lau<sub>2</sub>, SHON ni<sub>1</sub>-t<sup>h</sup>au<sub>2</sub>, PIN nik<sub>7</sub>-t<sup>h</sup>au<sub>2</sub> || ПВМ {日頭} \*nit<sub>7</sub>-t<sup>h</sup>au<sub>2</sub>

Старый корень 日 \*nit сохраняется по всему миньскому континууму, однако в ВМ диалектах всегда расширяется до бинорма с помощью продуктивного суффикса (КДК 頭 \*d<sup>h</sup>ō с исходным значением ‘голова’).

**83) «swim/плавать»:** а) {泗水} FCH θiu<sub>4</sub>-tsui<sub>2</sub>, GTN siu<sub>3</sub>-tsy<sub>3</sub> (LN), NND θeu<sub>2</sub>-tsy<sub>3</sub> (LN), FDN θiu<sub>2</sub>-tsui<sub>3</sub> (LN), FQN siu<sub>2</sub>-tsui<sub>3</sub>, CNL siu<sub>3</sub>-tsui<sub>3</sub>, YNT siu<sub>2</sub>-tsui<sub>3</sub>, FUA siu<sub>2</sub><sup>25</sup>-ʒi<sub>3</sub>; а.1) {泗潭} ZHN seu<sub>2</sub>-t<sup>h</sup>an<sub>2</sub> (LN), SHON siu<sub>2</sub>-t<sup>h</sup>an<sub>2</sub>, XNC θeu<sub>2</sub>-lam<sub>2</sub><sup>26</sup> || ПВМ {泗水} \*siu<sub>2</sub>-tsui<sub>3</sub>

Старый этимон 游 \*lu не сохраняется в восточной ветви, вытесняясь инновацией 泗 (КДК \*l<sup>h</sup>u) с близкой семантикой, которая начинает встречаться уже в текстах классической эпохи (Starostin 2019: 164). Этот же корень распространен и среди миньнаньских

<sup>23</sup> Накадзима указывает регистровый тон 3, который, скорее всего, соотносится с тоном 7 («светлый входящий» 陽入).

<sup>24</sup> См. предыдущую сноску.

<sup>25</sup> Линь Ханьшэн указывает для диалекта фуань регистровый номер 21, что предположительно соотносится с тоном 2 («светлый ровный» 陽平; Lin 2002: 68).

<sup>26</sup> Лексема также употребляется в значении ‘мыться’ (Hiroyuki 2020: 356).

диалектов (Лоренц 2020: 154). Во всех случаях лексема употребляется только в двусложной форме; в подавляющем числе диалектов расширительным компонентом выступает КДК 水 *\*tʷəj* ‘вода’ (см. разбор лексемы ‘вода’), но в двух диалектах представлен альтернативный компонент — КДК 潭 *\*l̥əm* ‘пучина’, ‘бездна’.

**84) «tail/хвост»:** а) {尾} FCH *mui*<sub>2</sub>, GTN *muoi*<sub>3</sub>, NND *møy*<sub>3</sub>, HUB *mli*<sub>3</sub>, XNC *møy*<sub>3</sub>, JDU *møy*<sub>3</sub>, ZHN *muai*<sub>3</sub>, FDN *muei*<sub>3</sub>, YNT *muoi*<sub>3</sub>, FUA *mui*<sub>3</sub>, PIN *moi*<sub>3</sub>; а.1) {尾溜 ~ 尾稍} FQN *mui*<sub>3</sub>-*lieu*<sub>4</sub>, SHON *muoi*<sub>4</sub>-*liu*<sub>1</sub>; а.2) {尾尾} CNL *mui*<sub>3</sub>-*mui*<sub>3</sub> || ПВМ {尾} *\*mui*<sub>3</sub>

**85) «that/тот»:** а) {許} CNL *hi*<sub>3</sub>, YNT *hy*<sub>3</sub>, SHON *хай*<sub>4</sub><sup>27</sup>, PIN *ха*<sub>3</sub>; а.1) {許隻 ~ 彼個} FCH *hi*<sub>2</sub>-*eik*<sub>6</sub><sup>28</sup>, GTN *xia*<sub>6</sub>-*ie?*<sub>6</sub>, HUB *ха*<sub>3</sub>-*lie?*<sub>6</sub>~*xei*<sub>3</sub>, XNC *ha*<sub>5</sub>-*cie?*<sub>6</sub>, JDU *ha*<sub>3</sub>-*cie?*<sub>6</sub>, FUA *heik*<sub>6</sub>; а.2) {許蜀隻 ~ 許一隻} NND *ха?*<sub>6</sub>-*søk*<sub>7</sub>-*ie?*<sub>6</sub>, FQN *hy*<sub>3</sub>-*syo*<sub>1</sub>-*tsia*<sub>4</sub>; а.3) {許隻} ZHN *хай*<sub>4</sub>-*it*<sub>6</sub>; а.4) {許個 ~ 許隻} FDN *xi*<sub>3</sub>-*koi*<sub>4</sub> || ПВМ {許隻} *\*hi*<sub>3</sub>-*tsiek*<sub>6</sub>

КДК этимон 彼 *\*páj* ‘тот’ вытесняется уже на общеминьском уровне ПДК диалектной инновацией 許 *\*hó* (Starostin 2019: 171; Лоренц 2020: 154–155), которая была характерна для ареала к востоку от Янцзы в эпоху Шести династий (220–589 гг.; Mei 1999: 8). Иногда при этом лексема продолжает записываться знаком 彼, оставленным для нее из-за своей исходной местоименной семантики.

Данные показывают, что чаще всего корень расширяется за счет счетного слова, восходящего к КДК лексеме 隻 *\*tek* ‘один (из пары)’. Иногда оно же (по аналогии с литературным языком) записывается другим графическим знаком — 個 (КДК *\*kājh* ‘штука’). На стыке слов инициаль счетного слова, как правило, выпадает, хотя отдельные фонетически архаичные диалекты, такие, как CNL и YNT (см. доп. примеры в Hiroyuki 2018: 374), ее сохраняют. Встречаются также трехсложные сочетания (а.2), где счетному слову ‘один из пары’ предшествует числительное ‘один’ (КДК 蜀 *\*zok*, см. выше ‘one/один’).

**86) «this/этот»:** а) {這個 ~ 這隻 ~ 此隻} FCH *tsi*<sub>2</sub>-*eik*<sub>6</sub> (*tsui*<sub>2</sub>), GTN *tsia*<sub>6</sub>-*ie?*<sub>6</sub>, FUA *tseik*<sub>6</sub>, HUB *tsa*<sub>3</sub>-*lie?*<sub>6</sub>~*tfei*<sub>3</sub>, XNC *tfa*<sub>3</sub>-*cie?*<sub>5</sub>, JDU *tsa*<sub>3</sub>-*cie?*<sub>6</sub>; а.1) {此 ~ 這} CNL *tsi*<sub>3</sub>, YNT *tsi*<sub>3</sub>, SHON *tsia*<sub>4</sub>, PIN *tse*<sub>3</sub>; а.2) {這蜀隻 ~ 者一隻} NND *tsa?*<sub>6</sub>-*søk*<sub>6</sub>-*ie?*<sub>6</sub>, FQN *tsie*<sub>3</sub>-*syo*<sub>1</sub>-*zja*<sub>5</sub>; а.3) {這一} ZHN *tsai*<sub>4</sub>-*it*<sub>6</sub>; а.4) {這個} FDN *tsi*<sub>3</sub>-*koi*<sub>4</sub> || ПВМ {這隻} *\*tsi*<sub>3</sub>-*tsiek*<sub>6</sub>

Представленные во всем ВМ ареале формы очевидно не могут восходить к КДК этимону 此 *\*c<sup>h</sup>é* (основная лексема для выражения ближнего дейксиса), но, вероятно, связаны с СК этимоном 𠄎 ‘этот’, фиксируемом в требуемом значении начиная с эпохи Пяти династий (Лоренц 2020: 155; Starostin 2019: 169). В поздне-СК это местоимение записывается знаком 這 (специальный графический вариант, пришедший на смену более раннему 者); в миньской транскрипции чаще всего используется он же, но может также встречаться как 者, так и 此 (последний — исключительно из-за семантической близости).

Местоимение часто фиксируется в биномиальной форме; наиболее широко распространено в качестве второго компонента счетное слово 隻 (КДК *\*tek*) ‘один (из пары)’. На стыке морфем инициаль второго слога в ряде диалектов апокопируется; иногда наблюдается полное стяжение в однослог, как в диалектах FUA и HUB. В единственном случае замечен вариант с альтернативным счетным словом а.4) 個 (КДК *\*kājh* ‘штука’).

В диалектах NND, FQN (а.3) и ZHN (а.4) обнаружено две расширяющих морфемы с семантикой ‘один’, разных по своему происхождению: в первом случае используется 蜀 (КДК *\*zok*; см. разбор лексемы ‘один’), во втором — лексема, восходящая к КДК — *\*?it* ‘один’.

<sup>27</sup> А. Хироюки в качестве синонима для шоуиньского диалекта указывает лексему 註 *tsyø*<sub>4</sub>; ее происхождение и семантическая специфика остаются неясными (Hiroyuki 2020: 683).

<sup>28</sup> М. Накадзима также фиксирует синонимичную форму *hui*<sub>2</sub>, которая, по всей видимости, представляет собой стяжение из 許回 *tsi*<sub>4</sub> *uoi*<sub>2</sub> (Nakajima 1979: 262; Liu & He 1998: 548).

87) «**thou/ты**»: {汝} FCH nø<sub>3</sub>, GTN ny<sub>3</sub>, NND ny<sub>3</sub>, HUB ny<sub>3</sub>, XNC ny<sub>3</sub>, JDU ny<sub>3</sub>, ZHN ny<sub>3</sub>, FDN ni<sub>3</sub>, FQN ny<sub>3</sub>, CNL ly<sub>3</sub>, YNT ny<sub>3</sub>, FUA ni<sub>3</sub>, SHON ny<sub>3</sub>, PIN ny<sub>3</sub> || ПВМ {汝} \*ny<sub>3</sub>

КДК этимон 汝 \*n<sup>h</sup>á стабильно сохраняется в ВМ ветви. Этот же корень широко распространен в южных и цюньвэньских диалектах (Лоренц 2020: 156).

88) «**tongue/язык**»: а) {嘴舌} FCH ts<sup>h</sup>ui<sub>4</sub>=liek<sub>7</sub>, HUB ts<sup>h</sup>ui<sub>4</sub>=let<sub>6</sub>, XNC t<sup>h</sup>ui<sub>4</sub>=lit<sub>6</sub>, JDU ɕy<sub>4</sub>=let<sub>5</sub>, FQN ts<sup>h</sup>ui<sub>4</sub>=(s)lie<sub>7</sub>, CNL ts<sup>h</sup>ui<sub>1</sub>=liek<sub>7</sub>, YNT ts<sup>h</sup>uoi<sub>1</sub>=lie<sub>7</sub>, FUA ts<sup>h</sup>i<sub>4</sub>=lik<sub>7</sub>, SHON ts<sup>h</sup>y<sub>4</sub>=tsie<sub>7</sub>, PIN ts<sup>h</sup>ui<sub>4</sub>=zeik<sub>7</sub>; а.1) {舌} GTN sie<sub>7</sub>, NND set<sub>7</sub>, ZHN sik<sub>7</sub>, FDN sie<sub>7</sub> || ПВМ {嘴舌} \*ts<sup>h</sup>ui<sub>4</sub>-liet<sub>7</sub>

КДК этимон 舌 \*lat ‘язык’ стабильно сохраняется по всему миньскому континууму, однако чаще всего встречается в форме а), расширенной до биннома с помощью компонента 喙 (КДК \*ʰwas) ‘рот’ (см. ‘mouth/рот’ выше; Лоренц 2020: 156).

89) «**tooth/зуб**»: а) {喙齒 ~ 喙齒} GTN ts<sup>h</sup>y<sub>4</sub>=i<sub>3</sub>, NND ts<sup>h</sup>i<sub>4</sub>=i<sub>3</sub>, HUB (ts<sup>h</sup>y) ts<sup>h</sup>ui<sub>4</sub>=i<sub>3</sub>, XNC ɕy<sub>4</sub>=ei<sub>3</sub>, JDU ɕy<sub>4</sub>=i<sub>3</sub>, ZHN ts<sup>h</sup>i<sub>4</sub>=i<sub>3</sub>, FQN ts<sup>h</sup>ui<sub>4</sub>=k<sup>h</sup>i<sub>3</sub>, FUA ts<sup>h</sup>i<sub>4</sub>=i<sub>3</sub>, SHON ts<sup>h</sup>y<sub>4</sub>=k<sup>h</sup>i<sub>3</sub>, PIN ts<sup>h</sup>u<sub>4</sub>=i<sub>3</sub>; б) {牙齒} FDN ɲa<sub>2</sub>-k<sup>h</sup>i<sub>3</sub>, FCH ɲai<sub>3</sub>, CNL ɲai<sub>3</sub>, YNT ɲai<sub>3</sub>, FQN ɲai<sub>3</sub> || ПВМ {喙齒} \*ts<sup>h</sup>ui<sub>4</sub>-k<sup>h</sup>i<sub>3</sub>

КДК этимон 齒 \*t<sup>h</sup>ǎ ~ \*k<sup>h</sup>ǎ<sup>29</sup> ‘зуб’ (применяется как к людям, так и к животным; Starostin 2019: 169) представлен во всех ВМ диалектах, но только в составе вариантов, расширенных до двусложного состояния. Наиболее частотным расширителем оказывается 喙 (КДК \*ʰwas) ‘рот’ (см. выше ‘mouth/рот’); при этом выпадение глухой придыхательной инициали k<sup>h</sup>- во втором слоге оказывается регулярным в случае, если финаль первого слога оканчивается на гласный (Hiroyuki 2018: 29).

В нескольких диалектах см. (б) КДК 齒 \*k<sup>h</sup>ǎ также образует бинном с КДК 牙 \*ɲrā ‘зуб животного’ (слово было исходно заимствовано в КДК из австроазиатских языков с исходным значением ‘бивень’, ‘зуб животного’, ‘клык’ в ходе торговых отношений и закупок изделий из кости северным Китаем, см. Starostin 2019: 169; Norman & Mei 2000: 475). В четырех диалектах эти два корня стягиваются в однослог ɲai<sub>3</sub>, на что однозначно указывает терминаль -i (Norman & Mei 2000: 480). Мы предполагаем, что в таком стяжении «лексический» акцент смещен на основу 牙 \*ɲrā.

90) «**tree/дерево**»: {樹} FCH ts<sup>h</sup>iu<sub>5</sub>, GTN ts<sup>h</sup>iu<sub>5</sub>, NND ts<sup>h</sup>eu<sub>5</sub>, HUB t<sup>h</sup>iu<sub>5</sub>, XNC ɕiu<sub>5</sub>, JDU t<sup>h</sup>eu<sub>5</sub>, ZHN ts<sup>h</sup>eu<sub>5</sub>, FDN ts<sup>h</sup>iu<sub>5</sub>, FQN ts<sup>h</sup>iu<sub>5</sub>#, CNL ts<sup>h</sup>iu<sub>1</sub>#, YNT ts<sup>h</sup>ieu<sub>1</sub>#, FUA ts<sup>h</sup>iu<sub>4</sub>#, SHON ts<sup>h</sup>iu<sub>1</sub>#, PIN ts<sup>h</sup>iu<sub>5</sub> || ПВМ {樹} \*ts<sup>h</sup>iu<sub>5</sub>

91) «**two/два**»: {兩} FCH naŋ<sub>5</sub>, GTN laŋ<sub>5</sub>#, NND laŋ<sub>5</sub>#, HUB laŋ<sub>5</sub>, XNC laŋ<sub>5</sub>, JDU laŋ<sub>5</sub>, ZHN laŋ<sub>5</sub>#, FDN laŋ<sub>5</sub>#, FQN laŋ<sub>5</sub>#, CNL laŋ<sub>6</sub># (LH), YNT laŋ<sub>6</sub># (LH), FUA laŋ<sub>6</sub># (LH), SHON laŋ<sub>6</sub># (LH), PIN laŋ<sub>5</sub> || ПВМ {兩} \*laŋ<sub>5</sub>

92) «**walk (go)/идти**»: {去} FCH k<sup>h</sup>o<sub>4</sub>, FQN k<sup>h</sup>ie<sub>4</sub>#, HUB k<sup>h</sup>yø<sub>4</sub>, XNC k<sup>h</sup>yø<sub>4</sub>, JDU k<sup>h</sup>io<sub>4</sub>, FUA k<sup>h</sup>ø<sub>4</sub> (H), SHON k<sup>h</sup>yø<sub>4</sub> (H), PIN k<sup>h</sup>œ<sub>4</sub> || ПВМ {去} \*k<sup>h</sup>ø<sub>4</sub>

93) «**warm (hot)/горячий**»: а) {熱(農)} FCH iek<sub>7</sub>, GTN ie<sub>7</sub>, NND et<sub>7</sub>, HUB et<sub>7</sub>, XNC it<sub>7</sub>, JDU et<sub>7</sub>, ZHN i<sub>7</sub>, FDN ie<sub>7</sub>, FQN ie<sub>7</sub>, SHON ie<sub>7</sub> (H), FUA i<sub>7</sub>-(nœuŋ<sub>0</sub>) (H), PIN eik<sub>7</sub>-(nœuŋ<sub>0</sub>); б) {•農~•暖} HUB kɔŋ<sub>2</sub>, JDU kɔŋ<sub>2</sub>, XNC kɔŋ<sub>2</sub>, FUA kɔuŋ<sub>2</sub>-(nœuŋ<sub>0</sub>) (H), PIN kɔuŋ<sub>2</sub>-(nœuŋ<sub>0</sub>) || ПВМ {熱} \*iet<sub>7</sub> и ПВМ \*kɔŋ<sub>2</sub>

В миньдунских диалектах существует два разных корня для передачи значения ‘горячий’; как правило, унаследованный от КДК этимон 熱 \*ɲet используется в сочетании с погодой, а альтернативный корень \*kɔŋ<sub>2</sub> (с неясной этимологией) обычно характеризует ‘горячие’ объекты. В данной ситуации мы выводим на прауровень оба корня в качестве «технических» синонимов.

<sup>29</sup> ДК реконструкция инициали неоднозначна; миньские диалекты стабильно отражают вариант с заднеязычным согласным.

Отметим, что внутри ВМ наблюдается четкое лексическое противопоставление между значениями ‘теплый’ и ‘горячий’: в значении ‘теплый’ употребляются формы, восходящие к ҚДК 暖 \*nōn (SHON nɔŋ<sub>3</sub> (H) ~ t<sup>h</sup>ɔŋ<sub>7-n</sub>æŋ<sub>0</sub> (H), HUB nɔŋ<sub>3</sub>, XNC nɔŋ<sub>3</sub>, JDU nɔŋ<sub>3</sub>, FUA nɔŋ<sub>3</sub> (H), PIN nouŋ<sub>3</sub> (H)). Любопытно, что эта же основа в некоторых диалектах может служить расширителем в составе бинома ‘горячий’ (т. е. ‘горячий’ = ‘горячий-теплый’).

94) «water/вода»: {水} FCH tsui<sub>3</sub>, GTN tsui<sub>3</sub>, NND tsui<sub>3</sub>, HUB tɕy<sub>5</sub>, XNC tɕy<sub>3</sub>, JDU tɕy<sub>5</sub>, ZHN tsui<sub>3</sub>, FDN tsy<sub>3</sub>, FQN tsui<sub>3</sub>, CNL tsui<sub>3</sub>#, YNT tsuoi<sub>3</sub>#, FUA tsi<sub>3</sub>, SHON tsy<sub>3</sub>, PIN tsuoi<sub>3</sub> || ПВМ {水} \*tsui<sub>3</sub>

95) «we/мы»: а) {我各儂 ~ 我各人} FCH ŋuai<sub>1</sub>-ko<sub>3</sub>-nøŋ<sub>2</sub><sup>30</sup>, GTN ŋuai<sub>3</sub>-ɔu<sub>6</sub>-nøŋ<sub>2</sub>, NND ua<sub>3</sub>-kɔ<sub>6</sub>-nøŋ<sub>2</sub>, HUB ua<sub>3</sub>-kœ<sub>6</sub>-nœu<sub>3</sub>, FQN ŋua<sub>3</sub>-ko<sub>4</sub>-nøŋ<sub>2</sub>, CNL ŋuoi<sub>3</sub>-o<sub>6</sub>-lœu<sub>2</sub>, PIN uai<sub>3</sub>-u<sub>6</sub>-nœu<sub>2</sub>; а.1) {我儕} ZHN uε<sub>3</sub>-lε<sub>2</sub>, FUA ŋo<sub>1</sub>-ε<sub>0</sub>, PIN uai<sub>3</sub>-(z)ε<sub>2</sub>; а.2) {我儕儂} HUB ua<sub>3</sub>-tɕε<sub>2</sub>-nœu<sub>2</sub>, XNC ua<sub>3</sub>-tɕε<sub>2</sub>-nœu<sub>2</sub>, JDU ua<sub>3</sub>-tsε<sub>2</sub>-nœu<sub>0</sub>; а.3) {我儂} FDN uε<sub>3</sub>-neŋ<sub>2</sub>; а.4) {儂家} FQN nøŋ<sub>2</sub>-(k-)ŋa<sub>1</sub>; а.4.1) {儂家依} YNT naŋ<sub>2</sub>-ŋa<sub>1</sub>-nøŋ<sub>2</sub>; а.5) {我儕家儂} SHON ua<sub>3</sub>-tsi<sub>2</sub>-ka<sub>4</sub>-neŋ<sub>2</sub> || ПВМ {我各儂} \*ŋuai<sub>3</sub>-kɔ<sub>6</sub>-nœu<sub>2</sub> и {我儕} \*ŋuai<sub>3</sub>-tsε<sub>2</sub>

В классических текстах этимон 我 \*ŋ<sup>h</sup>áj совмещал в себе семантику единственного и множественного числа. В миньдунских диалектах эта морфема сохраняется, однако при передаче значения ‘мы’ она всегда расширяется за счет различных суффиксальных элементов.

Шире всего в этой функции используется морфемная последовательность а) 各儂 (ҚДК \*kāk-nūŋ), букв. ‘каждый человек’; реже встречается вариант а.3) без морфемы 各 \*kāk ‘каждый’. Еще одна возможная суффиксальная морфема — известный с ханьского времени маркер множественного числа 儕 (ҚДК \*zrāj) с исходной семантикой ‘группа, категория, сверстники, товарищи’, инициаль которой ассимилируется на стыке основ. В диалектах HUB, XNC, JDU (см. а.2) этот суффикс вступает в сочетание с вышеупомянутой морфемой 儂 \*nūŋ ‘человек’.

Наконец, еще одним показателем множественности (а.4) может служить морфема 家 (ҚДК \*krā) ‘дом, семья’, которая также может сочетаться с 儂 \*nūŋ ‘человек’ (а.4.1), (а.5).

Очевидно, что на ПВМ уровне значение ‘мы’ выражалось той же корневой морфемой, что и ‘я’, но однозначно восстановить базовый для этого хронологического этапа суффиксальный расширитель не удастся. Как «технические синонимы» на ПВМ уровень можно вывести две лексемы: {我各儂} \*ŋuai<sub>3</sub>-kɔ<sub>6</sub>-nœu<sub>2</sub> и {我儕} \*ŋuai<sub>3</sub>-tsε<sub>2</sub>.

96) «what/что»: а) {甚毛 ~ 什毛} FCH θiek<sup>31</sup>-nok<sub>6</sub>, HUB θiŋ<sub>7</sub>-nɔ<sub>5</sub>, XNC (ɕit<sub>5</sub>)-nɔ<sub>6</sub>, JDU (ɕit<sub>6</sub>)-nɔ<sub>7</sub>; а.1) {乜毛 ~ 麼毛 ~ 呢毛 ~ 若毛} ZHN mi<sub>6</sub>-nɔ<sub>6</sub>, FQN mie<sub>4</sub>-nɔ<sub>5</sub>, FUA mi<sub>1</sub>-nɔk<sub>6</sub>, SHON ni<sub>1</sub>-nɔ<sub>4</sub>; а.2) {哪毛 ~ 毛毛} NND nɔ<sub>7</sub>-nɔ<sub>6</sub>, HUB nɔ<sub>6</sub>-nɔ<sub>4</sub>-nɔ<sub>6</sub>-nɔ<sub>0</sub>; FDN nɔ<sub>5</sub>-nɔ<sub>4</sub>, YNT nɔ<sub>5</sub>-nɔ<sub>6</sub>, PIN (nɔ<sub>1</sub>)-nɔ<sub>4</sub>; а.3) {甚毛 ~ 什毛} FCH θien<sub>2</sub>-nok<sub>6</sub>, CNL sien<sub>2</sub>-nɔ<sub>6</sub>; а.4) {甚毛} GTN sia<sub>4</sub>-nɔ<sub>6</sub><sup>32</sup> || ПВМ {乜毛} \*mi<sub>1</sub>-nɔ<sub>6</sub>

ҚДК этимон 何 \*g<sup>h</sup>āj был утрачен еще на общеминьском уровне, так как не сохранился ни в одной из ветвей современных языков-потомков (Лоренц 2020: 158). Вместо него в ВМ используются два бинома, происхождение и развитие каждого из которых достаточно проблематично.

Первый компонент бинома 什毛 (а), географически ограниченного центральным ареалом провинции Фуцзянь, несмотря на иероглифическую транскрипцию, невоз-

<sup>30</sup> По данным материалов Chen & Li (1991: 114) в фучжоуском диалекте используется форма 儂家 naŋ<sub>2</sub>-ŋa<sub>1</sub>, которую Накадзима (Nakajima 1979: 261) считает эксклюзивной.

<sup>31</sup> М. Накадзима указывает регистровый тон 3, который, вероятно, соотносится с тоном 7 («темный входящий» 陽入; Nakajima 1979: 262).

<sup>32</sup> В двух альтернативных источниках для диалекта гутянь указывается односложный вариант (哪)毛 nɔ<sub>6</sub> (Lin 2002: 75; Liu & He 1998: 548).

можно этимологически отождествлять с первым компонентом хорошо известной СК вопросительной лексики 什麼 *zim-mwá* ‘что?’ (впервые зафиксирована в произведении 壇經 «Сутра помоста шестого патриарха», VII–VIII вв. н.э; см. NYDC I: 1101) из-за наличия в нем финали \*-t. (Особо следует отметить форму *sia*- в диалекте GTN, которая из-за своей тональной характеристики не может быть этимологически тождественной корню \*sit, но при этом также не имеет понятной этимологии.) Однако именно этот компонент, скорее всего, представлен в формах группы а.3; допустимо его проникновение в диалекты FCH и CNL уже из литературного языка.

Второй компонент этого же бинорма 乜 \*nɔ̌<sub>6</sub> ‘вещь’ (условно записывается знаком ‘листик травы’, отобранным по принципу фонетической созвучности) не имеет исконно китайской этимологии. Не исключено его заимствование из австроазиатских языков, ср. такие фонетически близкие формы в бахнарической ветви, как бунор *na:w*, центральный роллом *nɔ*: (Shorto 1995: 553).

Сравнительно часто также встречается бинорм 乜乜, в котором первая лексема 乜 (СК *má*) может быть усеченной формой старого КДК корня 物 \*m<sup>h</sup>ət ‘вещь’, которая сама берет на себя функции вопросительного местоимения. Для записи этой морфемы используется целый ряд знаков: (а) 麼 (СК *mwa*, суффикс вопросительных и указательных местоимений); (б) 呢 (СК *ni*, конечная частица); (в) 若 (СК *nak* ‘такой’, ‘быть похожим на’), см. NYDC III: 312; IX: 328.

Наконец, иногда вопросительное местоимение образуется просто из редупликации лексемы 乜 ‘вещь’ (а.2).

Таким образом, объединяющим все перечисленные варианты элементом оказывается морфема второго слога \*nɔ̌<sub>6</sub> ‘вещь’, в то время как первый слог варьирует от диалекта к диалекту. Учитывая, что двуслог 乜乜 мог распространиться из-за влияния литературного языка, логично реконструировать на ПВМ уровне форму {乜乜} \*mi<sub>1</sub>-nɔ̌<sub>6</sub>.

**97) «white/белый»:** {白} FCH pak<sub>7</sub>, GTN paʔ<sub>7</sub>, NND paʔ<sub>7</sub>, HUB paʔ<sub>7</sub>, XNC paʔ<sub>7</sub>, JDU pa<sub>7</sub>, ZHN paʔ<sub>7</sub>, FDN pa<sub>7</sub>, FQN pa<sub>1</sub>#, CNL pa<sub>1</sub>#, YNT pa<sub>7</sub>#, FUA pa<sub>7</sub>#, SHON paʔ<sub>7</sub>#, PIN pa<sub>1</sub> || ПВМ {白} \*pak<sub>7</sub>

**98) «who/кто»:** а) {底儂 ~ 底人} FCH tak<sub>7</sub>-nøŋ<sub>2</sub>, GTN tøyŋ<sub>1</sub>, NND ti<sub>5</sub>-nøŋ<sub>2</sub>, ZHN tœŋ<sub>3</sub>, FQN tie<sub>1</sub>-nøŋ<sub>2</sub>, CNL tø<sub>1</sub>-løŋ<sub>2</sub>, YNT ti<sub>1</sub>-nøŋ<sub>2</sub>, FUA tœŋ<sub>3</sub>; б) {乜儂 ~ 哪儂} FDN no<sub>7</sub>-neŋ<sub>4</sub>, HUB nœŋ<sub>4</sub><sup>33</sup>, XNC nœŋ<sub>4</sub><sup>34</sup>, JDU nɔ̌<sub>6</sub>-nœŋ<sub>0</sub>~nœŋ<sub>4</sub>, PIN nœŋ<sub>3</sub>; б.1) {呢乜儂} SHON ni<sub>1</sub>-nɔ̌<sub>6</sub>-neŋ<sub>2</sub> || ПВМ {底儂} \*ti<sub>1</sub>-nœŋ<sub>2</sub> и ПВМ {乜儂} \*nɔ̌<sub>6</sub>-nœŋ<sub>2</sub>

КДК этимон 誰 \*dwəj ‘кто?’ (Starostin 2019: 164), не представленный в ВМ диалектах, заменяется на двуслог 底人 (СК *tiej-nōŋ*). Первый из двух компонентов здесь — вопросительное слово 底 (СК *tiej*), впервые зафиксированное в литературных памятниках начиная с эпохи Тан, а второй — уже известная нам лексема 儂 КДК \*nīŋ ‘человек’. В ряде диалектов двуслог стягивается в односложный вариант.

В другом композите б), который также сравнительно часто используется в миньдунских диалектах, выделяется первый компонент 乜 \*nɔ̌<sub>6</sub> ‘вещь’ (см. выше ‘what/что’). Бинорм 乜儂 также может стягиваться в однослог; помимо этого, в диалекте SHON наблюдается трехсложная форма, расширенная за счет вопросительного слова 乜 (СК *má*, см. выше ‘what/что’).

Исходя из связи между лексемами ‘who/кто’ и ‘what/что’, мы предполагаем, что формы, маркированные индексом (b), скорее всего появились под влиянием лексем

<sup>33</sup> Для диалекта хубэй в данном слове Хироюки указывает регистровый тон 52, который, возможно, следует относить к тональной категории 4 («темный падающий» 陰去).

<sup>34</sup> Хироюки указывает регистровый тон 53, который, вероятно, соотносится с тоном 4 («темный падающий», 陰去).



‘what/что’. Поскольку формы (a) {底儂}  $*ti_1-n\alpha\eta_2$  и (b) {乜儂}  $*n\alpha_6-n\alpha\eta_2$  имеют смешанную дистрибуцию, их можно вывести на прауровень в качестве технических синонимов; однако стоит учесть, что по крайней мере теоретически формы из группы (b) могли развиться вторично по аналогии с соответствующими неодушевленными местоимениями.

**99) «woman/женщина»:** a) {作母 ~ 晉母} NND  $ts\alpha m_6=mu_3$ , HUB  $t\dot{f}im_4=mu\alpha_3$ , XNC  $t\dot{c}in_4=mu_3$ , JDU  $t\dot{c}im_4=mu_3$ ; a.1) {做母儂 ~ 諸母儂} FDN  $ts\alpha\eta_5=mu\alpha_1-n\eta_2$ , SHON  $tsy\alpha_1=mu\alpha_3-n\eta_2$ ; b) {婦女儂 ~ 婦女人} GTN  $xu_5=ny_3-n\alpha\eta_2$ , ZHN  $xu_5=ny_3-n\alpha\eta_2$ , PIN  $xu_1=ny_3-n\alpha\eta_3$ ; b.1) {女子} FCH  $ny_6-y_3^{35}$ ; b.2) {女界} FCH  $ny_1-ai_4^{36}$ , CNL  $ny_1-ai_4$ ; c) {諸娘儂} FQN  $tsy_1=n\alpha\eta_2-n\alpha\eta_2$ , YNT  $tsy_1=ny\alpha\eta_2-ny\alpha\eta_2$ ; c.1) {諸娘} FUA  $tsi_1=n\alpha\eta_2$  || ПВМ {諸娘儂}  $tsy_1=ny\alpha\eta_2-n\alpha\eta_2$  и ПВМ {婦女儂}  $*hu_5=ny_3-n\alpha\eta_2$

Наши данные отражают три разных корня на территории распространения ВМ диалектов, включая архаичный КДК корень 女  $*nr\acute{a}$  ‘женщина’ (b), который, как правило, употребляется в составе композитов, причем чаще всего в связке с префигированным компонентом 婦  $*b\acute{a}$  (исходно — ‘жена, супруга’). В качестве третьего компонента обычно выступает КДК 儂  $*n\dot{i}\eta$  ‘человек’, но иногда вместо него используется десемантизированный суффикс 子 (КДК  $*c\acute{a}$ ) ‘сын, ребенок’ и лексема 界 (КДК  $*kr\acute{e}\acute{c}$ ) ‘граница → группа, класс’, где в обоих случаях происходит ассимиляция на стыке морфем. Корень преимущественно употребляется в центральной и северной частях провинции Фуцзянь.

Ареал распространения КДК корня 母  $*m\acute{a}$  ‘мать’, с достаточно логичным семантическим сдвигом (‘мать’ → ‘женщина’), ограничен преимущественно севером провинции Фуцзянь. Префикс, образующий композит с данным корнем, наблюдается, кроме восточной, также в *пусяньской* и в *южной* ветвях; как правило, он образует термины родства и гендерные лексемы. Согласно мнению Дун Чжунсы, префикс  $t(s)a-$  является реликтом языка древнего государства Минь-юэ и имеет корреляты в диалектах мяо-яо, чжуан и др., однако достоверно отследить историю его появления в миньских диалектах затруднительно (Dong 1993: 102). Единого, строго фиксированного иероглифического знака для этого префикса не существует; все эквиваленты подобраны по принципу фонетического сходства. Суффиксом в данных композитах, как и в случае с КДК 女  $*nr\acute{a}$  ‘женщина’, выступает лексема 儂  $*n\dot{i}\eta$  ‘человек’.

Наконец, третий корень (c) родственен СК основе 娘  $n\alpha\eta$  ‘девушка’, зафиксированной начиная с эпохи Тан; он также может образовывать композит с реликтовым префиксом и с суффиксом 儂  $*n\dot{i}\eta$  ‘человек’. Согласно дополнительным данным по источнику Liu & He 1998 эта основа имеет самую широкую дистрибуцию в ВМ ареале, будучи также представлена в следующих диалектах: {諸娘儂} лянцзян  $tsy_1=ny\alpha\eta_2-n\alpha\eta_2$ , миньцин  $tsy_1=ny\alpha\eta_2-n\alpha\eta_2$ , гутянь  $tsy_1=ny\alpha\eta_2-n\alpha\eta_2$ , пиннань  $tsy_1=ny\alpha\eta_2-n\alpha\eta_2$ , лоюань  $tsy_1=ny\alpha\eta_2-n\alpha\eta_2$  (Liu & He 1998: 512). Видно, что данные по диалекту гутянь и пиннань расходятся с теми, которые представлены в наших основных источниках.

Поскольку дистрибуция различных этимонов по географическим ареалам и ветвям носит достаточно хаотичный характер, на ПВМ уровень в качестве «технических» синонимов следует выносить трехсложные основы 婦女儂  $*hu_5=ny_3-n\alpha\eta_2$  и 諸娘儂  $tsy_1=ny\alpha\eta_2-n\alpha\eta_2$ .

<sup>35</sup> Данная форма (предположительно) появилась в фучжоуском диалекте под влиянием литературного языка.

<sup>36</sup> Ли Жулун в своем словаре по фучжоускому диалекту указывает также синонимичный вариант 諸娘儂  $tsy_1=n\alpha\eta_2-n\alpha\eta_2$  с полисемичным значением ‘женщина, жена’ (Li & Chen 1994: 334). Возможно, это указывает на существование двух конкурирующих основ.

**100) «yellow/желтый»:** {黃} FCH uoŋ<sub>2</sub>, GTN uoŋ<sub>2</sub>, NND uŋ<sub>2</sub>, HUB uoŋ<sub>2</sub>, XNC uŋ<sub>2</sub>, JDU uŋ<sub>2</sub>, ZHN uŋ<sub>2</sub>, FDN uoŋ<sub>2</sub>, FQN uoŋ<sub>2</sub>#, CNL uoŋ<sub>1</sub>#, YNT uoŋ<sub>1</sub>#, FUA wuŋ<sub>2</sub> (H), SHON uoŋ<sub>2</sub>#, PIN ouŋ<sub>2</sub> || ПБМ {黃} \*uoŋ<sub>2</sub>

**101) «far/далеко»:** {遠} FCH huŋ<sub>5</sub>, GTN xuŋ<sub>5</sub>, NND xun<sub>5</sub>, HUB xuŋ<sub>5</sub>, XNC hun<sub>5</sub>, JDU hun<sub>5</sub>, ZHN xun<sub>5</sub>, FDN xuŋ<sub>5</sub>, FQN huŋ<sub>5</sub>, FUA huŋ<sub>5</sub> (H), SHON xuŋ<sub>5</sub> (H), PIN xouŋ<sub>5</sub> || ПБМ {遠} \*huŋ<sub>5</sub>

**102) «heavy/тяжелый»:** {重} FCH tøyŋ<sub>5</sub>, GTN tøyŋ<sub>5</sub>, NND tøyŋ<sub>5</sub>, HUB tøyŋ<sub>5</sub>, XNC tøyŋ<sub>5</sub>, JDU tøyŋ<sub>5</sub>, ZHN tøyŋ<sub>5</sub>, FDN tøyŋ<sub>5</sub>, FQN tøyŋ<sub>5</sub> (H), FUA tøyŋ<sub>5</sub> (H), SHON tøyŋ<sub>5</sub> (H), PIN tøyŋ<sub>5</sub> || ПБМ {重} \*tøyŋ<sub>5</sub>

**103) «near/близко»:** {近} FCH køyŋ<sub>5</sub>, HUB køyŋ<sub>5</sub>, XNC køyŋ<sub>5</sub>, JDU køyŋ<sub>5</sub>, FQN køyŋ<sub>5</sub>, FUA køyŋ<sub>5</sub> (H), SHON kyŋ<sub>5</sub> (H), PIN køyŋ<sub>5</sub> || ПБМ {近} \*køyŋ<sub>5</sub>

**104) «salt/соль»:** {鹽}<sup>37</sup> FCH sieŋ<sub>5</sub> (LH)<sup>38</sup>, GTN sieŋ<sub>5</sub>#, NND sim<sub>5</sub>, HUB θiem<sub>2</sub>, XNC θim<sub>2</sub>, JDU sim<sub>2</sub>, ZHN sim<sub>5</sub>, FDN sieŋ<sub>5</sub>, FQN sieŋ<sub>2</sub>, CNL sieŋ<sub>5</sub> (LH), YNT sieŋ<sub>5</sub> (LH), FUA sieŋ<sub>5</sub> (LH), SHON sieŋ<sub>5</sub> (LH), PIN sieŋ<sub>5</sub> (LH) || ПБМ {鹽} \*sieŋ<sub>5</sub>

**105) «short/короткий»:** {短} FCH tøy<sub>3</sub>, GTN toi<sub>3</sub>, NND tøy<sub>3</sub>, HUB tai<sub>3</sub>, XNC tøy<sub>3</sub>, JDU tøy<sub>3</sub>, ZHN tai<sub>3</sub>, FDN toi<sub>3</sub>, FQN toi<sub>3</sub>#, FUA tai<sub>3</sub>, SHON tai<sub>3</sub>, PIN tai<sub>3</sub> || ПБМ {短} \*toi<sub>3</sub>

Здесь мы постулируем лексическую инновацию уже на общеминьском уровне, т.к. корень не восходит к КДК этимону 短 \*twán (примеров на отражение КДК финали \*-wán как \*-oi, очевидно, нет).

**106) «snake/змея»:** a) {老蛇} HUB lau<sub>2</sub>=lie<sub>2</sub>, XNC lau<sub>1</sub>=lie<sub>2</sub>, JDU lau<sub>2</sub>=lie<sub>2</sub>, FQN lo<sub>2</sub>=sia<sub>1</sub> (zia<sub>1</sub>), CNL lau<sub>1</sub>=lie<sub>2</sub>, YNT lau<sub>1</sub>=lie<sub>2</sub>, FUA lau<sub>1</sub>=e<sub>2</sub>, SHON lau<sub>1</sub>=sia<sub>2</sub>, PIN lau<sub>2</sub><sup>39</sup>=se<sub>2</sub>; a.1) {蛇} FCH θie<sub>2</sub>, GTN sie<sub>2</sub>, NND sie<sub>2</sub>, ZHN sie<sub>2</sub>, FDN sia<sub>2</sub> || ПБМ {老蛇} \*lau<sub>1</sub>=sie<sub>2</sub>

**107) «thin/тонкий»:** {薄} FCH pok<sub>7</sub>, HUB pɔʔ<sub>7</sub>, XNC pɔʔ<sub>7</sub>, JDU pɔʔ<sub>7</sub>, FQN pɔ<sub>1</sub>(H), FUA pɔʔ<sub>7</sub> (H), SHON pɔʔ<sub>7</sub> (H), PIN pɔ<sub>1</sub> || ПБМ {薄} \*pok<sub>7</sub>

**108) «wind/ветер»:** {風} FCH huŋ<sub>1</sub>, GTN huŋ<sub>1</sub>#(LN)<sup>40</sup>, NND xuŋ<sub>1</sub>#(LN), HUB xuŋ<sub>1</sub>, XNC huŋ<sub>1</sub>, JDU huŋ<sub>1</sub>, ZHN xuŋ<sub>1</sub>#(LN), FDN xuŋ<sub>1</sub>#(ln), FQN huŋ<sub>1</sub>#, CNL huŋ<sub>1</sub>#, YNT huŋ<sub>1</sub>#, FUA huŋ<sub>1</sub>#, SHON xuŋ<sub>1</sub>#<sup>41</sup>, PIN xuŋ<sub>1</sub> || ПБМ {風} \*huŋ<sub>1</sub>

**109) «worm/червяк»:** a) {猴蚓} NND kau<sub>2</sub>=ɔŋ<sub>3</sub>, HUB kau<sub>2</sub>=uɔŋ<sub>3</sub>, XNC kau<sub>2</sub>=ɔŋ<sub>3</sub>, JDU kau<sub>1</sub>=yŋ<sub>3</sub>, ZHN ka<sub>2</sub>=yŋ<sub>3</sub>, FDN kau<sub>2</sub>=oŋ<sub>3</sub>, CNL ka<sub>3</sub>=uŋ<sub>3</sub>, YNT kau<sub>1</sub>=uŋ<sub>3</sub>, FUA ka<sub>2</sub>=uŋ<sub>3</sub>, SHON kau<sub>2</sub>=xoŋ<sub>3</sub>; a.1) {牙蚓} FCH ŋauŋ<sub>3</sub>, GTN nie<sub>4</sub>=ɔuŋ<sub>3</sub>, PIN mi<sub>4</sub>=ɔuŋ<sub>3</sub>; a.2) FQN {琴蚓} k<sup>h</sup>iŋ<sub>2</sub>=ŋɔŋ<sub>3</sub> || ПБМ {猴蚓} \*kau<sub>2</sub>=oŋ<sub>3</sub>

Формы группы (a) внешне напоминают двусложный КДК этимон 蚯蚓 \*k<sup>h</sup>u-Lón ‘дождевой червь’, но никак не могут быть выведены из него фонетически; отсюда можно сделать заключение, что соответствующий двуслог мог иметь исконно несинитическое происхождение и параллельно (независимо) заимствоваться из неизвестного источника как в КДК, так и в миньские диалекты. Косвенно на это могут намекать и варианты,

<sup>37</sup> Большинство ВМ диалектов имеют литературное чтение: FCH sieŋ<sub>2</sub>, CNL sieŋ<sub>2</sub>, GTN sieŋ<sub>2</sub>, PIN sieŋ<sub>2</sub>, NND sim<sub>2</sub>, FUA sieŋ<sub>2</sub>, ZHN θim<sub>2</sub>, SHON sieŋ<sub>2</sub>, FDN sieŋ<sub>2</sub> (Liu & He 1998: 481–482).

<sup>38</sup> Лексема в фучжоуском диалекте указана согласно данным сборника Liu & He 1998: 481, так как в основном источнике (Nakajima 1979: 45) указано литературное чтение θieŋ<sub>2</sub>.

<sup>39</sup> Хироюки указывает регистровый тон 44, который, скорее всего, соотносится с тоном 2 («светлый ровный» 陽平; Hiroyuki 2020: 675).

<sup>40</sup> Линь Ханьшэн указывает регистровый номер 44, который, вероятно, соотносится с тоном 1 («темный ровный» 陰平; Lin 2002: 33).

<sup>41</sup> Аналогичная ситуация, см. сноску 39.

представленные в подгруппах а.1 и а.2, первый слог в которых отличается от первого слога в основном варианте; это может быть отражением префиксальной вариативности в источнике (источниках) заимствования.

110) «year/год»: {年} FCH *nien*<sub>2</sub>, GTN *nien*<sub>2#</sub>, NND *nin*<sub>2#</sub>, HUB *nien*<sub>2</sub>, XNC *nin*<sub>2</sub>, JDU *nin*<sub>2</sub>, ZHN *nin*<sub>2#</sub>, FDN *nien*<sub>2#</sub>, FQN *nien*<sub>2</sub>, CNL *lien*<sub>3</sub> (*nien*<sub>3</sub>)#, YNT *nien*<sub>2#</sub>, FUA *nien*<sub>2#</sub>, SHON *nien*<sub>2#</sub>, PIN *nien*<sub>2</sub> || ПВМ {年} \**nien*<sub>2</sub>

### Анализ

Сопоставление КДК базовой лексики и форм, реконструированных для ПВМ, показывает 63 случая (57%) несомненного совпадения корневых морфем. Помимо этого, 11 ПВМ реконструкций продолжают позднедревнекитайские инновации, постепенно заместившие КДК эквиваленты начиная с эпохи Хань (II в. до н.э. — III в. н.э.): ‘все’ ПДК 都 \**tā* → ПВМ 都 \**tu*<sub>1</sub>; ‘жечь’ ПДК 燒 \**šjetw* → ПВМ 燒 \**šiwu*<sub>1</sub>; ‘полный’ ПДК 滿 \**mān* → ПВМ 滿 \**miu*<sub>3</sub>; ‘хороший’ ПДК 好 \**háw* → ПВМ 好 \**ho*<sub>3</sub>; ‘голова’ ПДК 頭 \**d<sup>h</sup>əw* → ПВМ 頭 \**t<sup>h</sup>au*<sub>2</sub>; ‘слышать’ ПДК 聽 \**t<sup>h</sup>iēŋ* → ПВМ 聽 \**t<sup>h</sup>iaŋ*<sub>1</sub>; ‘красный’ ПДК 紅 \**ɣwōŋ* → ПВМ 紅 \**oŋ*<sub>2</sub>; ‘кожа’ ПДК 皮 \**b<sup>h</sup>e* → ПВМ 皮 \**p<sup>h</sup>ui*<sub>2</sub>; ‘дерево’ ПДК 樹 \**š<sup>h</sup>wo* → ПВМ 樹 \**t<sup>h</sup>iu*<sub>5</sub>; ‘два’ ПДК 兩 \**l<sup>h</sup>áŋ* → ПВМ 兩 \**laŋ*<sub>5</sub>; ‘идти’ 去 \**k<sup>h</sup>ò* → ПВМ 去 \**k<sup>h</sup>ø*<sub>4</sub>.

Далее, анализ отчетливо показывает, что в большинстве случаев, когда ПВМ отклоняется от языка письменных памятников КДК и ПДК, он наследует общеминьские инновации (общеминьские формы, ввиду ряда проблем с их фонетической реконструкцией, приводятся ниже с позднедревнекитайскими, а не собственно праминьскими чтениями):

- ‘кусать’: ПМ 咬 \**ɣáw* → ПВМ 咬 \**ka*<sub>5</sub>
- ‘черный’: ПМ 烏 \**?ō* → ПВМ 烏 \**u*<sub>1</sub>
- ‘ноготь’: ПМ 甲 \**kāp* → ПВМ 指甲 \**tsiŋ*<sub>3</sub>=*kap*<sub>6</sub>
- ‘земля’: ПМ 塗 \**d<sup>h</sup>ō* → ПВМ 塗 \**t<sup>h</sup>u*<sub>2</sub>
- ‘нога’: ПМ 骹 \**k<sup>h</sup>āw* → ПВМ 跤 \**k<sup>h</sup>a*<sub>1</sub>
- ‘давать’: ПМ 乞 \**k<sup>h</sup>it* → ПВМ 乞 \**k<sup>h</sup>yt*<sub>6</sub>
- ‘убивать’: ПМ 劊 \**d<sup>h</sup>āi* → ПВМ 劊 \**t<sup>h</sup>ai*<sub>2</sub>
- ‘колени’: ПМ 骹 \**k<sup>h</sup>āw* → ПВМ 骹肚頭 \**k<sup>h</sup>a*<sub>1</sub>-*lu*<sub>3</sub>-*lau*<sub>2</sub> ~ \**k<sup>h</sup>a*<sub>1</sub>-*lu*<sub>3</sub>-*t<sup>h</sup>au*<sub>2</sub>
- ‘лист’: ПМ 箬 \**háuk* → ПВМ 箬 \**niok*<sub>7</sub>
- ‘лежать’: ПМ 倒 \**tá* → ПВМ 倒 \**tø*<sub>3</sub>
- ‘рот’: ПМ 喙 \**č<sup>h</sup>waś* → ПВМ 喙 \**t<sup>h</sup>ui*<sub>4</sub>
- ‘шея’: ПМ 脰 \**d<sup>h</sup>òw* → ПВМ 脰項 \**tau*<sub>5</sub>-*uŋ*<sub>3</sub>
- ‘ночь’: ПМ 暝 \**miēŋ* → ПВМ 暝晡 \**maŋ*<sub>2</sub>-*mi*<sub>1</sub>
- ‘не’: ПМ 無 \**mwo* (\**ŋwo*) → ПВМ \**m*<sub>2</sub> (~\**ŋ*<sub>2</sub>)
- ‘один’: ПМ 蜀 \**šok* → ПВМ 蜀 \**suok*<sub>7</sub>
- ‘человек’: ПМ 儂 \**nāuŋ* → ПВМ 儂 \**næŋ*<sub>2</sub>
- ‘сказать’: ПМ 講 \**kōŋ* и 話 \**wāŋ* → ПВМ 講話 \**kōŋ*<sub>3</sub>-*ua*<sub>5</sub>
- ‘спать’: ПМ 睏 \**k<sup>h</sup>wəŋ* → ПВМ 睏 \**k<sup>h</sup>oŋ*<sub>5</sub>
- ‘стоять’: ПМ 企 \**k<sup>h</sup>jé* → ПВМ 倚 \**k<sup>h</sup>ie*<sub>5</sub>
- ‘тот’: ПМ 許 \**hó* → ПВМ 許隻 \**hi*<sub>3</sub>-*tsiek*<sub>6</sub>
- ‘этот’: ПМ 茲 \**cji* → ПВМ 這隻 \**tsi*<sub>3</sub>-*tsiek*<sub>6</sub>
- ‘кто’: ПМ 底儂 \**táj-nāuŋ* → ПВМ 底儂 \**ti*<sub>1</sub>-*næŋ*<sub>2</sub> и ПВМ 乜儂 \**nø*<sub>6</sub>-*næŋ*<sub>2</sub>
- ‘короткий’: ПМ 短 *tuei*<sub>3</sub> → ПВМ 短 \**toi*<sub>3</sub>
- ‘червяк’: ПМ 蚓 \**ún* → ПВМ 猴蚓 \**kau*<sub>2</sub>=*oŋ*<sub>3</sub>

Удалось выявить лишь три специфически восточноминьские инновации, не имеющие прямых аналогов ни в ҚДК/ПДК, ни в СК, ни в ПМ:

- ‘пепел’: ҚДК 灰 \**māj* → ПВМ 火焮 \**hui<sub>3</sub>=hu<sub>1</sub>*
- ‘много’: ҚДК 多 \**tāj* → ПВМ \**se<sub>6</sub>*
- ‘маленький’: 小 ҚДК \**séw* → ПВМ 嫩 \**nən<sub>5</sub>*

Единственным случаем внутри 100-словного списка, когда ПВМ этимон продолжает непосредственно ҚДК этимон, «минуя» ПДК состояние, оказывается уже упоминавшаяся выше лексема ‘собака’ (ҚДК 犬 \**k<sup>hw</sup>in* → ПВМ \**k<sup>h</sup>εη<sub>3</sub>*). Очевидно, что этого аргумента недостаточно для того, чтобы выделять восточноминьские языки в отдельную ветвь. Скорее можно было бы предполагать, что ‘собака’ унаследована ВМ ветвью от неизвестного нам «доминьского» диалектного субстрата, который некоторое время сосуществовал с носителями ВМ языков на территории современной провинции Фуцзянь; возможно, какие-то дополнительные аргументы в поддержку этой гипотезы удастся получить в ходе более глубокого изучения как базисной, так и культурной лексики миньдунской ветви.

Безусловный интерес вызывает небольшой слой предположительно австроазиатской лексики (см. ‘сухой’, ‘знать’; диминутив ‘ребенок, сын’ в составе двуслога ‘семя’ и др.), существование которого отмечено уже Дж. Норманом (Norman 1991). Кроме этого, в базисной лексике миньдунских диалектов также прослеживаются заимствования из тай-кадайской семьи, например: ‘кусать’, ‘убивать’, ‘полный’. Впрочем, перечисленные заимствования не являются исключительными для восточных диалектов: так, лексемы ‘кусать’, ‘сухой’, ‘убивать’ прослеживаются по всему миньскому ареалу, а ‘полный’ и ‘знать’, наряду с диминутивом ‘ребенок, сын’, представлены также в южной ветви (Лоренц 2020: 166). Таким образом, эти формы можно восстанавливать для праминьского состояния и считать их свидетельством в пользу тесных контактов с некитаеязычными носителями еще до разделения праминьского на отдельные ветви.

Отдельно перечислим основные замены в конкретных ВМ диалектах, отделяющих их от ПВМ состояния:

- PIN: ‘что’ {哪毛~毛毛} (*nɔ<sub>1</sub>*)-*nɔ<sub>4</sub>*; ‘давать’ *k<sup>h</sup>uai<sub>4</sub>#*; ‘лежать’ {筊倒} *ɕiη<sub>5</sub>=tɔ<sub>3</sub>*
- XNC: ‘лежать’ {筊倒} *æ<sub>3</sub>=tɔ<sub>3</sub>*; ‘один’ {蜀} *git<sub>7</sub>*
- SHON ‘видеть’ {映着} *ɕη<sub>5</sub>-tyø<sub>7</sub>*; ‘один’ {蜀} *sí<sub>7</sub>*
- FUA: ‘давать’ {錢} *tsiη<sub>2</sub>#*; ‘один’ {蜀} *sik<sub>7</sub>#*
- JDU: ‘лежать’ *æiη<sub>3</sub>*
- HUB: ‘что’ {哪毛~毛毛} *nɔ<sub>6</sub>*-*nɔ<sub>4</sub>*~*nɔ<sub>6</sub>*-*nɔ<sub>0</sub>*
- FDN: ‘что’ {哪毛~毛毛} *nɔ<sub>5</sub>*-*nɔ<sub>4</sub>*
- NND: ‘что’ {哪毛~毛毛} *nɔ<sub>7</sub>*-*nɔ<sub>6</sub>*
- YNT: ‘что’ {哪毛~毛毛} *nɔ<sub>5</sub>*-*nɔ<sub>6</sub>*

Новейшее генеалогическое дерево восточноминьской ветви (рис. 3), полученное после включения дополнительных диалектов и исправления ряда ошибок в списках, в целом подтверждает распределение конкретных диалектов по ветвям фунин и хоугуань, предложенное в LACD 2012, за исключением диалектов пиннань и чжоунин, которые оказываются включенными в противоположные ветви; действительно, пиннань, например, обнаруживает как общие инновации с фунин (‘человек’), так и с хоугуань (‘лежать’). Очевидно, что, ввиду неизбежных статистических погрешностей на столь близком уровне родства, вопрос уточнения классификации этих диалектов должен будет решаться на значительно большем количестве материала (как минимум на 200-словных списках базисной лексики).

Интересной особенностью стал факт, что непосредственно диалект ниндэ не вошел в т.н. «кластер ниндэ» (по А. Хироюки), образованный диалектами HUB, XNC, JDU. Анализ показывает, что диалект ниндэ лексически ближе к диалектам фудин и шоунин, чем к трем диалектам, распространенным в окрестностях г. Ниндэ и описанным в работах Хироюки. Ключевым разделяющим элементом в нашем случае оказывается лексема 'кто': 毛儂 ~ 哪儂 HUB  $næiŋ_4$ , XNC  $næŋ_4$ , JDU  $nɔ_6-næiŋ_0$ , но NND 底儂  $ti_5-nøŋ_2$  (она же сближает диалект ниндэ с гутянь и чжоунин: GTN  $təuŋ_1$ , ZHN  $tæiŋ_3$ ). Напротив, единственной эксклюзивной изоглоссой между диалектом ниндэ и (единственным из «кластера ниндэ») диалектом хубэй оказывается лексема 'что' 哪毛 ~ 毛毛: HUB  $nɔ_6-nɔ_4 \sim nɔ_6-nɔ_0$ , NND  $nɔ_7-nɔ_6$ ; впрочем, она же представлена и в диалектах FDN ( $no_5-no_4$ ) и YNT ( $no_5-no_6$ ). Помимо этого, диалект хубэй показывает одну эксклюзивную инновацию ( $k^hɔn_4$  'лежать'); также стоит обратить внимание на любопытную изоглоссу 'лежать' между цзюду, пиннань и сяньцунь: JDU  $æiŋ_3$ , PIN 筊倒  $ɔiŋ_5=tɔ_3$ , XNC  $æ_3=tɔ_3$ , из-за которой диалект пиннань, скорее всего, переместился из ветви фуин в хоугуань. Наконец, ср. лексему 'теплый', которая оказывается соединяющим звеном между диалектами хубэй, цзюду, сяньцунь и фуань: HUB  $kɔn_2$ , JDU  $kɔn_2$ , XNC  $kɔn_2$ , FUA  $kɔiŋ_2-(næiŋ_0)$ . Все это говорит в пользу того, что кластер хубэй — сяньцунь — цзюду действительно не является ближайшим родственником диалекта собственно г. Ниндэ.

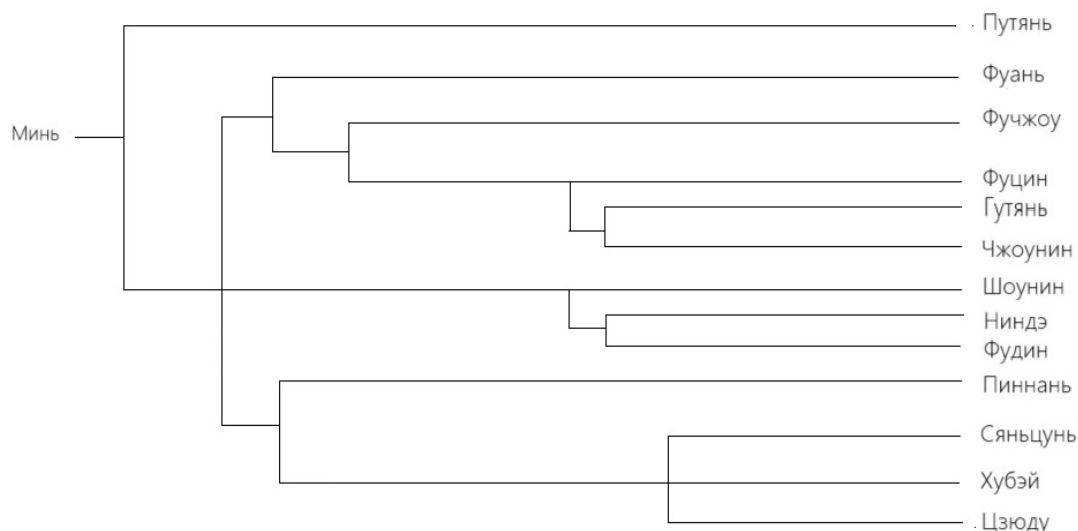


Рис. 3. Текущее генеалогическое дерево восточноминьских диалектов по данным 100-словных списков (включая д. Путянь как «outlier»).

Учитывая, что основные расхождения внутри 100-словного списка по диалектам проходят по линии оформления «главной» корневой морфемы различными префигурируемыми или суффигурируемыми «расширителями», в качестве дополнительного эксперимента было решено построить альтернативное дерево, в котором лексемы с одним и тем же главным корнем, но различающиеся вспомогательными морфемами, маркировались бы разными, а не одними и теми же индексами когнации, как если бы они представляли собой полноценные лексические замены. Результаты такого подсчета, намеренно идущего вразрез с постулатами классической лексикостатистики, приведены ниже (рис. 4).

Как видно, при таком подходе внутреннее деление на ветви фуин и хоугуань нарушается намного сильнее, чем дерево, представленное выше. Фактически на новом дереве остаются лишь три устойчивых кластера: (а) хубэй, сяньцунь, цзюду; (б) гутянь,

чжоунин; (в) ниндэ, фудин. Хотя этот результат и приходится считать негативным, он позволяет сделать важное предположение относительно того, что использование тех или иных «расширителей» в ходе трансформации лексической системы плохо подходит в качестве аргумента для генетической классификации (по крайней мере, применительно к китайским диалектам) — реально оно может отражать как ареальные связи, так и независимо протекающие «гомопластические» процессы. Впрочем, эта гипотеза, безусловно, нуждается в дальнейшей апробации на расширенном материале.

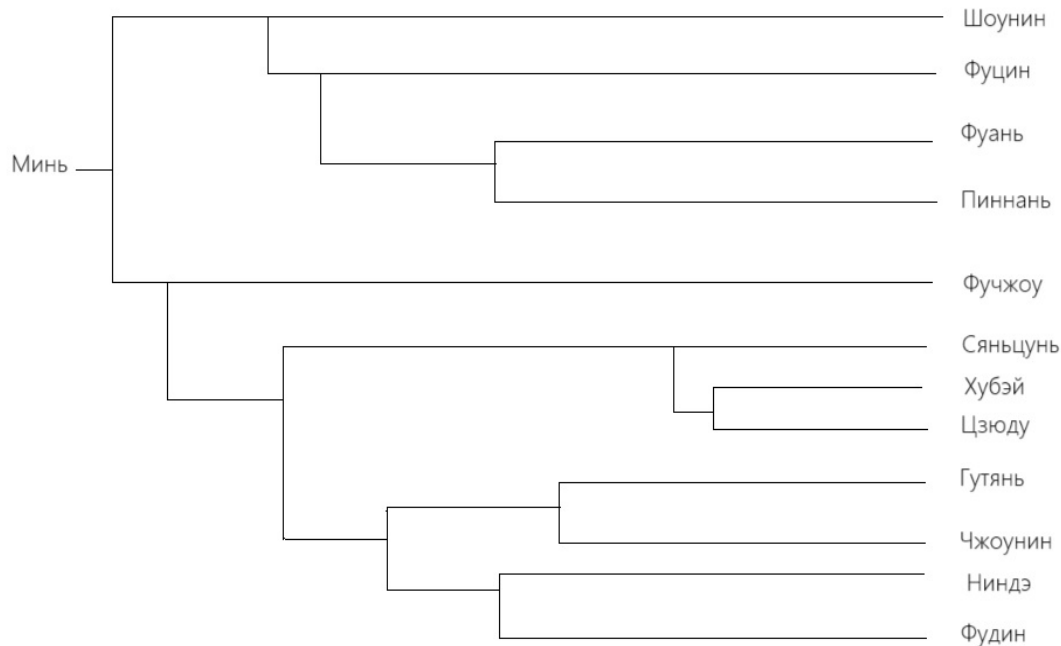


Рис. 4. «Альтернативное» генеалогическое дерево восточноминьских диалектов с различными индексами когнации для однокоренных, но разноморфемных лексем.

### Сопоставление с путяньским диалектом

Выше было упомянуто, что отдельный вопрос составляют взаимоотношения путяньского диалекта с миньдунскими диалектами, поскольку на общеминьском генеалогическом дереве он регулярно вклинивался в восточную ветвь. Для начала перечислим путяньские лексемы, совпадающие с общеминьскими<sup>42</sup> инновациями, чтобы эксплицитно подтвердить отнесение путяньского диалекта к собственно миньской ветви. Видно, что совпадений с общеминьским достаточно, чтобы отвести любые сомнения:

- ‘кусать’: 咬  $ko_4$  (ср. ПМ 咬  $*\gamma\acute{a}w$ )
- ‘черный’: 烏  $o_1$  (ср. ПМ 烏  $*?o$ )
- ‘ноготь’: 指掌甲  $ts^h i\eta_4 = li\eta_2 = \eta o_4$  (ср. ПМ 甲  $*k\grave{a}p$ )
- ‘земля’: 土  $t^h ou_2$  (ср. ПМ 塗  $*d^h\bar{o}$  и 泥  $*n\bar{a}j$ )
- ‘нога’: 跣  $k^h o_1$  (ср. ПМ 骹  $*k^h\grave{a}w$ )
- ‘давать’: 乞  $kek_6$ <sup>43</sup> (ср. ПМ 乞  $*k^h i\grave{t}$ )
- ‘слышать’: 聽  $t^h i e_1$  (ср. ПМ 聽  $*t^h i\bar{e}\eta$ )

<sup>42</sup> Реконструкция общеминьского состояния взята из статьи Лоренц 2020.

<sup>43</sup> Накадзима указывает регистровый тон 3, который, вероятно, может соотноситься с 6-м тоном («темный входящий» 陰入; Nakajima 1979: 193).

- ‘убивать’: 劊 *tai*<sub>2</sub> (ср. ПМ 劊 \**d<sup>h</sup>āi*)
- ‘колено’: 骹腹头 *k<sup>h</sup>v<sub>1</sub>-pv<sub>2</sub>-t<sup>h</sup>au*<sub>2</sub><sup>44</sup> (ср. ПМ 骹 \**k<sup>h</sup>āw*)
- ‘лист’: 葉 *niau*<sub>3</sub> (ср. ПМ 箬 \**ḡauk*)
- ‘лежать’: 倒下 *to<sub>1</sub>-xa*<sub>2</sub> (ср. ПМ 倒 \**tāw*)
- ‘рот’: 嘴 *tshui*<sub>1</sub> (ср. ПМ 嘴 \**cjwāj*)
- ‘шея’: 脰顛 *tau*<sub>2</sub>-*ly*<sub>1</sub> (ср. ПМ 脰 \**d<sup>h</sup>ōw*)
- ‘ночь’: 暗暝 ▪ *ak<sub>7</sub>=ma<sub>4</sub>-li*<sub>2</sub> (ср. ПМ 暝 \**miēŋ* и 暗 \**ʔəm*)
- ‘один’: 蜀 *ʔok*<sub>5</sub> (ср. ПМ 蜀 \**ʔok*)
- ‘человек’: 儂 *naŋ*<sub>2</sub> (ср. ПМ 儂 \**nāŋ*)
- ‘сказать’: 講話 *koŋ<sub>3</sub>-uo*<sub>6</sub> (ср. ПМ 講 \**kōŋ* и 話 \**wāj*)
- ‘спать’: 睏 *k<sup>h</sup>ui*<sub>1</sub> (ср. ПМ 睏 \**k<sup>h</sup>wə̀n*)
- ‘стоять’: 企 *k<sup>h</sup>æ*<sub>4</sub> (ср. ПМ 企 \**k<sup>h</sup>jé*)
- ‘тот’: 彼個 *hæk*<sub>3</sub><sup>45</sup>-*kei*<sub>2</sub> (ср. ПМ 許 \**hó*)
- ‘этот’: 這個 *tsek*<sub>3</sub><sup>46</sup>-*kei*<sub>2</sub> (ср. ПМ 茲 \**cjɿ*)

Однако при сопоставлении путяньских форм с восточноминьскими инновациями совпадений обнаружено не было, ср.:

- ‘много’: ВМ \**sε<sub>6</sub>*, но путянь 齊 *ʔei*<sub>4</sub>;
- ‘маленький’: ВМ 嫩 \**nɔn<sub>5</sub>*, но путянь 細 *ʔek*<sub>4</sub>.

Помимо этого, архаизм ‘собака’ (КДК 犬 \**k<sup>hwo</sup>ʔin*) в путяньском диалекте также отсутствует, что можно считать дополнительным свидетельством о непринадлежности его к миньдунской ветви.

В то же время путяньский диалект не разделяет и южноминьские инновации, ср.:

- ‘мясо’: ЮМ *bak*<sub>7</sub>, но путянь 肉 *næk*<sub>5</sub>;
- ‘шея’: ЮМ 頷 *am*<sub>3</sub>, но путянь 脰顛 *tau*<sub>2</sub>-*ly*<sub>1</sub>.

Напротив, немногочисленные эксклюзивные инновации скорее являются весомым аргументом для выделения диалекта путянь в независимую ветвь:

- ‘кто’: путянь 何儂~許儂 *hŋ<sub>1</sub>-ŋaŋ<sub>2</sub>*;
- ‘червяк’: путянь *niau*<sub>4</sub>=*æ*<sub>3</sub>.

Таким образом, как минимум на основании анализа 100-словного списка путяньский диалект не показывает никакой специфической близости ни к восточно-, ни к южноминьской ветви, что говорит о необходимости выделения его в особую подгруппу миньской клады. Возможно, какие-то более точные классификационные выводы удастся получить при расширении материала анализа как минимум до 200-словных списков.

## Литература

- Лоренц, М. М. 2020. Анализ базисной лексики диалектов группы Минь и реконструкция праиньского 100-словного списка. *Вопросы языкового родства* 18/1-2: 126–169.
- Старостин, Г. С. 2013. К проблеме двух собак в классическом китайском языке: *canis comestibilis* vs. *canis venaticus*? В: Н. П. Гринцер и др. (ред.). *Institutionis Conditore: Илье Сергеевичу Смирнову. Orientalia et Classica*, Vol. L: 253–267. Москва: РГГУ.
- Старостин, С. А. 1989. *Реконструкция древнекитайской фонологической системы*. Москва: Наука.

<sup>44</sup> Лексема указана согласно данным электронного словаря 興化語記 *Xinghua yuji*, ввиду лакуны в основном источнике.

<sup>45</sup> Автор указывает регистровый тон 5, который, вероятно, соотносится с тоном 3 (восходящий 上聲). На сайте электронного словаря по путяньскому диалекту *Xinghua yuji* также указывается тон 3.

<sup>46</sup> Аналогичная ситуация, см. выше.

## References

- Cài, Guómèi (蔡国妹). 2006. *Púxiān fāngyán dùxìng fāngyán de tèzhēng* (蒲仙方言渡性方言的特征): 申請文學博士。Fúzhōu: Fújiàn shīfàn dàxué (福建師範大學)。
- Chén, Zhāngtài, Rúlóng Lǐ (陳章太, 李如龍). 1991. *Mǐnyǔ yánjiū* (閩語研究). Beijing: Yǔwén chūbǎnshè (語文出版社)。
- Dèng, Xiǎohuá (鄧曉華). 1994. Nánfāng hànǔ zhōngde gǔnándǎo yǔ chéngfèn (南方漢語中的古南島語成分). *Minority Languages of China* (民族語文) 3: 36–40.
- Dǒng, Zhōngsī (董忠司). 1993. Yǒuguān táiwānhuà tsa1 pɔ1, tsa1 bɔ2 tànyuán wèntí — shìlùn kěnéng shì mǐnyuèyǔ cánliúde yīgè hénjì (有關臺灣話tsa1 pɔ1, tsa1 bɔ2 探源問題——試論可能是閩越語殘留的一個痕跡). In: *Dìsānjièguójì mǐnfāngyán tāolùnhuì huìyì lùnwénjí* (第三屆國際閩方言討論會議論文集): 99–106. Xianggang: Xiānggǎng zhōngwén dàxué (香港中文大學)。
- Féng, Àizhēn (馮愛珍). 1993. *Fúqīng fāngyán yánjiū* (福清方言研究). Beijing: Shèhuì kēxué wénxiàn chūbǎnshè (社會科學文獻出版社)。
- Hinghwa Dialect Digital Dictionary Project. Available on-line at: <https://hinghwa.cn>.
- Hiroyuki, Akitani (秋谷裕幸). 2018. *A phonological history of the Ningde dialect of Eastern Min*. Taipei: Institute of Linguistics, Academia Sinica.
- Hiroyuki, Akitani (秋谷裕幸). 2020. *Mǐndōng sì xiàn xiànshì fāngyán diào chá yánjiū* (閩東四縣市方言調查研究). Shanghai: Shànghǎi jiàoyù chūbǎnshè (上海教育出版社)。
- Hóu, Jīngyī (侯精一). 2002. *Xiàndài hànǔ fāngyán gài lùn* (現代漢語方言概論). Shanghai: Shànghǎi jiàoyù chūbǎnshè (上海教育出版社)。
- HYDC = *Hànǔ yǔ dà cídiǎn* (漢語大詞典). 1993. Shanghai: Shànghǎi císhū chūbǎnshè (上海辭書出版社)。
- Kassian, Alexei, George Starostin, Anna Dybo, Vasily Chernov. 2010. The Swadesh wordlist: an attempt at semantic specification. *Journal of Language Relationship* 4: 46–89.
- LACD = Cao Zhiyun (ed.). *The Linguistic Atlas of Chinese dialects. 2nd edition: Chinese dialect volume* (中國語言地圖集第二版: 漢語方言卷). 2012. Beijing: Shāngwù yìnshūguǎn chūbǎnshè (商務印書館出版社)。
- Li, Fang-kuei. 1977. *A handbook of comparative Tai*. Honolulu: University of Hawaii Press.
- Lǐ, Rúlóng, Yùzhāng Liàng (李如龍, 梁玉璋). 1994. *Fúzhōu fāngyán cídiǎn* (福州方言詞典). Fuzhou: Fúzhōu rénmín chūbǎnshè (福州人民出版社)。
- Lǐ Bīn (李濱). 2014. *Mǐndōng gǔtián fāngyán yánjiū* (閩東古田方言研究). Xiamen: Xiàmén dàxué chūbǎnshè (廈門大學出版社)。
- Lín, Hánshēng (林寒生). 2002. *Mǐn dōngfāng fāngyán cíhuì yǔfǎ yánjiū* (閩東方言詞匯語法研究). Kunming: Yúnnán dàxué chūbǎnshè (雲南大學出版社)。
- Liú Zǔbì, Hè Wèi (劉祖陞, 賀魏). 1998. *Fújiànshěngzhì fāngyánzhì* (福建省志方言志). Beijing: Fāngzhì chūbǎnshè (方志出版社)。
- Lorentz, Marina. 2020. Analysis of Min basic lexicon and reconstruction of the Swadesh wordlist for Proto-Min. *Journal of Language and Relationship* 18/1-2: 126–169.
- Macklay, Robert Samuel, Caleb Cook Baldwin. 1898. *An alphabetic dictionary in the Foochow dialect*. Foochow: Methodist episcopal mission press.
- Méi, Zǔlín (梅祖麟). 1999. Jǐge Táiwan mǐnnánhuà chángyòng xūcí de láiyuán (幾個台灣閩南話常用虛詞的來源). In: Ting, Pang-Hsin (ed.). *Contemporary Studies on the Min dialects. Journal of Chinese Linguistics Monograph series* 14: 1–41. Berkeley: University of California.
- Nakajima, Motoki. 1979. *A comparative lexicon of Fukien dialects* (福建漢語方言基礎語彙集). Tokyo: Tōkyō Gaikokugo Daigaku Ajia · Afurika gengo bunka kenkyūjo.
- Norman, Jerry, Tsu-lin Mei. 1976. The Austroasiatics in ancient South China: Some lexical evidence. *Monumenta Serica* 32: 274–301.
- Norman, Jerry. 1977–1978. *A preliminary report on the dialects of Mintung*. *Monumenta Serica*, Vol. XXXIII: 326–348.
- Norman, Jerry. 1979. The verb 治 — a note on Min etymology (閩語里的“治”字). *Fangyan* 3: 179–181.
- Norman, Jerry. 1983. Some ancient Chinese dialect words in the Min dialects. *Fangyan* 3: 202–211.
- Norman, Jerry. 1991. The Min Dialects in Historical Perspective. *Journal of Chinese Linguistics* 3: 323–358.
- Ngai, Sing-sing. 2015. On the origin of special numerals for 'one' in south-eastern China: [kei213] in the western Min dialect of Shaowu. In: H. Chappell (ed.). *Diversity in Sinitic languages*: 190–225. Oxford University Press.
- Schuessler, Axel. 2007. *ABC etymological dictionary of Old Chinese*. Honolulu: University of Hawaii Press.



- Shorto, Harry L. 2006. *A Mon-Khmer Comparative Dictionary*. Canberra: Pacific Linguistics.
- Starostin, George. 2013. К проблеме двух собак в классическом китайском языке: *canis comestibilis* vs. *canis venaticus*? In: N. P. Grintser et al. (eds.). *Institutionis Conditori: Ilje Sergeevichu Smirnovu. Orientalia et Classica*, Vol. L: 253–267. Moskva: RSUH Publishers.
- Starostin, George. 2019. Chinese basic lexicon from a diachronic perspective: implications for lexicostatistics and glottochronology. *Journal of Language Relationship* 17/1–2: 153–176.
- Starostin, Sergei A. 1989. *Rekonstruktsiia drevnekитайской fonologicheskoi sistemy [Reconstruction of the Old Chinese phonological system]*. Moskva: Nauka.

*Marina Lorentz*. Classification of Eastern Min dialects and reconstruction of the 100-item wordlist for Proto-Eastern Min

In this paper, I propose an updated classification of the Eastern Min subgroup of the Sinitic family, based on its basic vocabulary, along with the onomasiological reconstruction of the 110-item Swadesh list for Proto-Eastern Min. A layer of Austroasiatic vocabulary, detected in the Eastern Min basic vocabulary, confirms the possible existence of an archaic substrate in the modern area of distribution of the dialects in question, leaving traces at the lexical level. A new phylogenetic tree confirms, with some minor deviations, the two traditionally accepted sub-branches (Funing and Houguan), previously based on phonetic criteria. The research also examines the relationship of the Eastern group with Putian dialect, confirming its independent status due to the lack of common innovations with the Southern and Eastern branches of Min.

*Keywords*: dialectology; Old Chinese language; Middle Chinese language; Min dialects; Eastern Min dialects; Putian dialect; lexicostatistics; basic lexicon.