

Российский государственный гуманитарный университет
Russian State University for the Humanities



RSUH/RGGU BULLETIN

№ 1 (16)

Academic Journal

Series:

Philology. Journal of Language Relationship

Moscow 2018

ВЕСТНИК РГГУ

№ 1 (16)

Научный журнал

Серия

«Филология. Вопросы языкового родства»

Москва 2018

Редакционный совет серий «Вестника РГГУ»

Е.И. Пивовар, чл.-кор. РАН, д-р ист. н., проф. (председатель)

Н.И. Архипова, д-р экон. н., проф. (РГГУ), А.Б. Безбородов, д-р ист. н., проф. (РГГУ), Х. Варгас (Ун-т Кали, Колумбия), А.Д. Воскресенский, д-р полит. н., проф. (МГИМО (У) МИД России), Е. Вятр (Варшавский ун-т, Польша), Дж. Дебарделебен (Карлтонский ун-т, Канада), В.А. Дыбо, акад. РАН, д-р филол. н. (РГГУ), В.И. Заботкина, д-р филол. н., проф. (РГГУ), В.В. Иванов, акад. РАН, д-р филол. н., проф. (РГГУ; Калифорнийский ун-т Лос-Анджелеса, США), Э. Камия (Ун-т Тачибана г. Киото, Япония), Ш. Карнер (Ин-т по изучению последствий войн им. Л. Больцмана, Австрия), С.М. Каштанов, чл.-кор. РАН, д-р ист. н., проф. (ИВИ РАН), В. Кейдан (Ун-т Карло Бо, Италия), Ш. Кечкемети (Национальная Школа Хартий, Сорбонна, Франция), И. Клюканов (Восточно-Вашингтонский ун-т, США), В.П. Козлов, чл.-кор. РАН, д-р ист. н., проф. (ВНИИДАД), М. Коул (Калифорнийский ун-т Сан-Диего, США), Е.Е. Кравцова, д-р психол. н., проф. (РГГУ), М. Крэмер (Гарвардский ун-т, США), А.П. Логунов, д-р ист. н., проф. (РГГУ), Д. Ломар (Ун-т Кельна, Германия), Б. Луайер (Ин-т геополитики, Париж-VIII, Франция), С. Масамичи (Ун-т Чуо, Япония), В.И. Молчанов, д-р филос. н., проф. (РГГУ), В.Н. Незамайкин, д-р экон. н., проф. (Финансовый ун-т при Правительстве РФ), П. Новак (Ун-т Белостока, Польша), Ю.С. Пивоваров, акад. РАН, д-р полит. н., проф. (ИНИОН РАН), Е. ван Поведская (Ун-т Сантьяго-де-Компостела, Испания), С. Рапич (Ун-т Вупперталь, Германия), М. Сасаки (Ун-т Чуо, Япония), И.С. Смирнов, канд. филол. н. (РГГУ), В.А. Тишков, акад. РАН, д-р ист. н., проф. (ИЭА РАН), Ж.Т. Тощенко, чл.-кор. РАН, д-р филос. н., проф. (РГГУ), Д. Фоглесонг (Ун-т Ратгерс, США), И. Фолтыс (Политехнический ин-т г. Ополе, Польша), Т.И. Хорхордина, д-р ист. н., проф. (РГГУ), А.О. Чубарьян, акад. РАН, д-р ист. н., проф. (ИВИ РАН), Т.А. Шаклеина, д-р полит. н., проф. (МГИМО (У) МИД России), П.П. Шкаренков, д-р ист. н., проф. (РГГУ)

Серия «Филология. Вопросы языкового родства»

Редакционная коллегия серии

В.А. Дыбо, гл. ред., акад. РАН, д-р филол. н. (ИнСлав РАН / РГГУ), Г.С. Старостин, зам. гл. ред., канд. филол. н., доц. (РГГУ), Т.А. Михайлова, отв. сек., д-р филол. н., проф. (МГУ им. М.В. Ломоносова), А.В. Дыбо, д-р филол. н., член-корр. РАН (ИЯз РАН), С.В. Кулланда, канд. ист. н. (ИВ РАН), М.А. Молина (ИЯз РАН), М.Н. Саенко, канд. филол. н. (Институт славяноведения РАН), И.С. Якубович, д-р филол. н. (Марбургский университет имени Филиппа)

Ответственные за выпуск: Г.С. Старостин, И.С. Якубович

УДК 800(05)
ББК 80/84я5

Вопросы языкового родства: Международный научный журнал / Рос. гос. гуманитар. ун-т; Рос. акад. наук. Ин-т языкознания; под ред. В. А. Дыбо. — М., 2018. — № 1(16). — x + 78 с. — (Вестник РГГУ. Серия «Филология. Вопросы языкового родства»: Научный журнал).

Journal of Language Relationship: International Scientific Periodical / Russian State University for the Humanities; Russian Academy of Sciences. Institute of Linguistics; Ed. by V. A. Dybo. — Moscow, 2018. — No. 1(16). — x + 78 p. — (RSUH/RGGU Bulletin. Series: Philology. Journal of Language Relationship: Academic Journal).

ISSN 2073-6320

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

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

Подписано в печать 10.03.2018. Формат 60×90/8.

Бум. офсетная.

Печать офсетная. Тираж 1050 экз.

Заказ №219

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

Table of Contents / Содержание

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

Articles / Статьи

<i>Giorgio Orlandi</i> . Evaluating the Sino-Tibeto-Austronesian Hypothesis.	1
[Дж. Орланди. К оценке сино-тибето-австронезийской гипотезы]	
<i>George Starostin</i> . Lexicostatistical Studies in Khoisan I: The Ju-ǀHoan Relationship	19
[Г. С. Старостин. Лексикостатистические исследования по койсанским языкам I: родство между языками жу и ǀхоан]	
<i>M. V. Oslon</i> . Отражение южнославянского *r в заимствованиях из старорумынского в цыганский.	62
[Mikhail Oslon. The reflexes of South Slavic *r in borrowings from Old Romanian into Romani]	

Replies / Ответы

<i>A. I. Kogan</i> . Ответ на статью А. С. Крыловой «Лексикостатистика новоиндоарийских языков: взгляд полевого лингвиста»	73
[Anton Kogan. Reply to the paper by A. S. Krylova “Lexicostatistics and the New Indo-Aryan languages: a field linguist's perspective”]	

Contributors

Anton Kogan — candidate of sciences (Philology), researcher, Department of Asian and African languages, Institute of Oriental Studies of the Russian Academy of Sciences (Moscow), kogan_anton@yahoo.com

Giorgio Orlandi — postgraduate student, Department of Humanities, Xiamen University, thegaygenius@gmail.com

Mikhail Oslon — candidate of sciences (Philology), researcher, Institute of Linguistics, Russian Academy of Sciences (Moscow), neoakut@gmail.com

George Starostin — candidate of sciences (Philology), head of Department of the history and philology of the Far East, Institute for Oriental and Classical Studies, Russian State University for the Humanities (Moscow); senior researcher, Institute for Oriental and Classical Studies, Higher School of Economics (Moscow); external professor, Santa Fe Institute (New Mexico), gstarst1@gmail.com

Сведения об авторах

Коган, Антон Ильич — канд. филол. наук, науч. сотрудник Отдела языков Азии и Африки Института востоковедения РАН (Москва), kogan_anton@yahoo.com

Орланди, Джорджио — аспирант гуманитарного отделения Сямэньского университета, thegaygenius@gmail.com

Ослон, Михаил Владимирович — канд. филол. наук, науч. сотрудник Института языкознания РАН (Москва), neoakut@gmail.com

Старостин, Георгий Сергеевич — канд. филол. наук, зав. кафедрой истории и филологии Дальнего Востока ИВКА РГГУ, старший научный сотрудник ИКВиА ВШЭ (Москва), внештатный профессор Института Санта-Фе (США, Нью-Мексико), gstarst1@gmail.com

Note for Contributors

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

G. Starostin
Institute for Oriental and Classical Studies
Russian State University for the Humanities
125267 Moscow, Russia
Miuskaya Square, 6
E-mail: gstarst@rinet.ru

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

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

Будущим авторам

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

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

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

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

Evaluating the Sino-Tibeto-Austronesian Hypothesis

This paper analyses and evaluates the alleged genetic relationship between Sino-Tibetan and Austronesian, proposed by the French sinologist Laurent Sagart. The aim of the following paper is neither to prove, nor to disprove the Sino-Tibeto-Austronesian superphylum but to argue whether the data presented in favour of this proposed genetic relationship do or do not stand the scrutiny of a historical linguist. This paper also considers the hypothetical homeland of Proto-Sino-Tibeto-Austronesian people, with an eye towards competing hypotheses, such as Sino-Indo-European. It is concluded that Sagart's approach may be insufficient for proof of controversial cases of disputed genetic relationship, given the non-obvious relatedness of the languages he is comparing.

Keywords: Sino-Tibeto-Austronesian hypothesis; Sino-Tibetan languages; genetic relationship of languages; comparative method.

“Der erste mir bekannte Sprachvergleich im heutigen Sinne des Wortes ist der gelehrte Holländer Hadr. Relandus, der in seinem Dissertationes miscellaneae, Utrecht 1706–08, die weite Verbreitung des malaischen Sprachstammes, sogar Lautvertretungsgesetze zwischen Malaisch und Madegassisch nachweist” [Georg von der Gabelentz, *Die Sprachwissenschaft: ihre Aufgaben, Methoden, und bisherigen Ergebnisse*, 1891, p. 26].

Historical overview

Since the beginning of Western scholarly interest in both Sino-Tibetan and Austronesian, a panoply of claims concerning a wider genetic relationship of these language families has been proposed until our days. Among these various proposals, ranging from quasi-scholarly publications to more serious works, we may remember (1) the Austric hypothesis, first proposed by Schmidt (1906) and then re-proposed by Reid (1994); (2) the “Indo-Chinese” hypothesis of Conrady (1896, 1916); (3) the Sino-Caucasian hypothesis of Starostin (1984, 1989, 1991, 2005); (4) the Sino-Indo-European argued for by Ulenbrook (1967) and Pulleyblank (1995ab, 1996), and somewhat reinforced by Beckwith (2002); (5) the Sapirian Sino-Dene hypothesis (1915), further strengthened by Shafer (1952); (6) the Hokan-Austronesian hypothesis of Rivet (1925, 1926); (7) the Japanese-Austronesian hypothesis of Kawamoto (1977, 1978, 1980, 1984, 1993); (8) the Quechua-Austronesian hypothesis of Kempler-Cohen (2012); (9) the Sino-Austronesian hypothesis (Sagart 1993); (10) the Sino-Tibeto-Austronesian hypothesis (Sagart 2004, 2005, 2008, 2013, 2016). To be fair, it must be remembered that the Malayo-Polynesian languages, a subgroup of the Austronesian family, were originally drawn near the Indo-European family after the great German philosopher and linguist Wilhelm von Humboldt (1767–1835) noted that nearly half of the vocabulary of Kawi, the poetic language of the Old Javanese texts, derives from Sanskrit (1836–1839).¹ Nevertheless, Humboldt did not explicitly claim that Ma-

¹ Campbell and Poser (2008: 61) applauded Humboldt's emphasis on grammatical structure. In their book they conclude that Humboldt determined Kawi as “Malayan” (Austronesian), in spite of the vocabulary of San-

layo-Polynesian was a branch of Indo-European. Rather, this erroneous genetic relationship was argued for by Franz Bopp (1791–1867), who concluded that “[d]afs aber der Zufall bei allen diesen Wörtern sein Spiel getrieben habe, und bei allen anderen, die wir noch aus diesem Sprachgebiete dem Sanskrit gegenüberstellen können, ist unglaublich, zumal da bei einigen Wortklassen, nämlich bei den Pronomina und Zahlwörtern, worauf gerade bei Verwandtschaftsbestimmungen der Sprachen sehr viel ankommt, die Übereinstimmung fast durchgreifend erscheint” (1841: 7). Bopp’s view of the linguistic facts was utterly rejected by von der Gabelentz (1891: 164), who wrote:

Es ist schrecklich verführerisch in der Sprachenwelt umherzuschwärmen, drauf los Vocabeln zu vergleichen und dann die Wissenschaft mit einer Reihe neu entdeckter Verwandtschaften zu beglücken. Es kommen auch schrecklich viele Dummheiten dabei heraus; denn allerwaertssind unmethodische Köpfe die vordringlichsten Entdecker.

It is extremely important to bear this chronology of facts in mind in our analysis of the Sino-Tibeto-Austronesian (STAN henceforth) hypothesis, because it illustrates one of the many illicit modifications of the standard comparative method. In other words, it tells us how the comparative method does *not* and therefore should *not* work.

We owe the attempt to prove that Chinese and Austronesian are genetically related to Sagart (1990). In its earliest version, the theory linked only Sinitic, not Tibeto-Burman, and it seems that this early proposal was greeted by harsh critiques from South-East Asian specialists. Sagart has since modified his views, and now considers Chinese related to both Tibeto-Burman and Austronesian. This later proposal was, nonetheless, greeted by yet another round of criticism both from Austronesianists (Blust 1995, Li 1995) and Sinologists (Wang 1995, Pulleyblank 1996).² Leaving aside for a while the fact that any reader with at least half open mind would have to accept that something far more serious is afoot in the STAN than merely a few chance resemblances, the problem for *some* critics of the STAN (e.g. Pulleyblank 1996) is that their proposed explanation of the facts seems as implausible as their criticised linguistic scenario, perhaps even more so.³

In the following paragraphs, theories concerning the existence of the STAN superphylum and other relative problems — from basic lexicon to the question regarding the *Urheimat* — will be independently discussed and evaluated.

Methodological considerations

If it is a truism in historical linguistics that shared paradigmatic morphology, intended as the combination of morphological markers in a string, is the most stable system and, thus, has the

skrit origin it contains. However, this is not exactly the opinion of Blust, who writes: “He noted the obvious fact that much of the vocabulary has been borrowed from Sanskrit, but suspected that under this layer of relatively recent borrowing was a deeper layer indicative of genetic relationship” (2013: 659). What emerges directly from Humboldt’s work, however, is closer to Campbell & Poser’s depicted picture. Humboldt recognised the borrowings and, unlike Bopp, did not fall prey of lexical similarities. Rather, he utilised, as proof of kinship, both grammatical structures and sound correspondences (Humboldt 1836–9: 208-210, 217, 219-220, 221).

² STAN is partially accepted by Wolf (2010), who holds the view that Austronesian and Sino-Tibetan are two branches of the same “gigantophylum.”

³ On a personal note it should be premised that this author’s opinion toward the so-called “long-range comparisons” is unbiased. The two most promising superphyla which exist today, according to this writer’s opinion, are Indo-Uralic (Kortlandt 2010) and Yeniseic-Na-Dene, whose affiliation is prevalently built on paradigmatic morphology.

the strongest probatory force, then it is also true that any proposal of genetic relationship, no matter whether short-ranged or long-ranged, starts with the identifications of cognate forms in basic lexicon.⁴ Starting from 1990, Sagart has been busy in providing Sino-Austronesian cognates, and in modifying or incrementing his earlier proposals. In one of his latest proposals (Sagart 2005), for example, Sagart has eliminated some of the false cognates which appeared in his early works, such as 脯 *pú*, erroneously glossed as “breast.”⁵ It is laudable that Sagart is not afraid of revising his positions when new information forces him to rethink a problem. Nevertheless, his newer sixty-one basic vocabulary comparisons between AN and Old Chinese⁶ (OC henceforth) still remain problematic.

First of all, it would be more accurate to compare Proto-Austronesian (PAN hereafter) forms with Proto-Sino-Tibetan (PST), because the addition of Tibeto-Burmese cognates can strengthen any lexical match between PAN and OC; conversely, if such cognates cannot be found, one would be more prone to interpret these lexical matches as a result of areal contacts between OC and PAN.⁷

⁴ By “basic lexicon” is not necessarily meant, in this paper, the 100-word list proposed by Morrison Swadesh (1971). It must be remembered that this definition is not an absolute, and that in these years several scholars from the Moscow School have worked within this “paradigm,” trying to separate the “more stable” part from its “less stable” counterpart. For further knowledge, the interested reader should consult Dolgopolsky (1964) and Starostin (1984, 1991). See also the Leipzig-Jakarta list (2009).

⁵ The *Jíyùn* clearly indicates that the meaning of this character, a somewhat dialectal form, should be “chicken breast.” It follows that it does not match semantically the alleged Austronesian (AN hereafter) form *riba*.

⁶ The very notion of “Old Chinese” is somewhat misleading. By Old Chinese is generally meant the language from the Early Zhōu (1046–771 BC) to the Western Hàn (206 BC–9 AD) period. Nevertheless terms such as “Old Chinese,” “Middle Chinese,” “Modern Chinese,” *etc.* leave the reader with the impression of a linguistic unity in space and time throughout present-day Chinese territory, reinforced by certain lingering adumbrations of the somewhat involved and generally rather special political unity that has allegedly existed from then to now. In fact, the old “Chinese” language spoken by the Zhōu—not the earliest form of “Old Chinese”—was surely and undoubtedly different not only from the medieval language known as “Middle Chinese” but even from the so-called “Late Old Chinese” of the Hàn times. It is well-known that Hàn people, intended as the people of the Hàn dynasty—including its founder Liú Bāng—were descendant of Chǔ (1030 BCE–223), therefore it is not merely possible but very probable that Chǔ-Hàn Chinese (*i.e.* Late Old Chinese) was not only different from Zhōu Chinese (*i.e.* Early Old Chinese) but actually closer to Chǔ Chinese (*i.e.* the old language of the *Elegies of Chǔ*). Chǔ people were probably Hmong-Mien speakers, although the ruling class was undoubtedly Chinese-speaking. Middle Chinese is phonologically, typologically and lexically different from the Tibeto-Burman-looking Zhōu Chinese, let alone Shāng Chinese, and does not necessarily derive *directly* from the Old Chinese of the earliest classics (perhaps it comes from Chǔ-Hàn Chinese?). In this author’s opinion, the situation is, more or less, identical to the situation of India or of Italy and Greece. The Romans in Latium were undeniably speaking Latin after the Latino-Faliscan conquest of the Italian Peninsula. Today the people of Rome are speaking, nonetheless, a language which we now call ‘Italian.’ We have opted to use this term because vulgar Latin has evolved into a number of different “languages,” and not only into Italian, but this is a scholarly convenience. This is probably what happened in China. This view was somewhat presaged by Professor Branner, who writes that “early Chinese, the language of the classical texts, is perhaps truly a form of Tibeto-Burman; but medieval and modern Chinese belong to their own, distinct language family” (Branner 2000: 164). While this writer does not claim that Old Chinese and Middle Chinese are not related (they *are* and this fact is undeniable), we may agree on the fact that “the transition between early and medieval Chinese involved linguistic changes so fundamental that, for purposes of formal classification, we should consider the two types to have been different languages” (*Ibid.*).

⁷ Elsewhere, Sagart (2011c) cites few PST reconstructions, but they are too few and not easily verifiable. It is not clear why Proto-Tibeto-Burmese (PTB) **raŋ* < PST **gRaŋ* ‘chest,’ if not merely to account for PAN **baRaŋ*. Similarly, why PTB **kraŋ* < PST **t-gRaŋ*?

In addition, we must say that PAN was probably spoken during the sixth or fifth millennium BCE, while OC (which this author strongly suggests to rename as Zhōu Chinese) was spoken, more or less, in the second millennium BCE. This leaves a gap of three or two millennia after the purported split of Sinitic from its ancestor proto-language. Needless to say, many things may have happened in the meantime. Even assuming that the two language families are genetically related, one might expect the effects of coincidence, lookalikes and Wanderwörter to distort the evidence and create false positives. What if PAN and PST true commonalities had mutated separately beyond the point of being recognised as more akin to each other? Of course we cannot dismiss the entire STAN hypothesis with the *argumentum ex silentio*, but this is a serious problem which needs to be taken into consideration.

Furthermore, setting aside the fact that some Austronesianists (Li 1995, Blust 1995) have pointed out that PAN reconstructions do not belong to more general accepted reconstructed forms, another serious problem in this 61-word list is represented by the complete absence of numerals, pronouns, and basic kin terms.⁸ It is true that personal pronouns may have few marked sounds (and thus phonetic resemblance may just be fortuitous), but to make a point in favour of a genetic relationship without so much as including body parts, basic numerals and pronouns, in the absence of paradigmatic morphology, is something the present writer has not seen in serious linguistic works in many years of experience.⁹ Sagart's work is much more serious than a mere amateurish venture, therefore he should widen his circle of evidence and include more stable basic vocabulary.

There is another point which is likely to raise some eyebrows among the most conservative historical linguists. Sagart consistently compares OC monosyllables or sesquisyllables to the second part of AN disyllables (more rarely trisyllables). It is true that reconstructions fairly frequently involve tailoring or theoretical abstractions, but the manipulation of real language data is not methodologically admissible in comparative linguistics:

[T]he criteria which have usually been considered necessary for a good etymology are very strict, even though there may seem to be a high a priori probability of relationship when similar words in languages known to be related are compared. In the case of lexical comparisons it is necessary to account for *the whole word in the descendant languages, not just an arbitrarily segmented 'root,'* and the reconstructed ancestral form must be a complete word (emphasis added). [...] it is obvious that much stricter criteria must be applied to word-comparison between language whose relationship is in question. (Goddard 1975: 254–5, also quoted in Campbell 2003).

In Sagart (2005), instead, we still find comparisons between PAN **-qem* and OC **bʔim* (cloud, cloudy), or between PAN **-taq* and OC **athaʔ* (earth), PAN **-zem* and OC **btshimʔ* (sleep), etc. Leaving aside the fact that this method has been proven inadequate in comparative linguistics, some etymologies need to be adjusted. As the present writer's expertise does not lie in Austronesian linguistics, this paper is not in a position to evaluate, let alone falsify Sagart's hypotheses concerning PAN or AN forms. Hence, this paper will refrain from offering judgements about Austronesian.

⁸ Sagart cites very few body parts, most of which are strange and problematic (*i.e.* there are terms such as 'palm' but we do not find 'hand,' there are 'brain' and 'head' but no 'heart' and 'belly.' We find 'woman breast' but no 'chest' in general.). First, the character *fú* 扶 does not mean 'palm' but 'assist.' In the *Shuōwén* it is glossed as *zuǒ* 左, with *zuǒ* 左 being a graphic variant of *zuǒ* 佐 'assist.' Second, in **punuq* 'brain' we find an unaccounted for syllable **pu-*, while in **quluH* 'head' **qu-* is unmatched. **Kakay* 'foot' finds "correspondence" only with Burmese, not with Chinese.

⁹ It is well known since the times of Johannes de Laet (1581–1649) that basic vocabulary consists of numbers one to ten, kinship terms, body parts and natural terms.

A look to etymologies and lexical comparisons

In historical linguistics, the very notion of ‘basic vocabulary’ has played a major role as one source of evidence for genetic relationship. Of course, lexical similarities are not enough, and the testimony of morphological processes and sound correspondences is also required. Etymology, intended as the “historical search for earlier stages in languages and the origin of words” (Campbell, Poser 2008: 15), is also an important tool to establish linguistic relationship. During his academic career, Sagart (especially 1999) has provided many valid etymologies for OC. Nevertheless, in his lexical comparison between OC and PAN (2005), we find some far-fetched etymologies which cannot throw light on sound correspondences. A few examples are given below:

(1) *yīn* 陰: First of all, *yīn* 陰 does not mean ‘cloud,’ like in modern compounds, e.g., *yīn tiān* ‘cloudy day,’ but ‘dark.’ The *Shuōwén* glosses it as *àn* 闇, a graphic variant of the homophonous character *àn* 暗 ‘dark.’ According to the *Shì míng*, a Chinese dictionary compiled during Eastern Hàn times (25–220 AD) which employed phonological glosses, *yīn* 陰 has the same meaning of *yīn* 蔭 ‘shade.’ The normal and neutral OC equivalent for ‘cloud’ is *yún* 云, and a comparison with it would make far better sense, because *yīn* 陰 appears to be, instead, a descriptive predicate referring to shade.

(2) *dàn* 黧: Sagart (2005:164) compares PAN **-lem* with OC **^alīm?* ‘dark.’ The first problem which this author finds in this comparison lies in the semantic domain. The Chinese character *dàn* 黧 ‘dark’ is glossed as *sāngshèn zhī hēi* 桑葚之黑 ‘mulberry black.’ Another explanation is offered by the *Guǎngyùn*, where we find that the character means ‘dark cloud’ (*yún hēi* 雲黑). The second problem is that this rather rare character, which to the best of this author’s knowledge appeared for the first time in the spurious part of *Zhuāngzǐ* (3rd century BC), is hardly considerable a “basic lexeme.” The third problem is that we are presented with an unaccounted-for segment (if we exclude that another part of the PAN word has not been compared), i.e. OC **-?* vs PAN **-0?* In the opinion of the present writer, this could be instead a good place to suggest that, perhaps, the words *dàn* 黧 and *shèn* 葚 are etymologically related (they share the same phonetic component and essentially differ by type A/B syllable, which possibly reflects an old morphological connection). If so, it is most likely that this word for ‘dark’ simply goes back to ‘mulberry,’ further complicating Sagart’s Austronesian comparison. It must be remembered that this would not be the only case of a colour term which seems to have emerged from an object or from a surface property. For instance, the Sanskrit term for ‘red,’ *rudhirá* (< PIE **h₁rud^hrós*), is applied equally to the colour ‘red’ and ‘blood,’ suggesting that, perhaps, this colour term emerged from the word for blood. This could also be true for the words *dàn* 黧 and *shèn* 葚.

(3) *fú* 扶: As stated above, the character *fú* 扶 (OC: **m-[p](r)a*) does not mean ‘palm’ but ‘assist.’ In the *Shuōwén* it is glossed as *zuǒ* 左, with *zuǒ* 左 being a graphic variant of *zuǒ* 佐 ‘assist.’ Sagart (2009) insists that this word is cognate to Benedict’s PTB **pa > pwa* ‘palm,’ which makes it a Sino-Tibetan word, but this writer still remains unconvinced. The neutral OC equivalent for ‘palm’ is *zhǎng* 掌 (OC: **[k.t]aŋ?*), and is evidently unrelated to PAN **dapa*.

These few errors show nonetheless a certain negligence on Sagart’s part in his investigation of textual evidence. This writer thinks that the few remarks above suffice to demonstrate that a part of the data provided by Sagart (2005) sometimes contains rarely used or polysemantic words that should not be considered “basic lexicon”; at other times, the alleged cognancy appears to be semantically loose or imprecise.

A closer look to stan morphology and sound correspondences

In his famous work on comparative linguistics, Georg von der Gabelentz (1891: 168) stated that “Sprachvergleichung ohne Lautvergleichung ist gedankenlose Spielerei.” Every competent linguist agrees wholeheartedly with this statement, and so does Sagart (2005). Unfortunately, some of the sound correspondences which should prove in detail the existence of the STAN superphylum are not as regular as they might seem from a first look. It is true that, in Sagart (2005), sound correspondences appear to be generally genuine and quite regular, but there are also problems which deserve a mention (Cf. Tab. 1):

(1) OC glottal stop /ʔ/ seems to correspond to PAN *-q, *H₁ and *H₂, but sometimes it is unmatched, e.g., PAN *qiCeluR : OC *^aCə-lo[r]ʔ ‘egg,’ PAN *-lem : OC *^alimʔ ‘dark,’ PAN *di : OC *^bdiʔ ‘this,’ PAN *-danʔ : OC *^bdrangʔ ‘old,’¹⁰ PAN *-zem : OC *^btshimʔ ‘sleep,’ PAN *-kul : OC *^b[k](r)o[n]ʔ ‘curled,’ etc.;

(2) PAN velar nasal /ŋ/ in one case corresponds to OC *-ŋ and *-k without explained conditions, e.g. PAN *(q)uRung : OC *^ak-rok ‘horn,’ PAN *beCenʔ ‘foxtail millet’ : OC *^btsik ‘*Setaria italica*’; PAN *siNanʔ : OC *^blang ‘sunlight,’ etc.;

(3) PAN final syllable initials *-N- and *-l- correspond to OC *l- and *(h)l- without explained conditions, e.g. PAN *siNanʔ : OC *^blang ‘sunlight,’ PAN *daNum ‘water’ : OC *^bt-hlimʔ ‘liquid,’ PAN *b[e]lung : OC *^along ‘cave,’ PAN *bulay : OC *^bm-la[r] ‘snake,’ PAN *quluH₁ : OC *^bhluʔ ‘head,’ etc.;

(4) as argued before, AN roots occur only in combination with an initial syllable, therefore it is questionable to compare only the last syllable of PAN with the entire OC word, e.g. PAN *kupit : OC *^apit ‘close.’

Some points deserve a few remarks. First, the latest version of OC reconstruction (Baxter, Sagart 2014) seem to reject some of Sagart’s earlier proposals. For example, ‘far’ is reconstructed — rightly in this author’s opinion — with a final glottal stop which is unmatched in PAN. Second, if we accept the fact that Type A syllables had pharyngealised initials, then we must be aware that we are dealing with two different phonemes: thus, PAN *t- may correspond to OC *t-, *t^h-, *t^ʕ- and *t^{hʕ}-. This might be plausible, and this author does not exclude a priori these sound correspondences, but we must remember that plausibility by no means constitute proof (*a posse ad esse non valet consequentia*), and that the alleged sound correspondence, though somewhat regular, is quite unusual. Third, the Chinese character *hán* 含 ‘to hold in mouth’ deserves a special mention. This polyseme is reconstructed with a loosely attached iambic prefix *Cə- plus another *-m- affix before the initial consonant *-k^ʕ- which find no internal support. The new reconstruction undoubtedly matches much more the reconstructed PAN form,¹¹ however comparative evidence should not take the place of internal reconstruction. In historical linguistics it is not always recommendable to reconstruct morphemic boundaries in the language X on the basis of the same or of a similar morpheme which can be found on its relative language Y, let alone when the genetic relationship has yet to be established.

¹⁰ A reviewer has pointed out to me that the OC *-r- infix in this reconstruction is problematic, and that the word *danʔ matches much more the PAN string *-danʔ. The presence of an infix in OC seems to be confirmed by the dialects of Xiàmén and CháoZhōu, where this word is respectively pronounced as *tiŋ*⁶ and *tsian*⁴. This writer finds no elements here that would require us to reconsider Baxter and Sagart’s reconstruction.

¹¹ Another very important point needs to be clarified: in many cases striking similarities prove nothing. In the meantime, there are cases where real cognancy has been rendered almost unrecognisable based on today’s word shape by the accumulated multiplicity of sound changes. For example, Armenian *erku*, Italian *due* and German *zwei* ‘two’ share a common origin, despite their strikingly different shape. This is not to imply that Sagart is not aware of this fact, but that striking similarities between languages do not automatically imply genetic relationship.

Gloss	PAN	OC (Sagart 2005)	OC (Baxter & Sagrt 2014)
body hair	*gumuN	* ^b mu[r]	*mr[ə][r]
egg (dialectal)	*qiCelur	* ^a Cə-lo[r]?	*k.r ^f or?
snake	*bulay	* ^b m-la[r]	*Cə.lAj
sunlight	*siNan	* ^b lan	*lan
water	*daNum	* ^b t-hlim? (liquid)	* ^t [ə]m?
flow	*qaLur	* ^b hlu[r] (water)	*s.tur
horn	*(q)uRung	* ^a k-rok	*C.[k] ^f rok
chew	*paqpaq	* ^a m-paʔ-s	*[b] ^f ras
cut off	*[p,b]utul	* ^a to[r,n]	* ^t onʔ-s
far	*ma-dwiN	* ^b wa[r,n]	*C.ɔwanʔ
high	*-kaw	* ^a kaw	*Cə.k ^f aw
hold in mouth	*gemgem	* ^a gim (in mouth)	*Cə-m-k ^f [ə]m-s
curled	*-kul	* ^b N-k(h)ro[r,n]	*[k](r)o[n]?
speak	*kawaS	* ^a m-kw-r-at-s	*[g] ^{wi} rat-s
bone	*kukut	* ^a kut	*k ^f ut
head/brain	*punuq	* ^a nu?	*n ^f [u]?
elbow	*siku(H ₂)	* ^b t-r-ku?	*t-[k]<r>u?
hot	*qa(i)nget	* ^b nget	*C.nat/C.ɲet
put together	*pulung	* ^a long	*l ^f oŋ
ruin	*r[i]bas	* ^b bet-s	*[b]e[t]-s
chicken	*kuka	* ^a ke	*k ^f e
robe	*sabuk	* ^b buk	*[b]ək
broom	*CapuH ₁	* ^b t-pu?	*[t.p]ə?
Setaria	*beCeng	* ^b tsik	*[ts]ək
paddy/rice/grain	*Sumay	* ^a mij?	*(C.)m ^f [e]j?
stopper	*seŋseŋ	* ^a sik	*[s] ^f ək

Tab. 1. Unmatched sound correspondences between PAN and OC

Sagart's discussion on shared morphology, though very interesting — and in some cases even eye-opening — sometimes seems far-fetched. PAN nominaliser and goal focus marker *-ən* finds correspondence only in Tibetan and Lepcha nominalising suffix *-n*, there is nothing whatsoever similar in Chinese. It is very interesting, instead, that Early Middle Chinese had contrasting pairs of transitive verbs with voiceless stop initials versus intransitive verbs with voiced stop initials, such as 斷 **^twan* 'to cut' and 斷 **^dwan*¹² 'broken off.' EMC intransitive voicing is thought to reflect an OC prenasalised prefix **N-* (Sagart 1994, 1999, 2003, 2005; Baxter, Sagart 2014). This seems to correspond to PAN actor focus prefix and infix **m-/m-*.¹³ Leaving aside

¹² This author utterly rejects the usual practice, which seems to date back to Karlgren (1889–1978), of omitting asterisks with Early Middle Chinese forms. Even though the transcribed Late Middle Chinese (LMC) systems are coherent and agree with each other, EMC phonological system still remains highly theoretical, for it represents the literary reading of characters according to an assemblage of northern and southern dialects from various epochs. Hence, it represents a *mixtum compositum* and not a real language. On the other hand, the alphabetic and segmental scripts of written Old Tibetan can phonetically attest Late Middle Chinese forms, which thus require no asterisks.

¹³ The parallel is surely very suggestive, but it is so if and only if we remain confined within the framework of Sagart's hypotheses. In other reconstruction systems, such as Baxter (1992) or Starostin (1989a), there is no **N-*

the fact that Blust (1995: 287) finds problematic the correspondence between PAN “stative/attributive” marker **ma-* and OC transitive/intransitive prefix **N-*, it must be remembered that the transitive and intransitive prefixes in Sino-Tibetan do *not* always occupy the same slots of PAN actor focus prefixes and infixes. Furthermore, it must be stressed that we are dealing with single phonemes for both PAN (**m-/m-*) and OC (**N-*), and thus the risk of chance resemblance is dramatically high, especially when these isolated morphological markers (which can also be borrowed) are not integrated into a paradigmatic series.¹⁴ Moreover, we need to remember that grammatical affixes tend to utilise only a subset of consonants and vowels, and typically this subset is made up of less marked segments (Campbell, Poser 2008: 189). In particular, as Maddieson (1984: 70) informs us, nasals are often found in grammatical morphemes because they “are rarely subject to confusion with other types of consonants” (quoted also in Campbell, Poser 2008: 218). Comparisons between PAN instrumental/beneficiary focus prefix **si-* and ST valency-increasing **s-* is instead very suggestive. Nevertheless, we need to remember that the use of prefixes is a typologically commonplace, and that they can be easily explained in terms of areal typology, contacts or mere accident (*cf.* Tab. 2).

In brief, the evidence presented in favour of STAN is circumstantial and can easily be explained by non-genetic factors, therefore more effort is needed to guarantee a more solid plausibility for the proposed genetic relationship.

pronominal affix	Proto-Eastern Miwokan	Late Common IE
	<i>Declarative suffixes</i>	<i>Secondary affixes (active)</i>
1 sg	*-m	*-m
2 sg	*-s̥	*-s
3 sg	*-∅	*-t < **∅
1 pl	*-mas̥	*-me(s)/-mo(s)
2 pl	*-to-k	*-te

Tab. 2. Coincidence between Proto-Eastern-Miwokan and Late Common IE (quoted from Campbell, Poser 2008: 188).

To sum up, Sagart’s (2005, 2011a) lexical comparisons, sound correspondences and shared morphological derivations present no robust evidence which would require us to reconsider our current understanding of both the Sino-Tibetan and the Austronesian families.¹⁵

prefix we can compare with PAN **m-*. For example, what in Sagart (and later in Baxter and Sagart 2014) has been reconstructed as **N-q^{wh}ra* 華 ‘to flower,’ in Baxter (1992) is **wra*, in Schuessler (2009) is **(g)wrâ*, and in Norman (2014) is **gwrap*. They are all respectable reconstruction systems, and yet in none of them the prenasalised **N-* infix is reconstructed. Even those scholars who are broadly sympathetic with Sagart’s theories are not persuaded by the **N-* infix (Mei 2012: 14-16). In addition, transitivity is normally expressed syntactically without morphology.

¹⁴ In addition, there is another issue which does deserve a brief comment. The reduction to monosyllables and the maintenance of prefixation and infixation are serious issues, but just how short these issues are argued in Sagart (2005) is incredibly stunning. This writer thinks that Sagart’s paper, with its fifteen pages to fill, should have dwelt a bit more on this serious problem, which instead is crammed into a mere eleven lines. In short, Sagart (1993, 2005) argues that only polysyllabic words were affected by phonological erosion, so that monosyllables could then act as refuges for prefixes and infixes. The present writer knows no other attested case of phonological erosion where the sound change was “aware” of morpheme boundaries. To sum up, this writer finds this *ad-usum-delphini* explanation of a serious phonological issue—which this author had pinned his hopes on—quite disappointing.

¹⁵ A colleague from Taipei has drawn my attention toward another 210-words list available on the internet: <https://abvd.shh.mpg.de/austronesian/language.php?id=331> (University of Auckland, New Zealand). Setting aside that we find “nursery words” such as ‘mother’ and ‘father,’ this author finds several problems with the semantic

STAN *Urheimat* and Proto-STAN speakers

As any good textbook for historical linguistics will teach with ample and documented historical examples, language reconstruction is not a mere teleological exercise, and comparative linguistics does not develop in a vacuum. Every language has its own history, and the speakers of that language are culturally, socially and politically connected with the history of that language. Therefore, when the linguists start their investigation with the language of a people, they also need to find out when and where this people started their journey before setting in a given part of the world. In order to complete this task, collective work and dialogue between specialists from different domains of academic research is needed.¹⁶

Sagart presents many competing and stimulating ideas, which, though bring him into conflict with many Austronesianists, deserve special attention. In his earliest proposals (Sagart 1993: 2), Sagart first suggested that STAN should be associated with the Dàwènkǒu culture (4100–2600 BC) of today Shāndōng province. However, more recently, Sagart (2011) argues that the Yǎngsháo culture (5000–300 BC), a Neolithic culture which existed extensively along the Yellow River, corresponds to Proto-STAN homeland. It is interesting to see how Sagart's depicted picture fits in with different approaches. This writer thinks that Sagart's explanation of the historical and archaeological facts is very interesting but, unfortunately, not without problems.

First, it contrasts with the general scenario about the origin of PST people. The PST *Urheimat* seems to have been somewhere on the Himalayan plateau, where the great rivers of East and South-East Asia have their source (Matisoff STEDT, Blench & Post 2013). Furthermore, the Shāng dynasty (1600–1046 BC) probably originated somewhere near the Èrlǐgǎng cultural frontier, while the Zhōu dynasty (1046–771 BC) is associated with the Guāngshè culture of the second millennium BCE.¹⁷ Now, if we accept Sagart's scenario, we are forced to accept the fact

of many words. A few examples will suffice to confirm this point. The character *jiǎo* 腳 means 'foot' only in modern Chinese. The *Shuōwén* glosses it as *jìng* 脛 'calf of leg,' or even 'belly.' In the 64th Chapter of *Lǎozǐ* there is a phrase whose meaning in English is "a travel of a thousand *li* starts with a single step (literally "under the foot")," which in Chinese is rendered as *qiānlǐ zhī xíng shǐ yú zúxià* 千里之行始於足下 and not as **qiānlǐ zhī xíng shǐ yú jiǎoxià* *千里之行始於腳下. The character *kòu* 寇, arguably a quite rare word, doesn't mean 'to steal' in OC, which is rather expressed with the older and more common form *dào* 盜, but 'to invade.' In the *Yānzǐ chūnqū* [Annals of the Master Yan], a work which dates back no earlier than the III Century BCE, there is a passage which says: "The people who grow in Qi do not steal, they become burglars when they are [in the State of] Chu." In Chinese this phrase is written as *mín shēngchǎng yú Qí bù dào, rù Chǔ zé dào* 民生長於齊不盜, 入楚則盜. It is clear that even in Late Old Chinese the dominant form for 'to steal' was still *dào* 盜 and not *kòu* 寇. In other words, this author finds the alleged cognancy highly questionable, for it is made up of obscure, polysemantic, rare and imprecise words.

¹⁶ This, of course, does not imply that genetic relationship between languages can be established on extralinguistic grounds. Rather, as Gabelentz (1891: 157) has remarked: "Das einzig untrügliche Mittel, eine Verwandtschaft zu erkennen liegt in den Sprachen selbst."

¹⁷ Moreover, we must make one point clear at the outset: the study of Ancient China is perennially distorted by the desire, natural enough itself, of discovering not the origin of the civilisations *in* China, but the origin of the civilisation *of* China (Bagley 1999: 135), if this imprecise and rather modern term really means anything. We need to be aware that a Neolithic site in present-day Chinese territory was not necessarily the homeland of the so-called Huáxià people, an entity never well-defined but always imagined in quasi-historicistic terms drawn from later periods, or of some Chinese-speaking people. However, many Chinese archaeologists, instead of working back from definite evidence, starts with the assumption of the uninterrupted cultural unity of China and try to force all the archaeological discoveries to fit in with the traditional narrative (Jia 1980; Wu 1989; Liu & Xiu, 2007; Wu *et al.* 2016). It follows that we are left with the strong impression that they are searching continuously and desperately for any scrap of data that can be taken as evidence to support the traditional view. This is not to imply that Sagart was driven into this *cul-de-sac*, but that the archaeological picture in China is not as clear as the general audience might imagine.

that Proto-STAN inhabitants spoke their proto-language in the Hénán, Shǎnxī and Shānxī regions around 5000 BCE, and that PAN began to break up within the Taiwan Straits a millennium later, while PST people, essentially at the same time, break up into two branches, one of which (Proto-Sinitic) started their journey from the Himalayas and came back again in the vicinity of modern Zhèngzhōu, Pánlóngchéng and Yǎnshī cities, basically where it all began. This author may be not alone in thinking that this scenario is highly improbable. Sagart's solution to this problem is as follows: after the break up of Proto-STAN, the Western group (*i.e.* PST) developed *in situ*. Nevertheless, despite Sagart's efforts to link PST to Yǎngsháo culture, this scenario remains unproven. Yet, how such identifications can be reasonably made in first place, given that we have no linguistic evidence of PAN, PST and Proto-Sinitic speakers anywhere on the Asian continent until, respectively, a millennium and three millennia later, we do not learn. Moreover, the bases for distinguishing Proto-STAN from PAN or PST are also unknown to us. Although Sagart's scenario may be true, from the viewpoints of linguistics and human migrations this interpretation of the Proto-STAN *Urheimat* remains fundamentally problematic and tells us little about linguistic interactions and further migrations of PAN and PST speakers.

Sagart (2005, 2011a, 2016), in addition to archaeological researches, cites also anthropological and genetic studies in support of his thesis. In this author's opinion, these researches may be a useful tool, since they can — and they do — throw some light on prehistoric migrations. Nevertheless, we must remember that language is independent of phenotypic traits, because human groups change their languages for different reasons in different epochs, and biological findings have been irrelevant, in detecting language relationship, throughout the entire history of linguistics. To sum up, although genetic studies may wink at Sagart's hypotheses (*cf.* Ko et al. 2014),¹⁸ what this discipline is able to tell us is that contemporary populations do show some connections based on genetic flow, which, with all due respect to these serious scholars, we already knew. What instead they are *not* able to tell us, on a firm footing, is that the ancestors of these populations were speaking the same language more than seven millennia ago. Nevertheless, based on the origin of the haplogroup E, which — Ko et al. (2014) say — is developed out of the haplogroup M9 on the south-east coast near Fúzhōu (c. 8,136–10,933 years ago), Sagart (2016) claims that this scenario directly supports the STAN theory. This writer may be mistaken, but what Ko et al.'s genetic assumption seems to confirm is, instead, the scenario depicted by Bellwood (1997: 241–42):

During the late fifth or fourth millennium BC *colonists from the mainland of southern China (probably Zhejiang or Fujian) settled Taiwan* (emphasis added) [...] During the third millennium BC colonists moved into Luzon, and the Malayo-Polynesian subgroup now began its separation from the other primary subgroups of Austronesian which remained on Taiwan [...] By at least 2000 BC Proto-Malayo-Polynesian began to break up, probably with settlement expanding in various directions into the southern Philippines, Borneo, Sulawesi and the Moluccas.

¹⁸ “At around the time when haplogroup E developed from M9 in the population that later carried it into Taiwan, the surrounding regions gave rise to M9a lineages that are now carried by Sinitic speakers. Thus, the Han, Liangdao Man, and Formosan haplogroup M9a/E lineages can be traced to an ancestral M9 mtDNA lineage (Figure 3). Additionally, the Tibetans have a high frequency of M9a lineages that is shown to have coalesced during the Neolithic and there is a hypothesized linguistic link between Sino-Tibetan and Austronesian languages” (Ko et al. 2014: 239). The problem with the use of genetic studies in linguistics is that they can be interpreted as a Rorschach test, where facts are analysed using complex algorithms, intuitive and generally preconceived interpretations, or both. For example, Oppenheimer (2004) used mtDNA molecular clock calculations to conclude that Polynesians are derived from Palaeolithic eastern Indonesians, denying thus the generally accepted “out of Taiwan” theory. This scenario was rejected by Bellwood (2008).

The evidence of a migration from the Shāndōng Peninsula to south-east China coasts is lacking. The only evidence Sagart has is that foxtail millet and proso millet (two of the main Austronesian cereals) are unknown archaeologically in the Yangtze valley and south of it. This view is only partially correct. First, the earliest domestication of foxtail millet took place in the middle of Císhān culture (6500–5000 BC), which originated in modern Héběi (Stevens *et al.* 2016). Second, foxtail millet is also thought to have reached Primorsky Krai (Eastern Siberia) around 3620 BCE (Kuzmin, 2013): should we include Eskimo-Aleut languages¹⁹ into STAN, as well? Third, proso millet (*panicum miliaceum*) may have been domesticated independently in both Transcaucasian areas and East-Asian areas. Thus, the so-called “northern Chinese look” of early Austronesian agriculture does not prove convincingly the STAN hypothesis.

Furthermore, in support of his thesis, Sagart pointed out that those Neolithic cultures, which he identified as Proto-STAN speakers, practised tooth avulsion (2011a: 147). This is extremely stimulating but, in this author’s opinion, more attention to the anthropological works of the specialists of the field would have saved him from establishing a cause-and-effect relationship. Tooth avulsion is a worldwide practice which is still largely diffused today in Sub-Saharan Africa.²⁰ For these reasons, all these uncertainties give the present writer the impression that the STAN hypothesis is still standing infirmly on one leg.

Competing hypothesis: STAN vs Sino-Indo-European

Another criticism that has been levelled against STAN (Blust 1995) and that has apparently remained unanswered is that of competing hypothesis, which involves mutual exclusivity (*i.e.* if we accept, say, Sino-Indo-European or Sino-Caucasian, then STAN is an invalid superphylum and vice versa). In the specific case of Chinese, a panoply concerning a wider affiliation of Sinitic has been proposed at least since the 19th century. Despite the interesting and quite brave attempts of grouping Sinitic with other families there has been no general agreement on any of these stimulating or rather quite speculative proposals. For some scholars (*e.g.*, Beckwith 2002), the most promising proposal is the Sino-Indo-European hypothesis (Pulleyblank 1995, 1996), which is the taxonomic affiliation that argues for the connection between Indo-European and Chinese (or even Sino-Tibetan). Publications about this subject vary from amateurish attempts (Ulenbrook 1967, Chang 1988, Tan 2001, Zhōu 2002a, 2002b, Wei 2005b) to serious linguistic works (Pulleyblank 1996).²¹ Nevertheless, it will be briefly demonstrated in this paragraph that (1) all these attempts are situated in the same realm of non-demonstrable relationship, albeit some works are undoubtedly different in quality; (2) they are not better than the STAN hypothesis, actually much worse.²²

¹⁹ Eskimo-Aleut languages are now spoken in north-east Siberia. However, Vovin (2015) suggested that, since Northern Tungusic languages have loanwords which are not found in Southern Tungusic, the Eskimo-Aleut was once much more widely spoken in eastern Siberia. This author agrees wholeheartedly with Vovin’s view.

²⁰ In many cultures of Ethiopia, Sudan and East Africa, tooth avulsion appears to be related to the medical beliefs of those cultures, since canine, for example, are believed to be responsible for diarrhoea, fever and other diseases in children. For further understanding, the reader can consult this very insightful on-line paper by Mutai *et al.* 2010: <https://www.ncbi.nlm.nih.gov/m/pubmed/20476714/>. For a more technical discussion see also Schwartz (1946).

²¹ To be fair, both Chinese and Tibetan were wrongly regarded as Sanskrit, and thus as member of the Indo-European family, by Sir William Jones (1799a, 1799b).

²² The genetic relationship of Chinese and Indo-European has been proposed in several articles published on *Sino-Platonic Papers* (SPP) — a journal generally worth of praise for its critical power which unfortunately has

The Sino-Indo-European superphylum was mainly advocated by Pulleyblank (1995, 1996) on the basis of a somewhat identical ablaut system, *viz.* EMC **a/*ə*. Leaving aside the fact that Pulleyblank’s EMC ablaut system is based on the erroneous interpretation of the rather obscure terms of *nèizhuǎn* ‘inner turn’ and *wàizhuǎn* ‘outer turn,’²³ we must remember that ablaut and vowel mutations are not typologically unusual features and are, instead, diffused in many worldwide languages, such as Semitic languages, *etc.* (Campbell 1997: 196). Long before Pulleyblank, Sapir (1915, 1925) had already noted similarities in ablaut and causative prefixes between Na-Dene and Sino-Tibetan. Nevertheless, these vocalic alternations could also be found in Sapir’s “Penutian”, a distinct language family. Therefore, the sort of structural “evidences” both Sapir (1915, 1925) and Pulleyblank (1995, 1996) had in mind can be easily explained as independently developed typological similarities rather than inherited traits.²⁴

Many etyma discussed by Pulleyblank (1996) were accepted and expanded, though with very different reconstructions, by Beckwith (2002: 149–150). Unfortunately, Beckwith’s reconstructions are not methodologically explicit. For example, he reconstructs forms such as MOC **mîna* < **mbîna* < **C^mbêna* < PChi **Twêna* for *rǔ* 汝 ‘you’ (OC: **naʔ*), or MOC **mîna* < **mbîna* < **C^mbêna* < EOC **C^wêna* for *nǚ* 女 ‘woman’ (OC: **nraʔ*), or MOC dial. **rmay* < **mray* < EOC **marê* for *hǎi* 海 ‘sea’ (OC: **m^hǎʔ*). Beckwith’s reconstructions routinely omit the final glottal stop [-ʔ] which is strongly supported internally by many Sinitic languages. As it can be seen, all the real proposals of Sino-Indo-European that Beckwith might have had, unfortunately, are so buried under masses of obscure and unsupported reconstructions that it is quite hard to ferret them out.

given these quasi-scholarly works a positive judgement that undoubtedly suffers from excessive indulgence — by Zhōu (2002b, 2003) and Wei (2005a, 2005b). Nearly all these scholars say about Indo-European (IE)—a field in which they are evidently not at home—is false. The first methodical mistake which these scholars commit is that, in their attempt of inventing proofs where they do not exist, they do not compare Proto-Indo-European (PIE) with OC, but one of its many daughter languages with OC or with unsupported, obscure and invented proto-Chinese forms. They even claim to have discovered regular sound changes, such as OC **m-* and PIE **m-*, ignoring the fact that in PIE the **m-* stem is an active declension. Furthermore, when Zhōu compares Chinese *xùn* 殉 (OC: **s.[ç]ʷi[n]-s*) [2002b: 4], which means ‘be buried alive with the dead’ and not ‘die’ as he writes, with Latin *mor-t-* (and not **mort-*), he is ignoring the fact that Latin *mor-t-* ‘death’ (and not ‘die’) derives from PIE **m_ṛ-tó-*, which in turn is obtained by back derivation from the adjectival form **m_ṛ-to-* ‘immortal.’ Second, Zhōu’s own theoretical views on IE studies seem to have congealed somewhere in the beginning of the 19th century, when it would have been quite difficult to find an Indo-Europeanist who would have accepted any need to reconstruct the laryngeal consonants. In part, the general resistance by those scholars outside IE linguistics to accepting laryngeals is excusable in view of their complexity and their “algebraic appearance,” but a great amount of IE comparative linguistics is practically impossible without them. What Wei (2005b) adds to Zhōu’s dissertations is nothing but a further list of lexical comparisons riddled with incorrect etymologies, *Wortakrobatik* and other blatant errors. Given their lack of understanding of IE linguistics (see, *e.g.*, Wei 2005b: 1–72, Zhōu 2002b: 6, 2003: 3–17), it would seem that trying to make them understanding that, say, PIE **-t-* was not an isolated marker which can be paralleled with the OC entering tone final **-t* but is one of the various PIE verbal nouns (Beekes 1995: 249–251) would be as futile a task as explaining PIE verbal morphology to primary school students. If one thinks that it is important to stimulate younger scholars to take part in the discussion, then detecting these mistakes is of utmost importance, because the general audience may fall prey of these amateurish ventures into a field which should not allow the likes of Zhōu and Wei free range.

²³ The two terms do not indicate any quality of the vowel. See Lǐ (1994: 271–283), still the best reference on the subject.

²⁴ For a positive evaluation of the Sino-Dene hypothesis see Bengtson (1994). A relationship between Sino-Tibetan and Athabaskan was also advocated by Shafer (1952). For an evaluation of Shafer’s work, see Swadesh (1952a).

Be that as it may, there is one aspect of long-range comparisons which was nowhere mentioned by Beckwith and Pulleyblank, and that is never mentioned by many macro-comparativists who frequently invoke the Indo-European model in order to validate their reconstruction: workers in IE linguistics have demonstrated with abundant documentation that someone dealing with, say, Italic may benefit from someone working on Celtic or on Greek. On the contrary, a philologist working with classical documentation in Old Chinese will not find the solution to, say, a verb problem by looking at Latin, Slavonic or Tocharian.²⁵

Nevertheless, Beckwith (2002: 155) concludes his study on the subject in the following way:

In sum, it is certainly possible that Tibeto-Burman and Chinese are genetically related, but since the best-supported etymologies appear to be shared not only with Japanese-Koguryoic but also with Indo-European, it is likely that the relationship is either one of shared loan influence from the same donor or one of common descent from the same intrusive ancestor. *Further study is therefore needed to determine more precisely the history of the interrelationship of these four families* (emphasis added).

Absit omen!

Closing remarks

It is a truism in historical linguistics that non-relationship cannot be demonstrated. This author agrees wholeheartedly with this statement and, therefore, does not claim that Chinese (or Sino-Tibetan) and Austronesian *cannot* be related. Nevertheless, the genetic relationship between Sinitic and Austronesian still remains non-demonstrable due to the lack of common basic vocabulary, presence of unaccounted and unmatched segments and absence of *productive* sound correspondences.²⁶ Sagart's argumentations, though interesting and stimulating, are still unable to offer valid morphological comparisons, let alone *paradigmatic* morphological comparisons.²⁷ More emphasis is needed on morphology, both its overall structure and its idiosyncratic alternations. There are also other problematic issues, such as those concerning the STAN *Urheimat* and the competing hypotheses. This author, nonetheless, remains open-minded to STAN, and would like to suggest that the STAN controversy can be improved by abandoning emphasis on common etymologies and shifting the attention to (1) more solid predictive-productive morphological correspondences; (2) questions of chronology; (3) questions regarding the STAN homeland; (4) rejection of competing hypotheses.

²⁵ The present writer has learnt this thing from Alonso de la Fuente's review of M. Robbeets (2016).

²⁶ Sagart argues that AN *-ŋ regularly corresponds to OC *-ʔ, the source of the marked member among the two basic Chinese tones. However, those PAN examples which account for OC tonogenesis seem to be valid only for Chinese, while they add nothing to AN linguistics. Therefore, in this author's opinion, they cannot be considered productive sound correspondences.

²⁷ It is true, however, that morphosyntactic considerations are rarely invoked (or invoked in a quite tentative way) in support of the Sino-Tibetan hypothesis as well. The differences between Sinitic and Tibeto-Burman are quite striking and not always convincingly explained. It is up to the competent linguist to work within these divergences in both of these two categories, and to demonstrate that a genetic relationship is likely to exist between the two groups. Nevertheless, more efforts in reconstructing PST is needed. One of the most recent attempts (South Coblin 1986) is laudable but does contain methodological peculiarities. For example, far too many of South Coblin's "reconstructed" proto-forms consist of long string of phonemes which eventually merged into zero in the received daughter languages. The best treatment of the family, especially from the Tibeto-Burman side, is Matisoff (2003). Unfortunately, from the Sinitic side, Matisoff adopted Karlgren's system, which is obsolete and now quite inadequate. Curiously, in Greenberg's opinion, Sino-Tibetan seems to be "one of the most solidly based and universally accepted linguistic stock in the world" (Greenberg 1996: 134).

Literature

- Alonso de la Fuente, J. A. 2016. Martine Robbeets (2015): Diachrony of verb morphology. Japanese and the Trans-eurasian languages. *Diachronica* 33(4): 530–537.
- Bagley, Robert. 1999. Shang archaeology. In: M. Loewe, Edward L. Shaughnessy (eds.). *The Cambridge history of ancient China*: 124–231. Cambridge University Press.
- Baxter, William H. 1992. *A handbook of Old Chinese phonology*. Berlin: Walter de Gruyter.
- Baxter, William H., Laurent Sagart. 2014. *Old Chinese: a new reconstruction*. Oxford University Press.
- Beckwith, Christopher. 2002. The Sino-Tibetan problem. In: Christopher Beckwith. *Medieval Tibeto-Burman Languages*: 113–157. Leiden: Brill.
- Beekes, R. S. P. 1995. *Comparative Indo-European Linguistics. An Introduction*. Amsterdam: Benjamins.
- Bellwood, Peter. 1995. Austronesian prehistory in Southeast Asia: homeland, expansion and transformation. In: Peter Bellwood, J. Fox, D. Tryon (eds.) *The Austronesians*: 96–111. Canberra: Australian National University.
- Bellwood, Peter. 1997. *Prehistory of the Indo-Malaysian Archipelago* (2nd edition). Honolulu: University of Hawaii Press.
- Bellwood, Peter. 2005. *First farmers: The origins of agricultural societies*. Oxford: Blackwell.
- Bellwood, Peter. 2007. Southeast China and the Prehistory of the Austronesians. In: T. Jiao (ed.). *Lost Maritime Cultures: China and the Pacific*: 36–53. Hawaii: Bishop Museum Press.
- Bellwood, Peter. 2008. Austronesian cultural origins: out of Taiwan, via the Batanes Islands, and onwards to western Polynesia. In: Alicia Sanchez-Mazas, Roger Blench, Malcolm D. Ross, Ilia Peiros and Marie Lin (eds.) *Past human migrations in East Asia: matching archaeology, linguistics and genetics*: 23–39. London: Routledge.
- Bellwood, Peter. 2011. The checkered prehistory of rice movement southwards as a domesticated cereal – from the Yangzi to the equator. *Rice* 4(3–4): 93–103.
- Benedict, Paul K. 1991. Austric: an ‘extinct’ proto-language. J. H. C. S. Davidson (ed.). *Austroasiatic Languages: essays in honour of H. L. Shorto*: 7–11. London: Psychology Press.
- Bengtson, John D. 1994. Edward Sapir and the “Sino-Dene” Hypothesis. *Anthropological Science* 102(3): 207–230.
- Blench, Roger M. 2012. Almost everything you believed about the Austronesians isn’t true. In: Dominik Bonatz, Andreas Reinecke, Mai Lin Tjoa-Bonatz (eds.). *Crossing Borders: Selected Papers from the 13th International Conference of the European Association of Southeast Asian Archaeologists, Vol. 1*: 122–142. Singapore: NUS Press.
- Blust, Robert A. 1995a. The prehistory of the Austronesian-speaking peoples: a view from language. *Journal of World prehistory* 9(4): 453–510.
- Blust, Robert A. 1995b. An austronesianist looks at Sino-Austronesian. In: William S-Y. Wang (ed.). *The Ancestry of the Chinese Language. Journal of Chinese Linguistics Monograph Series 8*: 283–298. University Of California Press.
- Blust, Robert A. 1996. The Neogrammarian hypothesis and pandemic irregularity. In: Mark Durie, Malcolm Ross (eds.). *The Comparative Method reviewed: Regularity and irregularity in language change*: 135–156. New York / Oxford: Oxford University Press.
- Blust, Robert A. 2003. Three notes on early Austronesian morphology. *Oceanic Linguistics* 42(2): 438–478.
- Blust, Robert A. 2005a. Must sound change be linguistically motivated? *Diachronica* 22(2): 219–269.
- Blust, Robert A. 2005b. The Oceanic languages. *Oceanic Linguistics* 44(2): 544–558.
- Blust, Robert A. 2008. Is there a Bima-Sumba subgroup? *Oceanic Linguistics* 47(1): 45–113.
- Blust, Robert A. 2010. The greater north Borneo hypothesis. *Oceanic Linguistics* 49(1): 44–118.
- Blust, Robert A. 2013. *The Austronesian languages*. Canberra: Asia-Pacific Linguistics.
- Blust, Robert A. 2014. Some recent proposals concerning the classification of the Austronesian languages. *Oceanic Linguistics* 53(2): 300–391.
- Bopp, Franz. 1841. *Über die Verwandtschaft der malaisch-polynesischen Sprachen mit den indo-europäischen. Gelesen in der Akademie der Wissenschaften am 10. aug. und 10. dec. 1840*. Berlin: Dümmler.
- Branner, David P. 2000. *Problems in comparative Chinese dialectology: the classification of Miin and Hakka*. Berlin / New York: Walter de Gruyter.
- Campbell, Lyle. 1997. Genetic classification, typology, areal linguistics, language endangerment, and languages of the north Pacific rim. In: Osahito Miyaoka, Minoru Oshima (eds.). *Languages of the North Pacific Rim, Vol. 2*: 179–242. Kyoto University.
- Campbell, Lyle. 2003. How to show languages are related: Methods for distant genetic relationship. In: Brian D. Joseph, Richard D. Janda (eds.). *The Handbook of Historical Linguistics*: 262–282. Blackwell Publishing.

- Campbell, Lyle. 2013. *Historical linguistics*. Edinburgh University Press.
- Campbell, Lyle, William J. Poser. 2008. *Language classification. History and method*. Cambridge University Press.
- Chang, Tsung-Tung. 1988. Indo-European Vocabulary in Old Chinese. *Sino-Platonic Papers* 7: 1–54.
- Conrady, August. 1896. *Eine indochinesische Causativ-Denominativ-Bildung und ihr Zusammenhang mit den Tonaccenten: Ein Beitrag zur vergleichenden Grammatik der indochinesischen Sprachen, Insonderheit des tibetischen, barmanischen, siamesischen und chinesischen*. Leipzig: Otto Harrassowitz.
- Conrady, August. 1916. Eine merkwürdige Beziehung zwischen den austrischen und den indochinesischen Sprachen. In: *Aufsätze zur Kultur- und Sprachgeschichte vornehmlich des Orients: Ernst Kuhn zum 70. Geburtstag am 7. Februar, 1916*: 475–504. München: Verlag von M. & H. Marcus.
- Driem van, George. 2001. *Languages of the Himalayas: an ethnolinguistic handbook*. Leiden: Brill.
- Driem van, George. 2005. Sino-Austronesian vs. Sino-Caucasian, Sino-Bodic vs. Sino-Tibetan, and Tibeto-Burman as default theory. In: Yogendra Prasada Yadava, Govinda Bhattarai, Ram Raj Lohani, Balaram Prasain, Krishna Parajuli (eds.). *Contemporary issues in Nepalese linguistics*: 285–338. Kathmandu: Linguistic Society of Nepal.
- Edkins, Joseph. 1871. *China's Place in Philology: an Attempt to show that the Languages of Europe and Asia have a Common Origin*. London: Trübner & Co.
- Gabelentz von der, Georg. 1901. *Die Sprachwissenschaft*. Aufl. Leipzig.
- Goddard, Ives. 1975. Algonquian, Wiyot, and Yurok: Proving a distant genetic relationship. In: Dale M. Kinkade, Kenneth L. Hale, Otmar Werner (eds.). *Linguistics and anthropology: in honor of C. F. Voegelin*: 249–262. Lisse: The Peter de Ridder Press.
- Greenberg, Joseph H. 1996. In defense of Amerind. *International journal of American linguistics* 62(2): 131–164.
- Holman, E. W., C. H. Brown, S. Wichmann, A. Müller, V. Velupillai, H. Hammarström, S. Sauppe, H. Jung, D. Bakker, P. Brown, O. Belyaev. 2011. Automated dating of the world's language families based on lexical similarity. *Current Anthropology* 52(6): 841–875.
- Hudson, Mark J. 1999. Japanese and Austronesian: an archeological perspective on the proposed linguistic links. In: K. Omoto (ed.). *Interdisciplinary Perspectives on the Origins of the Japanese*: 267–279. Kyoto: International Research Center for Japanese Studies.
- Humboldt von, Wilhelm. 1836–1839. *Über die Kawi-Sprache auf der insel Java, nebst einer einleitung über die verschiedenheit des menschlichen sprachbaues und ihren einfluss auf die geistige entwicklung des menschengeschlechts*. Berlin: Druckerei der Königlichlichen Akademie der Wissenschaften.
- Jia, Lanpo. 1980. *Early man in China*. Beijing: Foreign Languages Press.
- Jones, Sir William. 1799a [delivered February 25, 1790]. Eighth anniversary discourse: on the Chinese. *Asiatick Researches* 2: 43–46.
- Jones, Sir William. 1799b [delivered February 23, 1792]. Ninth anniversary discourse: on the origin and families of nations. *Asiatick Researches* 3: 418–435.
- Kawamoto, Takao. 1977. Toward a comparative Japanese-Austronesian I. *Bulletin of Nara University Education* 26(1): 22–49.
- Kawamoto, Takao. 1978. *Minami kara kita Nihongo*. Tōkyō: Sanseidō.
- Kawamoto, Takao. 1980. *Nihongo no kenryū*. Tōkyō: Kōdansha.
- Kawamoto, Takao. 1984. Two sets of sound laws between Japanese and Austronesian. *Bulletin of Joetsu University of Education* 3: 31–50.
- Kawamoto, Takao. 1993. *Several similarities between Japanese and the Austronesian languages*. Paper presented at the Conference on Asia-Mainland/Austronesian Connections, University of Hawai'i at Mānoa, May 10–13.
- Kempler-Cohen, E. M. 2012. Austronesian cognates in Quechua – Part 1: Proto-Oceanic disyllabic cognates in Proto-Quechua. *The Philippine Journal of Linguistics* 43: 1–46.
- Kortlandt, Frederik. 2010. *Studies in Germanic, Indo-European and Indo-Uralic*. Brill / Rodopi.
- Krippes, Karl A. 1992. The genetic relationship between Japanese and Austronesian revisited. *Ural-Altische Jahrbücher* 64: 117–130.
- Ko, A. M. S., C. Y. Chen, Q. Fu, F. Delfin, M. Li, H. L. Chiu, M. Stoneking, Y. C. Ko. 2014. Early Austronesians: into and out of Taiwan. *The American Journal of Human Genetics* 94(3): 426–436.
- Kuzmin, Y. V. 2013. The beginnings of prehistoric agriculture in the Russian Far East: Current evidence and concepts. *Documenta Praehistorica* 40: 1–12.

- Li, Paul Jenkuei. 1995. Is Chinese genetically related to Austronesian? In: William S-Y. Wang (ed.). *The Ancestry of the Chinese Language. Journal of Chinese Linguistics Monograph Series 8*: 93–112. University Of California Press.
- Lǐ, Xīnkuí [李新魁]. 1994. *Lǐ Xīnkuí yǔyán xué lùnjí* 李新魁語言學論集 [Lǐ Xīnkuí's selected works in Linguistics]. Běijīng: Zhōnghuá shūjú.
- Liu, Li, Hong Xiu. 2007. Rethinking Erlitou: legend, history and Chinese archaeology. *Antiquity* 81(314): 886–901.
- Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge University Press.
- Matisoff, James A. 2003. *Handbook of Proto-Tibeto-Burman: system and philosophy of Sino-Tibetan reconstruction*. University of California Press.
- Mei, Tsu-Lin. 2012. The causative *s- and nominalizing *s- in Old Chinese and related matters in Proto-Sino-Tibetan. *Language and Linguistics* 13(1): 1–28.
- Mutai, J., E. Muniu, J. Sawe, J. Hassanali, P. Kibet, P. Wanzala. 2010. Socio-cultural practices of deciduous canine tooth bud removal among Maasai children. *International dental journal* 60(2): 94–98.
- Ness, Immanuel. 2014. *The global prehistory of human migration*. John Wiley & Sons.
- Norman, Jerry. 2014. A model for Chinese dialect evolution. In: Richard VanNess Simmons, Newell Ann Van Aucken (eds.). *Studies in Chinese and Sino-Tibetan linguistics: Dialect, phonology, transcription and text*: 1–26. Taipei: Institute of Linguistics, Academia Sinica.
- Pietrusewsky, Michael. 1994. Pacific-Asian relationships: a physical anthropological perspective. *Oceanic Linguistics* 33(2): 407–429.
- Pulleyblank, Edwin G. 1965. Close/open ablaut in Sino-Tibetan. *Lingua* 14: 230–240.
- Pulleyblank, Edwin G. 1991. The ganzhi as phonograms and their application to the calendar. *Early China* 16: 39–80.
- Pulleyblank, Edwin G. 1995a. Why Tocharians? *Journal of Indo-European Studies* 23(3–4): 415–430.
- Pulleyblank, Edwin G. 1995b. The historical and prehistorical relationships of Chinese. In: William S-Y. Wang (ed.). *The Ancestry of the Chinese Language. Journal of Chinese Linguistics Monograph Series 8*: 145–194. University Of California Press.
- Pulleyblank, Edwin G. 1996. Early contacts between Indo-Europeans and Chinese. *International Review of Chinese Linguistics* 1(1): 1–25.
- Reid, L. A. 1994. Morphological evidence for Austric. *Oceanic linguistics* 33(2): 323–344.
- Rivet, Paul. 1925. Les mélano-polynésiens et les australiens en Amérique. *Anthropos* 20: 51–54.
- Rivet, Paul. 1926. Les malayo-polynésiens en Amérique. *Journal de la Société des Américanistes de Paris* 18: 141–278.
- Sagart, Laurent. 1990. *Chinese and Austronesian are genetically related*. Paper presented at the 23rd International Conference on Sino-Tibetan Languages and Linguistics, Arlington, October 1990.
- Sagart, Laurent. 1993. Chinese and Austronesian: Evidence for a genetic relationship. *Journal of Chinese Linguistics* 21: 1–62.
- Sagart, Laurent. 1994. Old Chinese and Proto-Austronesian. *Oceanic Linguistics* 33(2): 271–308.
- Sagart, Laurent. 1995a. Some remarks on the ancestry of Chinese. In: William S-Y. Wang (ed.). *The Ancestry of the Chinese Language. Journal of Chinese Linguistics Monograph Series 8*: 195–223. University Of California Press.
- Sagart, Laurent. 1995b. Comments from Sagart. In: William S-Y. Wang (ed.). *The Ancestry of the Chinese Language. Journal of Chinese Linguistics Monograph Series 8*: 337–372. University Of California Press.
- Sagart, Laurent. 1995c. Questions of method in Chinese-Tibeto-Burman comparison. *Cahiers de linguistique Asie orientale* 24(2): 245–255.
- Sagart, Laurent. 1999. *The roots of old Chinese*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Sagart, Laurent. 2001. Vestiges of Archaic Chinese derivational affixes in modern Chinese dialects. In: Hilary Chappell (ed.). *Sinitic grammar: synchronic and diachronic perspectives*: 123–142. Oxford University Press.
- Sagart, Laurent. 2003. The vocabulary of cereal cultivation and the phylogeny of East Asian languages. *Bulletin of the Indo-Pacific Prehistory Association* 23(1): 127–136.
- Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2): 411–44.
- Sagart, Laurent. 2005. Sino-Tibetan-Austronesian: An updated and improved argument. In: Laurent Sagart, Roger Blench, Alicia Sanchez-Mazas (eds.). *The peopling of East Asia: Putting together archaeology, linguistics and genetics*: 161–176. London / New York: Routledge Curzon.
- Sagart, Laurent. 2006. On intransitive nasal prefixation in Sino-Tibetan languages. *Cahiers de linguistique Asie orientale* 35(1): 57–70.

- Sagart, Laurent. 2008a. The expansion of *Setaria* farmers in East Asia. In: Alicia Sanchez-Mazas, Roger Blench, Malcolm D. Ross, Ilia Peiros and Marie Lin (eds.) *Past human migrations in East Asia: matching archaeology, linguistics and genetics*: 133–157. London: Routledge.
- Sagart, Laurent. 2008b. Reply to Matisoff on the Handbook of Proto-Tibeto-Burman: System and philosophy of Sino-Tibetan reconstruction. *Diachronica* 25(1): 153–155.
- Sagart, Laurent. 2009. *PAn morphology in phylogenetic perspective*. Paper presented at 11-ICAL (11ème conférence internationale sur la linguistique austronésienne), Aussois, June 22–25, 2009.
- Sagart, Laurent. 2010a. Is Puyuma a primary branch of Austronesian? *Oceanic Linguistics* 49(1): 194–204.
- Sagart, Laurent. 2010b. *Why does the Tai-Kadai word for 'ten' end in -t?* Paper presented at the 23èmes Journées de Linguistique d'Asie Orientale, Paris, July 1–2, 2010.
- Sagart, Laurent. 2011a. The homeland of Sino-Tibetan-Austronesian: where and when? *Communication on Contemporary Anthropology* 2011(5): 143–147.
- Sagart, Laurent. 2011b. How many independent rice vocabularies in Asia? *Rice* 4(3–4): 121–133.
- Sagart, Laurent. 2011c. *Sino-Tibetan-Austronesian: Making sense of non-final syllables*. Paper presented at a symposium on Historical Linguistics in the Asia-Pacific Region and the Position of Japanese, Twentieth International Conference on Historical Linguistics Symposium, National Museum of Ethnology, Osaka, July 30, 2011.
- Sagart, Laurent. 2013. The higher phylogeny of Austronesian: A response to Winter. *Oceanic Linguistics* 52(1): 249–254.
- Sagart, Laurent. 2014. In defense of the numeral-based model of Austronesian phylogeny, and of Tsouic. *Language and Linguistics* 15(6): 859–882.
- Sagart, Laurent. 2016a. The wider connections of Austronesian: a Response to Blust (2009). *Diachronica* 33(2): 255–281.
- Sagart, Laurent. 2016b. Liangdao Man and Austronesian origins. *Journal of Chinese Linguistics* 44(1): 239–241.
- Sapir, Edward. 1915. The Na-Dene Languages, a preliminary report. *American Anthropologist* 17(3): 534–558.
- Sapir, Edward. 1925. The similarity of Chinese and Indian languages. *Science* 62(1607):xii.
- Schmidt, Wilhelm. 1906. *Die Mon-Khmer-völker: ein bindeglied zwischen Völkern Zentralasiens und Austronesiens*. Braunschweig: F. Vieweg und Sohn.
- Schwartz, J. 1946. The teeth of the Masai. *Journal of dental research* 25(1): 17–20.
- Shafer, Robert. 1952. Athapaskan and Sino-Tibetan. *International Journal of American Linguistics* 18(1): 12–19.
- Shafer, Robert. 1957. Note on Athapaskan and Sino-Tibetan. *International Journal of American Linguistics* 23(2): 116–117.
- Shafer, Robert. 1969. A few more Athapaskan and Sino-Tibetan comparisons. *International Journal of American Linguistics* 35(1): 67–67.
- Schuessler, Axel. 2009. *Minimal Old Chinese and later Han Chinese: a companion to Grammata serica recensita*. University of Hawaii Press.
- South Coblin, Weldon. 1986. *A Sinologist's handlist of Sino-Tibetan lexical comparisons*. Nettetal: Steyler Verlag.
- Starostin, S. A. 1984. Gipoteza o geneticheskix svjazjax sinotibetskix jazykov s enisejskimi i severnokavkazskimi jazykami. In: I. Vardul' (ed.). *Lingvističeskaja rekonstrukcija i drevnejšaja istorija Vostoka*: 19–38. Moscow: Nauka.
- Starostin, S. A. 1989a. *Rekonstrukcija drevnekitajskoj fonologičeskoj sistemy*. Moscow: Nauka.
- Starostin, S. A. 1989b. Nostratic and Sino-Caucasian. In: V. Shevoroshkin (ed.). *Explorations in language macrofamilies*: 42–66. Bochum: Universitätsverlag N. Brockmeyer.
- Starostin, S. A. 1991. On the hypothesis of a genetic connection between the Sino-Tibetan languages and the Yeniseian and North-Caucasian languages. In: V. Shevoroshkin (ed.). *Dene-Sino-Caucasian Languages*: 12–41. Bochum: Universitätsverlag N. Brockmeyer.
- Starostin, S. A. 2005. *Sino-Caucasian*. Unfinished MS. Available online at: <http://starling.rinet.ru> (Tower of Babel Project).
- Stevens, C. J., C. Murphy, R. Roberts, L. Lucas, F. Silva, D. Q. Fuller. 2016. Between China and South Asia: A Middle Asian corridor of crop dispersal and agricultural innovation in the Bronze Age. *The Holocene* 26 (10): 1541–1555.
- Swadesh, Morris. 1952a. "Athapaskan and Sino-Tibetan." Review of Shafer 1952. *International Journal of American Linguistics*: 178–181.
- Swadesh, Morris. 1952b. Lexicostatistic Dating of Prehistoric Ethnic Contacts. *Proceedings of the American Philosophical Society* 96: 452–463.
- Swadesh, Morris. 1955. Towards Greater Accuracy in Lexicostatistic Dating. *International Journal of American Linguistics* 21: 121–137.
- Swadesh, Morris. 1971. *The Origin and Diversification of Language*. Ed. post mortem by Joel Sherzer. Chicago: Aldine.

- Tabbada, K., J. Trejaut, J. Loo, Y. Chen, M. Lin, M. Mirazon-Lahr, T. Kivisild, M. De Ungria. 2009. Philippine Mitochondrial DNA Diversity: A Populated Viaduct between Taiwan and Indonesia? *Molecular Biology and Evolution* 27(1): 21–31.
- Tan, J. M. 2001. *Hànyǔ cíhuì de jìn yuán tàn mì* 漢語詞彙的近源探秘. Běijīng: Yuanzineng Publishing House.
- Ulenbrook, Jan. 1967. Einige Übereinstimmungen zwischen dem Chinesischen und dem Indogermanischen. *Anthropos* 3/4: 533–551.
- Vovin, Alexander. 2015. Eskimo Loanwords in Northern Tungusic. *Iran and the Caucasus* 19: 87–95.
- Wang, William S.-Y. (ed.). 1995. *The ancestry of the Chinese language. Journal of Chinese Linguistics Monograph Series 8.* University Of California Press.
- Wáng, X. Q., X. Liú. 1896. *Shì míng shū zhèng bǔ* 釋名疏證補. Shànghǎi gǔ jī Publishing House.
- Wei, J. L. 2005a. Review of Hànyǔ Yìn'ōuyǔ cíhuì bǐjiào. *Sino-Platonic Papers* 166: 1–22.
- Wei, J. L. 2005b. Counting and Knotting: Correspondences between Old Chinese and Indo-European. *Sino-Platonic Papers* 162: 1–72.
- Wolff, J. U. 2010. *Proto-Austronesian phonology with glossary.* Southeast Asia Program Publications. 2 vols. Ithaca, New York: Cornell University.
- Wu, R. K. 1989. *Gǔ rén lèi xué* 古人類學. Běijīng: Wénwù chūbǎn shè.
- Wu, Q., Z. Zhao, L. Liu, D. E. Granger, H. Wang, D. J. Cohen, X. Wu, M. Ye, O. Bar-Yosef, B. Lu, J. Zhang. 2016. Outburst flood at 1920 BCE supports historicity of China's Great Flood and the Xia dynasty. *Science* 353 (6299): 579–582.
- Zhōu Jíxù. 2002a. *Hànyǔ Yìn'ōuyǔ cíhuì bǐjiào* 漢語印歐語詞彙比較 [Word comparisons between Chinese and Indo-European]. Sìchuān Míngzú Publishing House.
- Zhōu Jíxù. 2002b. Correspondences of the Basic Words between Old Chinese and Proto-Indo-European. *Sino-Platonic Papers* 115: 1–8.
- Zhōu Jíxù. 2003. Correspondences of Cultural Words between Old Chinese and Proto-Indo-European. *Sino-Platonic Papers* 124: 1–19.
- Zohary, D., M. Hopf, E. Weiss. 2012. *Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin.* Oxford University Press on Demand.

Дж. Орланди. К оценке сино-тибето-австронезийской гипотезы.

В статье проводится анализ и оценка гипотезы о возможном генетическом родстве между сино-тибетскими и австронезийскими языками, предложенной французским синологом Лораном Сагаром. Статья преследует не столько цель доказать или опровергнуть существование сино-тибето-австронезийской макросемьи, сколько проверить, соответствуют ли данные, собранные Сагаром в ее поддержку, стандартным критериям исторического языкознания. Помимо этого, обсуждается ряд вопросов, связанных с гипотетической прародиной сино-тибето-австронезийцев, в том числе в контексте таких конкурирующих гипотез, как сино-индоевропейская. Мы приходим к выводу, что подход, избранный Сагаром, недостаточно доказателен для такого рода сложных и противоречивых гипотез о дальнем языковом родстве.

Ключевые слова: сино-тибето-австронезийская гипотеза; сино-тибетские языки; генетическое родство языков; сравнительно-исторический метод.

Lexicostatistical Studies in Khoisan I: The Ju-#Hoan Relationship¹

This paper is the first one in an intended series of publications on lexicostatistical relations between several linguistic groupings that have all been assigned by Joseph Greenberg to the hypothetical Khoisan macrofamily. Here, we examine the numbers and natures of various matches between the basic lexicon of two such groupings: the closely related cluster of Ju (North Khoisan) dialects and the Eastern #Hoan language, formerly considered an isolate but now widely regarded as the closest, and only non-controversial, genetic relative of Ju. Based on both superficial and etymological analysis of the data (including the reconstruction of a Swadesh wordlist for Proto-Ju), we conclude that there are from 32% to 44% matches between Proto-Ju and #Hoan (depending on the degree of strictness required from phonetic correspondences), which is translatable to a time depth comparable with such families as Fenno-Ugric and Kartvelian. Additionally, the distribution of cognates between the various stability layers of the basic lexicon is analyzed, leading to the conclusion that the matches are indeed indicative of genetic relationship rather than areal contact.

Keywords: Khoisan languages, Ju languages, Eastern #Hoan language, lexicostatistics, glottochronology, comparative-historical method.

Introduction

Despite significant progress that has been achieved over the past few decades in our understanding of the linguistic nature and historical relationships of the various «Khoisan» languages², there is still very little consensus on deep level genetic connections between low-level linguistic groupings, traditionally viewed as «Khoisan». Although practically all the researchers now working in the field seem to agree that Joseph Greenberg's «Khoisan», including all the non-Bantu and non-Cushitic click languages of South and Central Africa, has not been convincingly shown to constitute a valid genetic entity, judgements differ significantly on what might be the deepest identifiable genetic links between the three commonly accepted «Khoisan» families (Northern, or Ju; Southern, or Tuu; Central, or Khoe) and the four known «Khoisan» language isolates (#Hoan, Kwadi, Sandawe, and Hadza) — particularly because the criteria for testing the plausibility of such links often depend on the personal intuitions and preferences of researchers³.

¹ This research has been carried out as part of the Evolution of Human Languages project, supported by the Santa Fe Institute. Special gratitude goes to Dr. Bonny Sands, who provided the author with numerous insightful comments, corrections, and additional references at the early draft stage.

² For a relatively complete summary of all these developments, see Rainer Vossen's and Henry Honken's succinct overviews in Vossen (ed.) 2013: 1–24.

³ See Greenberg 1966 for a comprehensive set of arguments in favor of Khoisan as a genetic unity; Güldemann 2014 for an overview of the current state of Khoisan classification from a grammatical and typological angle. The principal conclusions («two hypotheses seem to be promising to pursue in the future. In the order of probability these would be to join Sandawe with Khoe-Kwadi, and Kx'a (= Ju-#Hoan /G.S./) with Tuu», p. 35) are conveniently consistent with the lexicostatistical conclusions in Starostin 2013, except that I would reverse the order of the two probabilities.

A general unified framework for evaluating the various hypotheses on such links has been suggested in Starostin 2003, 2008, and further refined in Starostin 2013. It combines a formal lexicostatistical approach, serving as a common evaluation standard for all such hypotheses (applicable not only to «Khoisan» lineages but, in theory, to any of the world's language families), with elements of the traditional comparative method and Greenberg's multi-lateral comparison, and allows for a rough probabilistic ranking of competing hypotheses.

However, the framework has not really been fully applied to all relevant data. In particular, inter-group comparison in Starostin 2013 has only been conducted (a) based on the abridged 50-item, rather than the complete 100-item variant of the Swadesh wordlist; (b) based on an automated algorithm of comparison (utilising Aron Dolgopolsky's «consonant class method» of evaluating phonetic similarity⁴), rather than the more fine-grained and historically significant method of establishing recurrent patterns of phonetic correspondences. Both of these decisions were intentional and technically inevitable within the scope of a general preliminary survey, resulting in a first-approximation classification scheme that should then be subject to additional revisions and refinements.

The present paper is the first attempt at such a refinement, and is intended to provide additional insights into one of the most reliable and closest linkages confirmed by the overall survey, namely, the genetic connection between the Ju, or North Khoisan, cluster of closely related languages (or, perhaps, dialects of a single macro-language), and (Eastern) †Hoan (= †Hòã or †Huã), an isolated language of Botswana. Such a connection could already be suspected from the lexical comparisons presented by Anthony Traill in his pioneering study of †Hoan (Traill 1973), and seemed plausible even to such a notorious «splitter» in the field of Khoisanology as E. O. J. Westphal (1974). Since then, the main proponents of a specific genetic (rather than areal) connection between Ju and †Hoan have been Henry Honken (1977, 1988)⁵ and George Starostin (2003, 2008)⁶. The most recent attempt for a comprehensive survey of the evidence relating the two small taxa is Heine & Honken 2010, where the authors provide their own reconstruction of the phonological system for Proto-Ju-†Hoan⁷, illustrating it with numerous lexical examples. All in all, the total amount of lexical and grammatical isoglosses between Ju and †Hoan, coupled with numerous attested phonetic regularities between the observed etymological parallels, makes the Ju-†Hoan relationship proposal one of the most reliable and highly probable historical hypotheses about Khoisan languages in general.

Nevertheless, in order to complete the formal testing of the hypothesis and to provide a stronger foundation for the *genetic*, rather than *areal*, interpretation of the evidence, we find it useful to present a detailed lexicostatistical evaluation of Ju-†Hoan, in accordance with the basic methodological guidelines laid out in Starostin 2013. A first attempt at such an evaluation

⁴ For an up-to-date description of the Dolgopolsky method, see Kassian et al. 2015: 307.

⁵ In his first publication on historical Khoisan linguistics (1977), Honken classifies †Hoan and Ju|'hoan as «Ž1» and «Ž2» respectively, implying their close relationship without specifically commenting on it. In Honken 1988: 59, he explicitly states: «I have put Eastern †huã firmly in the Zhu family unlike Traill who regards it as a link between Zhu and Taa», providing several examples of lexical and phonetic isoglosses to strengthen his case.

⁶ In both of these sources, the primary argument for a close relationship between Ju and †Hoan is made on the basis of lexicostatistical analysis. However, Starostin 2008: 356–363 also presents a first approximation for a reconstruction of the «Proto-North-†Hoan» phonological system.

⁷ Heine and Honken have suggested the short name «Kx'a» to denote this taxon, based on the identical phonetic shape for the word 'earth' in both Ju and †Hoan. This seems to be more of a mnemonic tactic than a substantial decision (why should Ju-†Hoan be thought of as the 'earth family?'), and it also bypasses the fact that the same word for 'earth' (*kx'a*) is also used in the unrelated extinct Kwadi language of Angola, which would complicate the matter even further. We prefer to stick to 'Ju-†Hoan' as a slightly more complex, but more accurate designation for this taxon.

was already carried out in Starostin 2003, where Proto-Ju (= Proto-North Khoisan) and ǀHoan were found to have 43% matches on the Swadesh list. However, that comparison was merely a part of a much larger general study, not allowing the author to focus on specific issues of finding and evaluating lexicostatistical and etymological matches between the two taxa; moreover, it did not properly take into account the possible effects of areal diffusion, and employed somewhat lax and properly undefined criteria for establishing phonetic correspondences. Another important limitation is that it relied too heavily on limited and not wholly accurate lexical data for ǀHoan, not being able to take into consideration a lot of data that have only been published over the past ten years (see our main sources below).

A significant improvement has been offered in Starostin 2013, which already made use of much better data for both ǀHoan and the different varieties of Ju. However, that study was also a general lexicostatistical evaluation of phonetic similarities (rather than regularities) between the different Khoisan lineages; and while the study itself, limited to the «ultra-stable» 50-item half of the Swadesh wordlist, *confirmed* the existence of a special link between Ju and ǀHoan, it did not truly *explore* that link the way a thorough joint lexicostatistical-etymological study should have done. Consequently, this paper is an attempt to remedy that situation and provide a *definitive* lexicostatistical evaluation of the evidence for Ju-ǀHoan, one that would allow us to formulate explicit historical statements about the relative chronology of these families, some particularities of their divergence, and their areal connections with other varieties of «Khoisan».

Data

Complete and most up-to-date versions of the 110-item wordlists⁸ for six different languages/dialects of the Ju group and for (Eastern) ǀHoan, accompanied by detailed annotations, are currently available at the *Global Lexicostatistical Database* (<http://starling.rinet.ru/new100>). The Ju lists differ significantly in quality, since only two of them are drawn from relatively recent sources that benefit from greater phonetic and semantic accuracy⁹: Ju'hoan, based on Patrick Dickens' dictionary (Dickens 1994), and Northwestern (Ekoka) !Xun (!Xung), based on the glossary in König & Heine 2008 (and largely coinciding, pending certain phonetic discrepancies, with the data in Heikkinen 1986).

The availability of both these sources today is a strong advantage, since Ju'hoan and Ekoka represent two different sub-clusters of Ju dialects and are about as far removed from each other lexically as any two languages/dialects of Ju can be. However, for the sakes of etymological and lexicostatistical accuracy, and as a necessary condition for a reliable reconstruction of the basic lexicon for Proto-Ju, it is imperative to also make use of older data, namely, the vocabularies collected by Lucy Lloyd, Dorothea Bleek, and Clement Doke, all of them eventually integrated in D. Bleek's monumental comparative dictionary (Bleek 1956). We have specifically selected four varieties:

⁸ The 110-item wordlist is a slightly expanded version of the standard 100-item Swadesh wordlist (with 10 additional items from the earlier 200-item version) commonly used in lexicostatistical studies conducted by the Moscow school of comparative linguistics. For specific details on the semantics of individual items and on the handling of potential synonyms, see Kassian et al. 2010.

⁹ This should not be understood as implying that these works are completely free of phonetic errors: in most cases, it makes sense to compare transcriptions by different specialists where they are available. Nevertheless, qualitative differences between most of the modern sources look relatively negligible when compared with the first systematic attempts at transcribing Ju (and other Khoisan) phonologies in the early 20th century.

(a) Lucy Lloyd's «!Kung», recorded in 1879–1880 from four young informants from around lake Ngami; typically correlated with what has been termed the «Central Dialect Cluster» in Snyman 1997, Treis 1998, and Sands 2010, but more recently re-aligned with the «North-Central» cluster by Florian Lionnet (2009) because of specific lexical, phonetic, and grammatical isoglosses;

(b) Clement Doke's «!Hũ» of Grootfontein (research originally published as Doke 1925), also typically grouped in the Central cluster (despite some significant discrepancies with Lloyd's data, although it is often hard to understand if these discrepancies are real or due to inaccurate fixation);

(c) Dorothea Bleek's «||K'au||en» or «‡Au kwe» (the most modern transcription in Vossen 2013: 9 puts the dialect's name as †x'áó-||'àèh), recorded in the early 1920s at Sandfontein; this dialect is typically assumed to belong to the Southern cluster as well (Treis 1998: 468), although the issue remains open due to lack of modern data from the same region (Sands 2010)¹⁰;

(d) Dorothea Bleek's «!O!kung», recorded in Central Angola in 1925; this dialect is lexically and phonetically very close to Ekoka !Xun, as well as to «Angolan !Xũ», a brief account of which was published as Snyman 1980.

All four of these sources share the same advantages (sufficient in size to allow for a relatively complete and representative set of Swadesh-type wordlists; recorded a hundred or so years ago in communities slightly less linguistically susceptible to Khoe, Bantu, and European influence than they are today) and flaws (generally poor quality of transcription and possible semantic inaccuracies). In the case of this particular study, however, phonetic inaccuracies are not a significant problem as long as the necessary adjustments are made (i.e. there is a general understanding of what kinds of errors are typical for Lloyd's and Bleek's data); semantic inaccuracies are far more harmful for lexicostatistical data and can severely influence classifications and datings, but as long as the data may be compared with data from more recent and accurate sources, most of the potential errors may be successfully filtered out on the way from modern data to the reconstructed proto-wordlist¹¹.

No «official» dictionary has so far been published for Eastern †Hoan, but enough lexical data have become available in the past few decades to make the language perfectly acceptable for lexicostatistical comparison. Most of that data have been collected by Jeffrey Gruber (G) and Chris Collins (C), our main source being the relatively recently published comparative grammar of the language (Collins, Gruber 2014), well illustrated by lexical and textual examples, and also heavily drawing upon previously published papers by the same authors (Collins 2001, 2002; Bell, Collins 2001; Gruber 1975). A few lacunae had to be filled in by data from the first ever published wordlist of †Hoan that was put together by Anthony Traill (1973); overall comparison of Traill's data with Gruber's and Collins' materials shows that, while the quality of Traill's transcription leaves a lot to be desired, his elicitation of †Hoan lexical equivalents for basic semantic notions was largely correct.

Naturally, some data sources for other Khoisan languages have to be taken into consideration as well, since any serious study on the etymology or lexicostatistics of Ju-†Hoan has to take the areal factor into account. In particular, †Hoan is known to have been in tight contact

¹⁰ Work on the documentation of †x'áó-||'àèh is currently being conducted by Lee Pratchett (2017), but outside of several papers dealing with specific phonetic and grammatical issues, no comprehensive data collections have yet been made publicly available.

¹¹ For the basic principles of reconstructing an «optimal» Swadesh-type proto-wordlist from attested lexical data (in accordance with which we reconstruct the Proto-Ju wordlist in this paper), see Starostin 2016.

with |Gui, a Kalahari Khoe language, and through it (and, perhaps, directly as well), also with !Xóǀ, a Taa language (Traill & Nakagawa 2000); although some of the resemblances between ǀHoan and !Xóǀ are not to be ruled out as potential evidence for genetic relationship on a deeper level than Ju-ǀHoan (Starostin's «Peripheral Khoisan»), specific binary isoglosses between the two languages without any parallels in the rest of «Peripheral Khoisan» are most likely explainable as results of diffusion. Most of the references to !Xóǀ lexicon will be given according to Traill 1994; Kalahari Khoe references will be provided according to the reconstructions in Vossen 1997, except where specially noted.

For the sakes of general convenience, we utilize here a unified system of transcription as is currently adopted for the purposes of the *Global Lexicostatistical Database* project; for the most part, it does not differ from IPA, except for a few details (such as the use of single-graph vs. digraph transcriptions for affricates: IPA $ts = c$, IPA $tʃ = č$, IPA $tɕ = ɕ$, etc.). In our transcription of click accompaniments, we also follow the old transcriptional convention by Rainer Vossen (1997), where voiced clicks are transcribed as $ǀ$, $ǂ$ etc. (instead of $gǀ$, $gǂ$ or $|g$, $|g$, etc.) and nasalized are transcribed as $ǀ̃$, $ǂ̃$ etc. (instead of $nǀ$, $nǂ$ or $|n$, $|n$ etc.).

Comparative procedure

For the sake of historical accuracy, lexicostatistical comparison between Ju languages and ǀHoan has to be carried out on the level of protolanguage reconstruction in the case of Ju¹². Although some details of Proto-Ju and the phonetic laws that tie it to its modern descendants still remain poorly understood (mostly in the sphere of tonology and non-productive/fossilized nominal morphology), all the dialects are close enough to provide evidence for the basic phonetic shape of the protoforms, particularly with the aid of precious comparative data in J. Snyman's (1997) dialectal survey. It is very important not to rely exclusively on a single source, such as Patrick Dickens' exhaustive dictionary of Juǀ'hoan, which, paradoxically, sometimes provides *too much* data for an accurate lexicostatistical analysis (for instance, many basic terms, such as body parts, are often represented in that dictionary by doublet forms — one inherited from Proto-Ju, one recently borrowed from Khoe; external comparison with other Ju dialects helps sort the situation out very easily).

Although a definitive areal/historical classification for Ju dialects is still lacking, it seems clear from both phonetic and lexical evidence that the sharpest dividing line separates the Southern cluster, represented most prominently by Juǀ'hoan, from the Northern cluster, represented by Ekoka !Xun. The lexicostatistical implications are such that, quite often, one finds a binary opposition between Juǀ'hoan (and related dialects) and Ekoka (and related dialects), where simple distributional considerations are not enough to understand which of the two roots is a better candidate for the respective «Swadesh meaning» on the Proto-Ju wordlist. In such cases, we resort to «extra-distributional rules»¹³ to help resolve the situation, wherever they are applicable. When no reasonable choice can be made, we may count two roots as «technical synonyms» and subject both of them to comparison with ǀHoan.

¹² Theoretically, it is also possible to subject ǀHoan data to the reconstruction procedure, since we know of at least one additional dialectal variety, Sasi, somewhat divergent from ǀHoan proper; however, data on Sasi are extremely limited and, at best, show it to be slightly more archaic in terms of certain phonetic features, but not in terms of lexical stability. For more details on the differences between the two dialects, see Collins & Gruber 2014: 17–20.

¹³ For a complete list of said rules, illustrated by examples, see Starostin 2016. These typically have to do with internal etymologization or external analysis (checking for borrowings, etc.).

Matches between Proto-Ju reconstructions and †Hoan forms are evaluated on a somewhat fine-grained scale, allowing for a more insightful final analysis. The «evaluation marks» are as follows:

+ : *Definitive lexicostatistical matches*. To get a + mark, both parts of the comparison have to be reliably attested or reconstructed in the appropriate Swadesh meaning, *and* be phonetically compatible, i.e. agree with the basic correspondence patterns, identified in Starostin 2008 and in Heine & Honken 2010 (see below on the comparison between the two systems). «Phonetic compatibility» does not necessarily imply complete historical transparency of the correspondences between each of the segmental and suprasegmental features, but it does imply that the majority of segmental alignments should display pattern-like behavior¹⁴.

± : *Potential lexicostatistical matches*. These pairings, also reliably attested or reconstructed in the appropriate Swadesh meaning, typically display a remarkable degree of phonetic similarity, but also feature at least one (preferably not more than one) major segmental discrepancy that cannot be explained according to our current understanding of the historical phonology of Ju-†Hoan. Such matches cannot be taken as direct evidence for relationship and should not be included into the main round of lexicostatistical calculations, but since we cannot claim to know everything there is to know about regular vs. sporadic developments from Proto-Ju-†Hoan to their modern descendants, it makes perfect sense to make note of such potential matches and include them in an alternate set of lexicostatistical calculations (see below).

≈ : *Etymostatistical (etymological) matches*. Since this study is carried out on the data of a compact, binary taxon, tied together by sets of phonetic correspondences, it makes sense to expand the strict lexicostatistical analysis (demanding exact semantic matches between compared items) by also taking into consideration those situations where a Proto-Ju Swadesh item finds a good phonetic/semantic match in †Hoan (or vice versa), but the meanings are semantically related rather than semantically identical. Based on typological (and simply logical) arguments, in any situation of language relationship we should be able to find such matches in addition to direct lexicostatistical ones, and comparing their numbers and their character to those of direct lexicostatistical matches should provide additional insight into the degree and nature of their relationship.

- : *No matches*. There are no hitherto detected parallels between the compared items. (Given the deficiency of our knowledge on Ju and especially †Hoan lexicon, any of these pairings could turn out to be etymostatistical matches in the future, but it is highly unlikely that they will ever turn out to be direct lexicostatistical matches).

? : *Insufficient data*. These are the cases where the respective item is not attested in our sources on †Hoan (e.g. ‘bark’), or is insufficiently well attested in Ju idioms to be reconstructible (e.g. ‘round’). In all such cases, the Swadesh item is excluded from calculations, and any percentages are calculated out of the remaining items. The same also applies to a few cases where either the †Hoan item (e.g. ‘salt’) or the Proto-Ju (or, rather, «Common Ju») item (e.g. ‘fish’) is highly likely to have been borrowed from a third source, such as !Xóõ or Khoe.

An additional factor to be taken into consideration is the distribution of detected cognates across the wordlist. In accordance with the well-known and empirically well confirmed «Yakhontov principle» (genetically related languages will share more matches on the more stable sub-

¹⁴ For instance, the exact factors determining the lack or presence of voicing during click articulation in Ju-†Hoan cognates remain obscure; however, recurrent examples are available for all four types of possible correlations, confirming their regularity. In other words, it is impossible at present to offer unequivocal Proto-Ju-†Hoan reconstructions for such items (due to insufficient data or incomplete analysis of all the factors that could be involved), but it is possible to regard them as reliable cognates.

section of the lexicostatistical wordlist, while languages in contact will share more matches on the less stable sub-section), we separate the 100-item list in two halves and compare the numbers for all types of matches (definitive, potential, etymostatistical) separately, so that the nature of relationship between Ju and #Hoan could be assessed according to that parameter — and so that the results could also contribute to establishing a general benchmark for all such types of situations.

Correspondences

Unlike deeper level lexicostatistical comparisons, where comparanda still have to be evaluated on the basis of phonetic similarity rather than phonetic correspondences, Proto-Ju and #Hoan forms have the benefit of actually being linked together by recurrent phonetic isomorphisms, as shown in Starostin 2008 and Heine & Honken 2010. Due to data limitations and certain unresolved issues with Proto-Ju itself, these isomorphisms have not yet been processed to the stage of a definitive, all-encompassing phonological reconstruction of Proto-Ju-#Hoan, but enough of them have been observed for us to be able to confidently propose common Ju-#Hoan etymologies even in certain cases where the forms do not at all look alike.

In the notes section for each individual comparison, we typically comment on the degree of regularity that may be inferred for specific Ju-#Hoan segments, particularly when these segments are not phonetically identical. Where necessary and/or possible, additional examples to confirm the recurrent nature of the pattern are drawn upon from the available corpus of Ju-#Hoan etymological comparanda (most of it published either in Starostin 2008 or in Heine & Honken 2010). The complete list of correspondences observed between Ju and #Hoan basic lexicon items is given in the Appendix, with each correspondence enumerated so that it can be briefly referred to in the main section of the paper.

A detailed description of the phonological systems of (Proto-)Ju and #Hoan lies well beyond the scope of the current paper. See Miller 2013 for an up-to-date brief account of Ju phonology and phonetics, Honken 2013 for the same concerning Eastern #Hoan, and the above-mentioned papers by Starostin and Heine/Honken for comparisons between the two.

Abbreviations

Language names: PJ = Proto-Ju; Ek. = Ekoka !Xun; Ju. = Ju'hoan; Kg. = (Lucy Lloyd's) !Kung; Kx. = #X'áo-||'àè; OK. = (Dorothea Bleek's) !O!Kung.

Sources: C = Ch. Collins (for #Hoan); G = J. Gruber (for #Hoan); S = B. Sands (for #Hoan); HH = Heine, Honken 2010; SH = Sands, Honken 2014.

Ju/#Hoan comparative wordlist

1. ALL (+)

- PJ: ***wòe-še** (Ju. *wè-šè*, Kx. *oá-si*, Gr. *we:še-sn̄*, OK. *wì-sè ~ wè-šè*, Ek. *wòhē-šē*). ◊ Preserved in all daughter dialects. No alternate stems. The reconstruction follows the Ekoka variant as phonetically more archaic in its vocalism; the variant **wè-še* is also possible. Extra low tone in Ekoka is not, however, confirmed by the rest of the data. The form is morphologically complex: the derivation is transparently seen in Ekoka, cf. *wòhà* 'for-

ever', *wðhè* 'some time ago, already, just'. The meaning of the suffix **-še*, however, remains unknown.

- †Hoan: *ùē* (G).
- Ju-†Hoan: A phonetically similar and compatible match¹⁵. Since initial **w-* in PJ is not prothetic, we have to suppose simplification in †Hoan (**woe* → *ue*). The suffix **-še* may have been a PJ innovation. ◊ HH: 14.

2. ASHES (-)

- PJ: **tō̃* (Ju. *tō̃*, Kx. *tx*, Kg. *tx*: ~ *tó*). ◊ Not attested in the Northern dialects, except for the reduplicated variant *tàõ-tàõ* in Snyman 1980: 33. OK. *ɬàā* 'ashes' = Ju. *ɬàā^h* 'soap', both forms probably having been borrowed from Khoe sources, cf. Proto-Khoe **ɬoa* 'ashes' (Vossen 1997: 417); Ek. *ɬðhà* 'ashes' probably belongs here as well, but the click correspondence is irregular (possibly a transcription error).
- †Hoan: *ɬoe* (T). ◊ Only attested in Traill's records (as *ɬoe* ~ *ɬue* ~ *ɬue^h*), so the precise phonological shape is uncertain; however, the word is clearly not a possible match for PJ.
- Ju-†Hoan: No lexicostatistical or etymological matches.

3. BARK (?)

- PJ: **ɬõʔrV_A* (Ju. *ɬõʔòrò*, Ek. *ɬùli*). ◊ Reliably reconstructible for the PJ stage, although in many dialects, particularly those found in Bleek's dictionary, the meaning 'bark' is usually merged with 'skin' (Kx. *ɬo-si*, Kg. *ɬò* ~ *ɬò* ~ *ɬwa*; see 'skin' below).
- †Hoan: Not attested.

4. BELLY (+)

- PJ: **ɬú* (Ju. *ɬú*, Kx. *ɬú*, Kg. *ɬù* ~ *ɬú*, Gr. *ɬú*, OK. *ɬú*, Ek. *ɬú*). ◊ Preserved in all daughter dialects. No alternate stems. Straightforward reconstruction.
- †Hoan: *!ō* (C, G); *!úv* ~ *!óō* ~ *!^hú*: (S).
- Ju-†Hoan: A phonetically similar and compatible match (see corr. #12, #38a). ◊ HH: 17¹⁶.

5. BIG (≈)

- PJ: **ɬ!àʔà* (Ju. *ɬ!àʔà*, Kx. *ɬ!a:/*, Kg. *ɬ!a*, Gr. *ɬ!a*, OK. *ɬ!a* ~ *ɬ!á* ~ *ɬ!aa*, Ek. *ɬ!á* ~ *ɬ!àʔà*). ◊ Preserved in all daughter dialects. Correspondences are regular, indicating an original retroflex nasalized click and a glottal stop between the vowels.
- †Hoan: (a) *ɬ^hĩ*, (b) *ɬ!ào* (C, G). ◊ Both of these words are consistently glossed as 'big' in available sources, but textual examples offer no hint at their semantic differences.
- Ju-†Hoan: No direct matches. However, (a) is a transparent etymological match with PJ **ɬ^hi* 'much, many' (Ju. *ɬ^hái*, OK. *ɬ^hi*, etc.; see MANY); the semantic shift 'big' ↔ 'many' is quite trivial, although the direction of the shift remains unclear in this case.

¹⁵ The vocalic correspondence is unique (if we are talking about the coda as a whole), but there are not a lot of cases of Proto-Ju **-oe* with reliable parallels in †Hoan. At least one attested case also involves †Hoan *-ui* (PJ **šoe* 'to take out / take off' = †Hoan *šui* 'to drop off').

¹⁶ In Sands, Honken 2014: 252, the connection is put in doubt because of the incompatibility of Ju *ɬ-* and †Hoan *!^h-*, but it is not certain that the aspirated accompaniment is phonologically primary in this case (most of the other sources agree on zero accompaniment, and even Sands herself records phonetic variation between *!-* and *!^h-*).

6. BIRD (-)

- PJ: ***c̣ā(m)-mà** (Ju. *c̣àmà*, Kx. *cama*, Kg. *caba*, Gr. *c̣auà*, OK. *cama*, Ek. *čámà*). ◇ Preserved in all daughter dialects. The word is morphologically complex; the second component is clearly identifiable as PJ *-*ma* ‘small; diminutive suffix’, cf. the corresponding plural form *čá-m̃^hè* in Ek. Given the additional presence of Ek. *čám* ‘poultry, bird, aeroplane’, it is possible, but not certain, that the original root shape was **c̣ām* rather than **c̣ā* (with subsequent contraction *-*mm-* → -*m-* in most dialects).
- ǀHoan: **ǀǎ-sì**: (G). ◇ The suffix -*si* is a diminutive morpheme.
- Ju-ǀHoan: H. Honken (1988: 60) quotes the ǀHoan form *c̣hà:^fma* ‘bird’, apparently taken from Gruber’s formally unpublished field records; if it really exists and has an ornithological meaning, it is clearly related to the PJ equivalent. However, no additional sources confirm this, and all text examples that can be elicited from existing sources clearly show that *ǀǎ-sì* is the most common and neutral generic term for ‘bird’ in this language. Pending further publications of data, we prefer to disregard this form for the time being.

7. BITE (±)

- PJ: ***ǀāē** (Ju. *ǀái*, Kx. *ǀà ~ ǀe(:)*, Kg. *ǀē: ~ ǀè:*, Ek. *ǀāē-m̃*). ◇ The only divergent form is OK. *ǀā*, unless Bleek’s transcription of the dental click is erroneous (not highly likely). The Ek. form (a compound with *m̃* ‘eat’ as the second part) is essential for the reconstruction, since this is the only dialect in which the preglottalized nasalized click has been explicitly elicited. Vocalic correspondences point to the diphthong **ae* rather than **ai* as the original constituent.
- ǀHoan: **ǀái** (C).
- Ju-ǀHoan: Despite obvious phonetic similarity (click influx and vocalism match perfectly), the two forms cannot be considered a solid etymological match, since the preglottalized nasal click in PJ always corresponds to a nasal click in ǀHoan as well (see HEAD below). Nevertheless, with two features matching out of three, the unique correspondence between click effluxes may reflect some undetected contamination, or even be part of a regular pattern, undetectable due to lack of data. We count this as a potential match with low probability.

8. BLACK (-)

- PJ: ***ǀō** (Ju. *ǀó*, Kx. *ǀǒ:*, Kg. *ǀǒ ~ ǀò ~ ǀǒ̀*, OK. *ǀo ~ ǀu:*, Ek. *ǀō*).
- ǀHoan: **ǀkxau** (C).
- Ju-ǀHoan: No lexicostatistical or clearly defined etymological parallels.

9. BLOOD (+)

- PJ: ***ǀVŋ** (Ju. *ǀánŋ*, Kx. *ǀĩ*, Kg. *ǀĩ ~ ǀĩŋ*). ◇ Vocalic reconstruction is uncertain. This word is not attested in the Northern cluster, where the corresponding equivalents fluctuate between **ǀo^rru* (Ek. *ǀǒ^rlú ~ ǀú^rlú*; cf. also Kg. *ǀóru ~ ǀoru*) and **yaru* (OK. *yalo ~ yalu*, quoted as *yàlò ~ yúlà* in Snyman 1980: 34). The former of these is probably inherited, but its dialectal distribution is quite sparse compared to **ǀanŋ*; the latter has a phonetic shape that is highly atypical of Ju languages (with an initial *y-*) and is most likely of non-Ju origin.
- ǀHoan: **ǀqí** (C), **ǀí**: (S).
- Ju-ǀHoan: A plausible lexicostatistical match with recurrent phonetic correspondences. ◇ SH: 238.

10. BONE (-)

- PJ: ***!ʔú** (Ju. *!ʔú*, Kx. *!ʔú*, Kg. *!ù ~ !ú*, Gr. *!ʔú*, OK. *!ʔu ~ !ʔo*, Ek. *!ʔú*). ◇ Preserved in all daughter dialects. Ju. shows the rare extra high tone on this root, possibly an archaic feature.
- †Hoan: **ǂáˀ** (C) ← ***táˀ**. ◇ Internal phonetic reconstruction derives all palatal plosives from original coronal stops.
- Ju-†Hoan: No lexicostatistical matches; the closest etymological connection to PJ ***!ʔú** in †Hoan may be *!ʔui* ‘spine’ (C), but only provided that final *-i* can be explained away as an old suffixal extension, which is currently unclear. The †Hoan word has no known Ju equivalents. ◇ In HH: 15, the comparison of PJ ***!ʔú** to †Hoan *!ʔui* is justified by reconstructing an obscure diphthong ***-Vi** (cf. PJ ***!xɔ** = †Hoan *!xúú* ‘elephant’ for extra support), but this is not a phonologically viable explanation; it is more likely that morphological reasons are responsible for both cases.

11. BREAST (= CHEST /male/) (-)

- PJ: ***!ǂʔá** (Ju. *!ǂʔá*, Kx. *!wa*). ◇ The reconstruction is approximate due to lack of data (initial click could have been ***!ʔ** instead of ***ǂ**). Essentially an isogloss between Ju. and Kx.; a much less stable root than the far more widespread and perfectly reconstructible ***ku** ‘female breast; milk’ (→ Ju. *kù*, Ek. *kūú* etc.). Nevertheless, the only alternate candidate for PJ ‘/male/ chest’ is Ek. *ǂǂǂ*, corresponding to various forms with the meaning ‘lungs’ in Ju dialects (Gr. *sūǂǂ*, Cuito/Cuando *šǂǂ*, etc.) and probably representing the results of a metonymic semantic shift.
- †Hoan: **!ǂǂˀmā** (C, G). ◇ The meaning of this word is glossed as ‘chest (of humans)’ in Collins 2001: 458; according to the same source, this item is lexically opposed to *!ǂǂe* ‘chest (of animal)’ and *ǂxūū* ‘breast (of a non-human animal)’.
- Ju-†Hoan: No lexicostatistical matches, although, interestingly enough, both words have parallels in Taa — †Hoan **!ǂǂˀmā** is practically the same item as **!Xóǂ** *!ǂǂˀma* ‘sternum’ (the parallel may reflect either a genetic or an areal connection), while PJ ***!ǂʔá** is formally comparable with **!Xóǂ** *!ǂǂ* ‘chest’).

12. BURN (tr.) (?)

- PJ: ***!kūʔú** (Ju. *kùʔú*, Kx. *kou ~ kau*, Kg. *kuú ~ kóú ~ kóú*, Ek. *kūʔú*). ◇ Preserved in all daughter dialects, often with polysemy ‘to burn / to roast’.
- †Hoan: Not attested in reliable sources. Traill (1973: 29) quotes two different forms, *θui* and *ǂǂam*, both with the meaning ‘burn’. He does not specify, however, whether these stems are transitive or intransitive, and their existence has not yet been confirmed in published sources.

13. CLAW (= NAIL) (±)

- PJ: ***!ǂúʔú** (Ju. *!ǂúʔú*, Kx. *!uru*, Kg. *!uru*, Gr. *!uʔú*, OK. *!ulu ~ !ǂnu*, Ek. *!ǂú*). ◇ Preserved everywhere. Reliably reconstructed with a retroflex click, although the Ek. reflex **!-** instead of **ǂ-** is irregular; it may reflect a secondary contamination with ***!uru** ‘quiver’ (= Ju. *!ǂú*, etc.). Glottalic articulation in word-medial position is less certain (only attested in Ju.).
- †Hoan: **!ǂˀ** (HH), **!ǂˀo** (C).
- Ju-†Hoan: The potential relation between these two forms is problematic. The correspondence between Ju ***!ǂ** and †Hoan **!** is supported by at least two more significant examples (PJ ***!ǂǂma** ‘to enter’ — †Hoan *!ǂm* ‘to enter /plural action/’; PJ ***!ǂai** ‘puff-adder’ —

17. DIE (+)

- PJ: ***!ae** (Ju. !áí, Kx. †é: ~ †éi ~ †í, Kg. †é ~ †è, Gr. !áí, OK. †é ~ †e, Ek. †āē ~ †ē). ◊ Preserved in all daughter dialects. Singular subject action verb; the corresponding plural form is ***!ao** (Ju. !àò, Kx. †au, Kg. †au, OK. †au, Ek. †āō). Both stems are reliably reconstructed with the retroflex click, and it is tempting to trace them back to a single root (***!a-**) with different vocalic extensions. However, there is not a single other example that could hint at the productivity of this morphological operation on the PJ level; considering that all other known pairs of singular vs. plural action verbs in PJ are completely suppletive, etymological relationship between ***!ae** and ***!ao** cannot be reliably demonstrated without supporting external data.
- †Hoan: **šǐ** (G, C). ◊ The plural action equivalent is a composite form: **šǐ-ǎ̃**.
- Ju-†Hoan: The correspondence between the PJ voiced/voiceless retroflex click and the †Hoan voiced/voiceless palatal fricative is recurrent (see more examples in the entries for ‘hand’, ‘water’, corr. #35b); vocalic correspondences are also easily reconciled, and cases where sporadic nasalization in †Hoan is missing in Ju. are well known (cf. PJ ***!a** ‘to stand /plural action/’ = †Hoan **ǎ̃** id., corr. #18). This is a sufficient basis to regard both forms as etymological and lexicostatistical matches. However, the singular/plural action suppletivism of Ju finds no parallels in †Hoan.

18. DOG (-)

- PJ: ***†^ho-ĩ** ~ ***†^ho-e** (Ju. †úú ~ †òà, Kx. !ɔ, Kg. †wé ~ †wé, Gr. †^hwi: ~ η†^hwi:, OK. †wé, Ek. †hōē). ◊ Preserved in all dialects; however, there are at least three different morphological variants of this stem, with ***†^ho-ĩ** and ***†^ho-e** being the most frequent ones, and ***†^ho-a** only found in Ju. Although the origins of this diversity are unclear (probably a reflection of Pre-Proto-Ju's morphological productivity), the evidence seems to point to ***†^ho-** as the original root for all these forms.
- †Hoan: **ǎ̃amà** (C, G) ← ***tema**. ◊ The old non-palatalized variant **tēmà** is still preserved in the Sasi dialect.
- Ju-†Hoan: No lexicostatistical or clearly defined etymological parallels.

19. DRINK (±)

- PJ: ***č^hĩŋ** (Ju. č^hì, Kx. čí, Kg. šĩŋ ~ čĩŋ, Gr. š^hŋ:, OK. čí, Ek. šĩŋ). ◊ Preserved in all daughter dialects. Coda reconstruction is not fully secure, but loss of the final velar nasal in Ju. is a recurrent phenomenon, so all the listed forms are unquestionably related.
- †Hoan: **čū** (C, G).
- Ju-†Hoan: Despite some obvious phonetic similarity, it is hard to reconcile the codas: although cases in which a final velar nasal in PJ seems to leave no trace in †Hoan are relatively numerous (see corr. #16), the vocalic correspondence «PJ ***-i(ŋ)** : †Hoan **-u**» remains unique. However, precise behavior of vowels in such specific contexts (between a palatal affricate and an unstable velar nasal coda) can hardly be predicted at the present time, and this means that the parallel can be provisionally accepted as a potential «weak» match.

20. DRY (-)

- PJ: ***!kxau** (Ju. !kxáú, Kg. !áo ~ †áo ~ †ào ~ †kxáo, Gr. !^hau, Ek. †kxāō). ◊ Preserved in all daughter dialects. Original retroflex click safely reconstructed based on the correspondence between Ju. and Ek.
- †Hoan: **ǎ̃q^hau** (C).

- Ju-ǀHoan: No lexicostatistical or clearly defined etymological parallels. (An alternate synonym in Ju., /ɔ̃ ‘dry’, looks quite compatible with the form in ǀHoan, but has to be discarded as non-reconstructible for the PJ level and most likely recently borrowed from a Khoe source, along with numerous other secondary equivalents for basic terms in Ju.).

21. EAR (+)

- PJ: *|^huí (Ju. /^huí, Kx. /wí, Kg. /úí, Gr. /^hwí, OK. /wí, Ek. /^huí). ◊ Preserved in all daughter dialects. PJ reconstruction relies on the full coincidence of the Ju. and Ek. forms.
- ǀHoan: |q^hǒē (C, G), |q^hǔĩ (SH).
- Ju-ǀHoan: Correspondences are completely regular, with the uvular component of the efflux in ǀHoan matching the glottalization in Ju (for a fully identical supporting example, cf. PJ *|^hũ ‘steenbok’ = ǀHoan |q^hǒǒ id.); examples of «ǀHoan *oe* = Ju. *ui*» are scarce, but examples of «ǀHoan *o* = Ju *u*» are not (see corr. #12), so the vocalism is hardly problematic. ◊ HH: 29; SH: 253.

22. EARTH (+)

- PJ: *kxà (Ju. *kxà*, Kx. *kxa*, Kg. *kxá*, Gr. *kxá*, OK. *kxa*, Ek. *kxà*). ◊ Preserved in all daughter dialects. Reconstruction is based on the completely coinciding forms in Ju. and Ek. Neither any of the modern dialects nor, as may be inferred, PJ itself make any clear lexical differentiation between ‘earth’ and ‘sand’, due to specific landscape conditions of the Ju people.
- ǀHoan: *kxà* (C, G).
- Ju-ǀHoan: An obvious match. Complete phonetic identity between both forms is interpretable in terms of regular phonetic correspondences, i.e. there is no need to assume areal diffusion, particularly since the isogloss is exclusive to Ju and ǀHoan, but not to Tuu or Khoe. Of note, however, is the presence of the same word for ‘earth’ in Kwadi (Westphal 1966: 144), a language that is unrelated or very distantly related to Ju-ǀHoan, so in this case areal diffusion is a likely scenario. ◊ HH: 13, 24.

23. EAT (+)

- PJ: *ǀm (Ju. ?*m*, Kx. *m*: ~ *m*, Kg. *m*: ~ *m̃*, Gr. ?*m*, OK. *m* ~ *m*, Ek. *m*). ◊ Preserved in all daughter dialects. The root, just as it is attested in most dialects, should be reconstructed with a high-toned syllabic **m* preceded by a glottal stop (or a single preglottalized nasal consonant).
- ǀHoan: ǀám (C, G).
- Ju-ǀHoan: A perfect match. ǀHoan, unlike Ju languages, seems to generally lack syllabic nasal consonants, so the shift **m* → *am* is more probable than the opposite. ◊ HH: 14 (advocating for the reconstruction **am*).

24. EGG (-)

- PJ: *ǀu (Ju. *ǀu*, Kx. *ǀu*, Kg. *ǀú*, OK. *ǀu* ~ *ǀú*, Ek. *ǀũũ* ~ *ǀōú*). ◊ Preserved in all daughter dialects.
- ǀHoan: k^hǒǔē (C, G). ◊ The alternate form *ǀxui* ~ *ǀ^hxui*, found in Traill 1973: 29, is not confirmed in newer sources.
- Ju-ǀHoan: No lexicostatistical or clearly defined etymological parallels.

25. EYE (+)

- PJ: ***ǰaʔā** (Ju. *ǰàʔá*, Kx. *ǰa*, Kg. *ǰá ~ ǰáá ~ ǰá*, Gr. *ǰəʔa ~ ǰa ~ ǰaʔa*, OK. *ǰa*, Ek. *ǰàʔā*). ◇ Preserved in all daughter dialects. Initial vowel is occasionally reduced, creating an odd «voiced glottalized» click effect (although such transcriptions are only attested in old sources).
- †Hoan: **ǰōā**, pl. **ǰōē** (C, G).
- Ju-†Hoan: Despite the lack of phonetic similarity, both forms can be reconciled by means of recurrent correspondences. Examples of the †Hoan labial click corresponding to PJ dental */ also include such basic and semantically stable terms as ‘head’ q.v., ‘sky’ (†Hoan *ǰōā* = PJ **ǰaʔa*), ‘duiker’ (†Hoan *ǰəu* = PJ **ǰau*), and possibly ‘one’ q.v. For the lack of voiced articulation in †Hoan, see corr. #38a. As for the diphthong *oa* in †Hoan, labial articulation here, judging by all attested cases of words with labial clicks, is automatic after such a click (corr. #1a)¹⁷. The word should probably be reconstructed as **ǰaʔ-*, perhaps with an original paradigm of sg. **ǰaʔ-a*, pl. **ǰaʔ-i/N/*, levelled in PJ. ◇ HH: 18, 27.

26. FAT (-)

- PJ: ***ǰí** (Ju. *ǰáí*, Kx. *ǰí*, Kg. *ǰí ~ ǰai ~ ǰáie*, Gr. *ǰáí*, OK. *ǰí*, Ek. *ǰéí ~ ǰí*). ◇ Preserved in all daughter dialects (sometimes glossed as ‘fat’, sometimes as ‘oil’; there seems to be no lexical differentiation between the two meanings). Original *-i diphthongized in Ju. and several other dialects.
- †Hoan: **ǰui** ~ **ǰui** (T). ◇ Not very reliable (attested only in A. Traill's old publication).
- Ju-†Hoan: If Traill's notation for †Hoan is correct (although the strange variation between click effluxes makes it doubtful), the form is incompatible with the Ju. equivalent.

27. FEATHER (-)

- PJ: Same word as ‘hair’ q.v. (sometimes used in conjunction with ‘bird’ q.v.).
- †Hoan: Same word as ‘hair’ q.v.
- Ju-†Hoan: Same lack of lexicostatistical/etymological matches as in ‘hair’ q.v.

28. FIRE (-)

- PJ: ***dàʔá** (Ju. *dàʔá*, Kx. *dà*, Kg. *da: ~ d:ʔa ~ daá*, Gr. *dəʔa ~ daʔa*, OK. *dà ~ dàa*, Ek. *dàʔà*). ◇ Preserved in all daughter dialects. First vowel sometimes gets reduced (see ‘eye’ for the same structure).
- †Hoan: *ǰoa* (C, G).
- Ju-†Hoan: No lexicostatistical or clearly defined etymological parallels.

29. FISH (?)

- PJ: Technically reconstructible for the PJ stage as ***ǰau** (Ju. *ǰàù*, Kg. *ǰau:*, OK. *ǰau*, Ek. *ǰáú*). However, all attested forms are plausibly interpreted as borrowings from a Khoe source (cf. Proto-Khoe **ǰau* ‘fish’); these borrowings may have taken place either before the disintegration of PJ or already after, but there is no reason to think of them as inherited from a Proto-Ju-†Hoan, let alone earlier, stage of development. Ek. *ǰlō* ‘fish’ is different, but etymologically obscure.

¹⁷ In HH: 18, labial articulation of the vowel is considered to be primary in such cases, with the authors reconstructing Proto-Kxa **-oCa* → †Hoan *-o-a*, Ju **-a-a*. However, since all of their examples involve items with †Hoan labial clicks or, at least, another labial consonant in the vicinity, it makes more sense to assume secondary labialization of the vowel in †Hoan.

- ǀHoan: Not attested.
- Ju-ǀHoan: Not reconstructible. The generic term ‘fish’ may not have existed in the protolanguage at all, given the geographical localisation of its descendants.

30. FLY (-)

- PJ: Technically not reconstructible; a slightly more probable candidate for proto-status is PJ *ǀom (Ju. ǀòǀm ‘to fly’, ǀòǀm-ǀ ‘to fly over (a village)’ = Kx. ǀo:a ← *ǀom-a ‘to fly about /of birds’, OK. ǀoa ‘to mount up (in the sky)’), although all the parallels to the Ju. form are only attested in D. Bleek’s old records and are somewhat questionable, both phonetically and semantically. In Ek., no separate lexical root for the meaning ‘to fly’ is attested; cf., perhaps, ǀāō ‘to wake up, rise, stand up, fly up, jump up’ = Ju. ǀáú ‘to rise, get up’. Kx. tē, OK. tē: (as in *songu tē*: ‘the arrow flies’) are not confirmed in modern sources.
- ǀHoan: **kālā** (C, G). ◊ This form is clearly related to !Xóō *kála* ‘to go round, circle as vultures’, but the nature of the relationship (genetic? borrowing? if yes, in what direction?) remains unclear.
- Ju-ǀHoan: No lexicostatistical or etymological parallels. Overall, an unstable item that may not have had its own unique lexical representation at the Ju-ǀHoan level.

31. FOOT (≈)

- PJ: *ǀkxái (Ju. ǀkxái, Kx. ǀe ~ ǀxe:, Kg. ǀkxe ~ ǀkxi ~ ǀxí ~ ǀe ~ ǀí, Gr. ǀáí ~ ǀxái, OK. ǀkxe ~ kxe, Ek. ǀkxái). ◊ Preserved in all daughter languages.
- ǀHoan: **ǀáǀū** (C, G). ◊ Plural: ǀǎǎǎ.
- Ju-ǀHoan: No lexicostatistical parallels. However, the form in ǀHoan is comparable to PJ *ǀu^h ‘track, footprint’ (Ju. ǀù^h, Ek. ǀǎ); consonantal correspondences here are recurrent (see corr. #35, #37a), and although the vocalic correspondence is rare (corr. #5), it is not totally unique (cf. also at least ǀHoan *ǀu* = PJ *ǀau ‘duiker’), so we do not have any firm grounds to definitively reject the comparison. Acceptance of this etymology would imply that the ǀHoan form is more archaic in the meaning ‘foot’, since the semantic development ‘footprint’ → ‘foot’ is typologically far less likely than the opposite.

32. FULL (?)

- PJ: *ǀǎǎǎ (Ju. ǀǎǎǎ, Kx. ǀē ~ ǀēī, Ek. ǀǎǎǎ). ◊ Preserved in all daughter dialects.
- ǀHoan: Not attested.

33. GIVE (-)

- PJ: *ǀaǎa (Ju. ǀǎ, Kx. ǀǎ ~ ǀa:, Kg. ǀǎ ~ ǀǎ ~ ǀǎ ~ ǀá, OK. ǀa ~ ǀa:, Ek. ǀǎǎ ~ ǀǎ). ◊ This is the most common equivalent for ‘give’ in most of the dialects. Vocalic reconstruction is unclear: technically, the coda -aǎa accounts for most of the attested variations, but some of the developments would still have to be irregular (such as the contraction *ǀaǎa → ǀa in some of the dialects). It is also unclear whether nasalization of the vowel has to be set up as a PJ feature or if it appears in Ju. and some other dialects secondarily. An additional PJ root is *ǀna ~ *ǀneǎe, whose functions seem to be restricted to the imperative throughout: Ju. ǀnà, Kx. *na*, Gr. *na:*, OK. *na*, Ek. *nèǎè*. Finally, Kx. and Gr. yield evidence for a third root, *ǀau ‘give’, whose semantic difference from *ǀaǎa cannot be established from available sources; strange enough, it is not confirmed at all by more modern and reliable sources on Ju dialects.
- ǀHoan: **ǀú** (C, G).
- Ju-ǀHoan: No lexicostatistical or clearly defined etymological parallels.

34. GOOD (-)

- PJ: ***ǰã** (Ju. *ǰã*, Kg. *ǰã ~ ǰa*, Gr. *ǰa: ~ ǰa:*). ◇ This is the most common and probably the inherited term for ‘good’ in Ju dialects. Several other phonetically similar forms, such as Kx. *!ãĩ*, Ek. *kãhĩ*, Gr. *gãĩ*, etc., do not correspond regularly to each other and are most plausibly explained as borrowings from various Khoe sources; cf. Proto-Khoe **!ãĩ* ‘good’ → Nama *!ãĩ*, !Ora, Naro *!ãĩ*, East Khoe **kãĩ*, etc. (Vossen 1997: 445).
- †Hoan: **q^hǎě** (C, G).
- Ju-†Hoan: No lexicostatistical or clearly defined etymological parallels. The †Hoan form is clearly the same word as !Xóó *qáĩ* ‘pretty, beautiful, nice’, but whether the similarity is due to common ancestry or recent borrowing remains unclear.

35. GREEN (-)

- PJ: ***!ǎŋu** (Ju. *!ǎũ^h*, Kx. *!ǎũ*, Kg. *!ǎŋ ~ !ǎŋ ~ !ǎŋ ~ !ũŋ*, OK. *!ǎŋ*, Ek. *!ǎŋ*). ◇ Judging by available semantic notation, the root must have denoted the entire ‘blue/green/yellow’ spectrum in PJ. Reconstruction of the coda **-aŋu* is set up to account for the correspondence between Ju *-ǎũ* and Ek. *-aŋ*.
- †Hoan: **za^hʔa** (T). ◇ Attested only in A. Traill's old publication, so somewhat dubious.
- Ju-†Hoan: No lexicostatistical or clearly defined etymological parallels.

36. HAIR (-)

- PJ: ***!kxúí** (Ju. *!kxúí*, Kx. *!kxwe ~ !kxwi ~ !kwi*, Kg. *!kxwé ~ !kxwí*, Gr. *!kxwi*, OK. *kxwi ~ !wi*, Ek. *!kxúí*). ◇ Preserved in all daughter dialects. No lexical difference between ‘head hair’ and ‘body hair’.
- †Hoan: **ǰu** (C), **ǰù** (SH).
- Ju-†Hoan: No lexicostatistical or clearly defined etymological parallels.

37. HAND (+)

- PJ: ***!ǰau** (Ju. *!ǰáú*, Kx. *!lou ~ !lau*, Kg. *!ǰau*, Gr. *!ǰàù*, OK. *!ǰau*, Ek. *!ǰàò ~ !ǰāò*). ◇ Preserved in all daughter dialects. Correspondences indicate an original voiced retroflex click, still preserved in the Grootfontein dialect.
- †Hoan: **ǰíu** (C, G), **síu** (SH). ◇ The more archaic Sasi form is *ǰáú*. Odd consonantal gradation in the plural form: *ǰěō-qà* (C, G).
- Ju-†Hoan: PJ **!ǰau* and †Hoan *ǰíu* are tied together by recurrent correspondences (#5a, #35b, #38a) and can be reliably traced back to the same proto-root. Cf. a perfect near-homonymous example in Ju. *!àú* ‘to dig’ (← **!ǰau?*) = †Hoan *ǰíu* (C) ‘to dig’. ◇ The possible connection is mentioned, but rejected in HH: 17, because the authors have not located the additional evidence for this correspondence.

38. HEAD (+)

- PJ: ***ǰǎĩ** (Ju. *ǰáí*, Kx. *ǰé ~ ǰé ~ ǰí ~ ǰí*, Kg. *ǰé*, Gr. *ǰé*, OK. *ǰé*, Ek. *ǰǎĩ*). ◇ Preserved in all daughter dialects. Preglottalized nasal click is reconstructed based on Ekoka data.
- †Hoan: **ǰǎú** (C, G), **ǰǎú** (SH).
- Ju-†Hoan: Despite the lack of phonetic similarity, Ju and †Hoan forms are connected by recurrent correspondences. The labial click in †Hoan corresponds to the dental click in Ju (corr. #32a), while the preglottalized nasalized efflux in Ek. and †Hoan coincide precisely. Labial vowel articulation in †Hoan is automatic after a labial click, and nasalization of the vowel may be secondary (influence of the nasal click, or a trace of some old

morphological feature). The hypothetical protoform would presumably look like $*\tilde{\theta}e-$ on the Proto-Ju-ǀHoan level¹⁸.

39. HEAR (+)

- PJ: $*s\grave{a}\tilde{\theta}\bar{a}$ (Ju. *càṽá*, Kx. *cá ~ čṽa*, Kg. *sá ~ ssá ~ ssaá ~ saá*, Gr. *sóṽá*, OK. *sáa ~ saa*, Ek. *čà ~ čàṽá*). ◊ Preserved in all daughter dialects. Fluctuation between affricate (*c-*) and fricative (*s-*) articulation is resolved in favor of the fricative articulation as original; affricativization probably occurs under the influence of the glottal stop, especially considering that the first half of the complex vowel sequence is frequently reduced or even completely deleted in the actual articulation.
- ǀHoan: $c\bar{a}$ (C, G).
- Ju-ǀHoan: The forms are perfectly compatible (ǀHoan *c-* is a regular correspondence for PJ $*s-$; lack of the glottal stop in ǀHoan is the same as in ‘eye’ q.v.). ◊ HH: 23 (reconstructed with $*c-$).

40. HEART (-)

- PJ: $*!kx\bar{a}$ (Ju. *!kxá*, Kx. *!a*, Kg. *!kxá ~ !xá*, Gr. *!ṽa*, OK. *kxa*, Ek. *!kxā*). ◊ Preserved in all daughter dialects. The velar affricate efflux is transcribed inconsistently in old sources, but these inconsistencies are not enough to amend the reconstruction, based on modern data from Ju. and Ek. Most of the dialects also reflect polysemy ‘heart/inside’, likely inherited from the PJ state.
- ǀHoan: $!q\tilde{\theta}$ (C, G).
- Ju-ǀHoan: There are no other plausible cases where PJ $*kx$ (either as a non-click phoneme or as a click efflux) could be correlated with ǀHoan $*q\tilde{\theta}$; vocalic correspondences cannot be properly resolved, either, implying that the two forms are not related.

41. HORN (+)

- PJ: $*!h\acute{u}$ (Ju. *!hú*, Kx. *!u: ~ !ú*, Kg. *!ú ~ !hú ~ !xú*, Gr. *!!ú*, OK. *!hú*, Ek. *!hú*). ◊ Preserved in all daughter dialects. Doke's transcription of a retroflex click for the Grootfontein dialect, instead of an alveolar one, is most likely erroneous, since it is not supported by any data outside that source.
- ǀHoan: $!h\check{o}$ (G). ◊ Also attested in the reduplicated variant $!h\check{o}-!h\check{o}$.
- Ju-ǀHoan: A perfect etymological/lexicostatistical match with regular correspondences. ◊ HH: 28.

42. I (+)

- PJ: $*m\acute{i}$ (Ju. *mí*, Kx. *m ~ me ~ mi*, Kg. *m ~ mé ~ mi ~ mĩ*, Gr. *mí*, OK. *m ~ me ~ mi*, Ek. *mí ~ mā*). ◊ Preserved in all daughter dialects, along with the emphatic stem $*mi-hi$. A very rare case of word-initial $*m-$, reliably reconstructed for the PJ level.
- ǀHoan: **ma** (C, G). ◊ Cf. also the possessive form: *ṽam* ‘my’.
- Ju-ǀHoan: A clear match, although the vocalism remains unclear. Considering that both *mí* and *mā* are encountered in Ek., partially distributed depending on syntactic function (König & Heine 2001: 49), it is possible that both variants were already present in Proto-Ju-ǀHoan. ◊ HH: 14.

¹⁸ In [Sands, Honken 2014: 249] it is tentatively suggested that the ǀHoan form may be related to $!X\acute{o}\tilde{\theta} \tilde{\theta}\acute{u}:$ ‘louse’ as a loan. Despite the phonetic similarity (involving a relatively rarely encountered labial click), a semantic shift from ‘head’ to ‘louse’ or vice versa is so completely unprecedented that the Ju-ǀHoan etymology must take precedence here.

43. KILL (+)

- PJ: *^hú (Ju. ^hú, Kx. !ú, Kg. !úŋ ~ !^hú ~ !^húŋ ~ !xú, Gr. !^hú:, OK. !ú ~ !xú, Ek. !^húŋ ~ !^hú).
 ◇ Singular action stem; the corresponding plural action stem is harder to reconstruct, since the two main attested forms, Ju. !^hóǎ (← *!^hóǎ, cf. Kx. //^hoǎ id.) and Ek. //^híj, do not properly correspond to each other. Amendment of the reconstruction to *!^huŋ (cf. the variation in Ek.) is not out of the question, but on the whole, the correspondences seem rather suggestive of a nasalized vowel as the original coda.
- †Hoan: !^hǒ (G). ◇ Singular action stem; the plural correlate is *θōā* (C, G).
- Ju-†Hoan: Singular action stems correspond to each other precisely; their plural action correlates seem to be less stable and are historically incompatible. ◇ HH: 19, 28.

44. KNEE (≈)

- PJ: *^hxòǎ (Ju. ^hxòǎ, Kx. !*wa-ǎ*, Kg. !óǎ ~ !xóǎ, Gr. !*xwa*, Ek. ^hxòǎ). ◇ In OK., the only attested equivalent for ‘knee’ is *ǎm* ~ *gǎm* (the second variant shows irregular click loss) = Ek. *ǎm* ‘knee-cap’; this is possibly an archaic root with this meaning, whereas most of the modern dialects use the compound form ‘knee-head’ (e.g. Ek. ^hxòǎ *ǎé*; cf. also the form in Kx.) instead.
- †Hoan: †^hemē (C, G). ◇ Cf. also //*dam* ‘to kneel’.
- Ju-†Hoan: No lexicostatistical matches. However, the †Hoan form is comparable to Ju. †^hòm ‘to kneel’ (no known parallels in other Ju dialects); discrepancies in vocalism may imply that the †Hoan form is an old derivative from the verb ‘to kneel’ — *^hòm-e → †^hem-ē with assimilation (of note is the ultra-low tonal characteristics in both languages).

45. KNOW (-)

- PJ: *!^hǎ (Ju. !^hǎ, Kg. !^hǎ ~ !^hǎ, Gr. !^hǎ). ◇ This root is only preserved in the Southern dialect cluster. Its main alternative is OK. *ǎi*, Ek. *ǎhī* ~ *ǎhī* ~ *ǎhī*, which corresponds to the widespread (but not attested in Ju.) root **ǎi* (Snyman 1997: 94) that means ‘to be able to, to know how (to do smth.)’ in several other dialects. This lexical distinction may be set up for the PJ level (*!^hǎ ‘to know smth.’ — **ǎi* ‘to know how to do smth.’).
- †Hoan: *ǎi* (C, G). ◇ Probably the same word as ‘to see’ q.v.; distinct from *ǎi* ‘to know how’.
- Ju-†Hoan: PJ **ǎi* is clearly the same root as †Hoan *ǎi*, reflecting a common Proto-Ju-†Hoan root with the meaning ‘to be able, to know how’ (HH: 28). However, there are no direct lexicostatistical matches for the required meaning ‘to know (smth.)’: †Hoan has seemingly merged this meaning with ‘to see’, whereas PJ *!^hǎ is either archaic or may be itself borrowed from a Khoe source (cf. Proto-Khoe *!^hǎ ‘to know’).

46. LEAF (-)

- PJ: (?) *!^hoa (Kg. !*wá*, Ek. !*ǎ*). ◇ The basic form for ‘leaf’ in Ek. corresponds to Ju. *ǎ* ‘wet leaf’ (with a slight irregularity, since there is no pharyngealization in Ju.) and possibly to OK. *gǎ*: ‘leaf’ with irregular click loss. The following alternate roots have been excluded from comparison for various reasons: (a) Ju. *dǎrà*, Kx. *dǎra* ‘leaf’; this is an areal isogloss with Naro *toǎrà* ‘leaf’, a word that also lacks a proper Common Khoe etymology and should probably be ascribed to some local substrate, affecting geographically contiguous areas occupied by Ju|hoan and Naro speakers; (b) Kg. *ǎǎbu*, Gr. *ǎǎbu* ‘leaf’ = Ju. *ǎǎbu* ‘to come into leaf (vb.); leaf, cabbage (n.)’, a root that is hard to separate from the phonetically similar *ǎǎbu* ‘to swell, be bloated’ and whose primary meaning is most likely verbal.

- ǀHoan: $zò^{\prime}bā$ (C, G). ◊ Cf. Sasi $dò^{\prime}βā$ id., reflecting the original coronal articulation.
- Ju-ǀHoan: No lexicostatistical matches. It is tempting to compare Ju. $dò^{\prime}rà$ with Sasi $dò^{\prime}βā$, since the first syllable of both words is identical (right down to the pharyngealized articulation of the vowel); however, the second syllable is a stark mismatch, and given the observations about the areal connection between Ju|hoan and Naro, it is possible that the ǀHoan form was borrowed separately from a distinct dialect of the same substrate (e.g. if $*do^{\prime}ra$ and $*do^{\prime}ba$ were morphologically different variants in these dialects). In any case, unless more correspondences between nominal structures $*CV-ra$ and $*CV-ba$ are discovered between Ju and ǀHoan, it is premature to speak about common inherited lexemes in this particular case. ◊ In HH: 19, the Ju-ǀHoan match is accepted, but no explanation is provided for the morphological differences, and the areal distribution of the Ju form is not taken into account either.

47. LIE (-)

- PJ: $*šú$ (Ju. $šú$, Kx. $šu \sim šú$, Kg. $šú \sim šù$, Gr. $šú$, OK. $šú \sim ču$, Ek. $šú$). ◊ Singular action verb; its plural action correlate is PJ $*šà$ (Ju. $šà$, Ek. $šà$). Preserved in all daughter dialects.
- ǀHoan: $ǀqíʔi$ (C, G). ◊ Singular action verb; plural action correlate is $!q^hǎu$.
- Ju-ǀHoan: No lexicostatistical matches. It is possible to compare ǀHoan $!q^hǎu$ ‘to lie (pl.)’ with PJ $*šà$ ‘to sit (pl.)’ (q.v.), with a slight semantic shift and generally regular phonetic correspondences; however, no clear etymological parallels for ǀHoan $ǀqíʔi$ have been found in PJ.

48. LIVER (-)

- PJ: $*čij$ (Ju. $čij$, Kx. $čĩ$, Kg. $čij$, Gr. $šij$, OK. $čĩ$, Ek. $šij$). ◊ Preserved in all daughter dialects. The correspondence between Ju. $č^{(h)}$ - and Ek. $š$ - is irregular, possibly reflecting a specific development $*č \rightarrow š$ before a syllabic nasal.
- ǀHoan: $kúí$ (C, G).
- Ju-ǀHoan: No lexicostatistical or clearly defined etymological parallels.

49. LONG (-)

- PJ: $*ǀaʔŋ$ (Ju. $ǀǎʔí$, Kx. $ǀē$, Kg. $ǀaʔáŋ \sim ǀáʔŋ \sim ǀaʔíŋ$, Gr. $ǀaʔŋ \sim ǀǎʔŋ$, Ek. $ǀǎʔŋ$). ◊ Preserved in the majority of daughter dialects; the only deviation is found in OK., where Bleek lists $ǀxana$ as the main equivalent for ‘long’ = Ju. $ǀxá$ ‘far’, perhaps with additional suffixation. Coda reconstruction is not entirely secure, but the presence of a segmental nasal $*-ŋ$ most certainly is.
- ǀHoan: $čǎʔa$ (C, G). ◊ Usually glossed as ‘tall’, but also as ‘long’ in Traill 1973: 30; there is no reason to suspect any lexical differentiation between the two submeanings.
- Ju-ǀHoan: No additional evidence has been uncovered so far for the possibility of click affricativization $*ǀ \rightarrow č$ in ǀHoan, so the two forms have to be judged as incompatible.

50. LOUSE (+)

- PJ: $*cʰij$ (Ju. $cʰij$, Ek. $šij$). ◊ Attested only in modern sources, but well reconstructible for PJ based on the Ju.-Ek. isogloss, although the voicing in Ek. is irregular (additional dialectal forms in Snyman 1997 show that voiceless $*cʰ$ is primary).
- ǀHoan: $cʰí$ (G). ◊ After Honken 1988: 64, cf. also $cʰi$: in Traill 1973: 30.
- Ju-ǀHoan: Both forms are perfectly compatible (see corr. #16 for the lack of the nasal coda in ǀHoan).

51. MAN (-)

- PJ: *!ʰᵃᵗᵃ̃ (Ju. !ʰᵃᵗᵃ̃, Kx. !wᵃ̃, Kg. !ᵗ ~ !ù ~ !ùᵗ ~ !ʰᵗ, OK. !ᵗ). ◇ In some dialects, the only form attested in the meaning ‘man = male human being / husband’ is PJ *!ᵗᵃ̃ (Gr. !ᵗᵃ̃, Ek. !ᵗᵃ̃); in Ju. and several other dialects, however, the two roots are well distinguished, so that *!ʰᵃᵗᵃ̃ has the more narrow meaning ‘male human being’ and *!ᵗᵃ̃ has the wider meaning ‘male’ (including, or sometimes restricted to, male animals). This is likely to have been the situation in PJ. Phonetically, the reconstruction *!ʰᵃᵗᵃ̃ is problematic; attested variants presuppose at least four irreconcilable variants (Ju. !ʰᵃᵗᵃ̃, Kx. *!ᵗᵃ̃, Kg. *!ᵗ, OK. /Snyman/ !xᵗ = Ek. !xᵗᵗ ‘!Xun person’). Mechanistically tracing them back to four different proto-entries does not make sense; it is more likely that the Ju. variant is the most archaic, whereas the old sources fail to properly transcribe aspiration. As for the velar fricative efflux -x- in the Northern cluster, it may have appeared secondarily through contamination with *!xuni ‘to live, reside’ (cf. Ek. !xᵗᵗᵗ, etc.). Finally, worth noting is the suppletive plural form *!ᵗᵃ̃ᵉᵗ ‘men, males’: Ju. !ᵗᵃ̃ᵉᵗ, Gr. !ᵗᵃ̃ᵉᵗ, Ek. !ᵗᵃ̃ᵉᵗ.
- †Hoan: !ᵗᵃ̃ᵉᵗ-!ᵗᵃ̃ᵉᵗ (C, G). ◇ Suppletive plural: !ᵗᵃ̃ᵉᵗ ‘men’. The second part of this compound by itself (!ᵗᵃ̃ᵉᵗ) is used in the meaning ‘husband’.
- Ju-†Hoan: No lexicostatistical or clearly defined etymological parallels.

52. MANY (≈)

- PJ: *!ʰᵃᵗᵃ̃ (Ju. !ʰᵃᵗᵃ̃, Kx. !ʰᵃᵗᵃ̃, Kg. !ʰᵃᵗᵃ̃ ~ !ʰᵃᵗᵃ̃ᵉᵗ ~ !ʰᵃᵗᵃ̃ᵉᵗᵉᵗ, Gr. !ʰᵃᵗᵃ̃, OK. !ʰᵃᵗᵃ̃ ~ !ʰᵃᵗᵃ̃ᵉᵗ ~ !ʰᵃᵗᵃ̃ᵉᵗᵉᵗ, Ek. !ʰᵃᵗᵃ̃). ◇ Preserved in all daughter dialects. Click efflux is reconstructed as simple aspiration, despite the (probably erroneous) transcription with a glottal stop in Doke's Grootfontein materials. The original vowel is *-i, undergoing regular diphthongization in Ju.
- †Hoan: !ᵗᵃ̃ᵉᵗᵉᵗ-!ᵗᵃ̃ᵉᵗᵉᵗ (C, G). ◇ Cf. the form without the plural prefix in Traill 1973: 30: !ᵗᵃ̃ᵉᵗᵉᵗ ~ !ᵗᵃ̃ᵉᵗᵉᵗᵉᵗ.
- Ju-†Hoan: No lexicostatistical parallels, but PJ *!ʰᵃᵗᵃ̃ is cognate with †Hoan !ʰᵃᵗᵃ̃ ‘big’ q.v.

53. MEAT (≈)

- PJ: *!ʰᵃᵗᵃ̃ (Ju. !ʰᵃᵗᵃ̃, Kx. !ʰᵃᵗᵃ̃ ~ !ʰᵃᵗᵃ̃ᵉᵗ ~ !ʰᵃᵗᵃ̃ᵉᵗᵉᵗ, Kg. !ʰᵃᵗᵃ̃ ~ !ʰᵃᵗᵃ̃ᵉᵗ ~ !ʰᵃᵗᵃ̃ᵉᵗᵉᵗ, Gr. !ʰᵃᵗᵃ̃ ~ !ʰᵃᵗᵃ̃ᵉᵗ, OK. !ʰᵃᵗᵃ̃, Ek. !ʰᵃᵗᵃ̃). ◇ Preserved in all daughter dialects; correspondences indicate an original retroflex click, still preserved in the Grootfontein dialect.
- †Hoan: !ᵗᵃ̃ᵉᵗᵉᵗ (C, G). ◇ Cf. also !ᵗᵃ̃ᵉᵗᵉᵗ: ‘animal’ (Traill 1973: 29): possibly the same root or even the same word (mistranscribed?), considering the natural polysemy ‘meat/animal’ in South African languages.
- Ju-†Hoan: In HH: 14, the †Hoan form is compared with Ju. !ᵗᵃ̃ᵉᵗᵉᵗ ‘to slaughter; to cut meat’; this is acceptable if the †Hoan noun originally meant something like ‘stripe/slice of meat’, i.e. represented a nominalization of the original verb. However, this is obviously not a lexicostatistical match.

54. MOON (≈)

- PJ: *!ᵗᵃ̃ᵉᵗᵉᵗ (Ju. !ᵗᵃ̃ᵉᵗᵉᵗ, Kx. !ᵗᵃ̃ᵉᵗᵉᵗ, Kg. !ᵗᵃ̃ᵉᵗᵉᵗ ~ !ᵗᵃ̃ᵉᵗᵉᵗᵉᵗ ~ !ᵗᵃ̃ᵉᵗᵉᵗᵉᵗᵉᵗ, Gr. !ᵗᵃ̃ᵉᵗᵉᵗ, OK. !ᵗᵃ̃ᵉᵗᵉᵗ ~ !ᵗᵃ̃ᵉᵗᵉᵗᵉᵗ, Ek. !ᵗᵃ̃ᵉᵗᵉᵗ). ◇ Preserved in all daughter dialects. Correspondences clearly indicate a retroflex click in PJ.
- †Hoan: !ᵗᵃ̃ᵉᵗᵉᵗ (C, G).
- Ju-†Hoan: No lexicostatistical parallels. However, the †Hoan form is comparable with Ju. !ᵗᵃ̃ᵉᵗᵉᵗᵉᵗ ‘shiny’; all correspondences are recurrent (for the possible assimilative development *CaCI → CiCI in †Hoan, see ‘knee / to kneel’ above plus additional examples, e.g. †Hoan !ᵗᵃ̃ᵉᵗᵉᵗᵉᵗ ‘dove’ = Ju. !ᵗᵃ̃ᵉᵗᵉᵗᵉᵗᵉᵗ id., †Hoan !ᵗᵃ̃ᵉᵗᵉᵗᵉᵗᵉᵗ ‘to refuse’ = Ju. !ᵗᵃ̃ᵉᵗᵉᵗᵉᵗᵉᵗ ‘to dissuade’,

corr. #1b), and the semantic shift from ‘to shine, shiny’ to ‘moon’ belongs to the trivial type. This would imply that the PJ equivalent for ‘moon’ may be more archaic. An alternate (but, it must be noted, not necessarily mutually exclusive) comparison for Ju. ǀàǀàbè is ǀHoan ǀàǀnna ‘white’ q.v.

55. MOUNTAIN (-)

- PJ: *ǀòm (Ju. ǀòm, Kx. ǀum, Kg. ǀum, Gr. ǀù:m, OK. ǀum, Ek. ǀòm). ◊ Preserved in all daughter dialects. There is, however, an unresolved problem connected with the polysemy ‘stone/mountain’. Both of the major sources on Ekoka (König & Heine 2008; Heikkinen 1986) agree that this word, glossed as ‘hill’, is pronounced with a preglottalized nasal click rather than the regular nasal click, while ‘stone’ just has the regular nasal click — i.e. that we are dealing with two different roots. This is a very odd observation, considering the frequent and natural character of the ‘stone/mountain’ polysemy in African (and world) languages and, at the same time, the impossibility to explain this difference in terms of morphological derivation. It may be further noted that C. Doke also marks a difference between the two words, but in his notation it is purely tonal (ǀù:m ‘mountain’ vs. ǀū:m ‘stone’), and while such a differentiation may be easier to explain in terms of historical derivation (tonal alternations actually exist in Ju), it can hardly be correlated with the difference in click effluxes as observed in Ekoka. Naturally, since Ekoka is the only dialect in which the difference between the two types of nasal clicks has been systematically observed and notated, in all the other dialects *ǀòm ‘mountain’ and *ǀòm ‘stone’ would be expected to look completely identical.
- ǀHoan: ǀu (C).
- Ju-ǀHoan: No lexicostatistical or etymological parallels.

56. MOUTH (±)

- PJ: *cǀi (Ju. cǀi, Kx. ci ~ ci:, Kg. ci ~ cì, Gr. cǀi:, OK. cí ~ cí:, Ek. čǀi). ◊ Preserved in all daughter dialects.
- ǀHoan: ǀǐ: (C, G).
- Ju-ǀHoan: Although the basic consonantal structures (as well as the main vowel) for both forms are perfectly compatible, and nasalization in ǀHoan is negligible, proper common etymologization is seriously hampered by the fact that PJ *cǀ- would be expected to correspond to ǀHoan cǀ- rather than ǀ- (see ‘sleep’, ‘tooth’), whereas ǀHoan ǀ-, in turn, corresponds to either PJ *ǀ- or a retroflex click, but not *cǀ-. The only way to circumvent this issue would be to set up a more complex protoform, e.g. *siǀi, with an irregular (or a contextually unique) reduction + affricativization → *sǀi → *cǀi in PJ and contraction + palatalization → *si: → ǀi: in ǀHoan. Whether this scenario can be plausibly justified remains to be seen; however, it is not out of the question, and given the undeniable phonetic similarity between the two words, we can count this entry as a potential «weak» match.

57. NAME (+)

- PJ: *ǀú (Ju. ǀú, Kx. ǀú ~ ǀú:, Kg. ǀù ~ ǀú ~ ǀú:, Gr. ǀú, OK. ǀú, Ek. ǀú). ◊ Preserved in all daughter dialects. Unexpected dental (rather than alveolar) click transcription in Doke's Grootfontein data might simply be a misprint.
- ǀHoan: ǀo (C, G).
- Ju-ǀHoan: A perfect etymological and lexicostatistical match. ◊ HH: 17, 25 (reconstructed with the diphthong *ou to reflect the regular correspondence between Ju *u and ǀHoan *o).

58. NECK (-)

- PJ: ***!àŋi** (Ju. *!àṅ*, Kx. *!ēṅ*, Kg. *||āŋ ~ ||āũŋ*, Gr. *!!āŋ*, OK. *||āŋ*, Ek. *||āŋ*). ◇ Preserved in all daughter dialects. Retroflex click securely reconstructed, based on regular correspondences between Ju., Gr., and Ek. Coda correspondences fall under the recurrent pattern «Ju. *-āṅ ~ -āũ* : Ek. *-aŋ*» that we provisionally mark as reflecting PJ **-aŋi* and **-aŋu* respectively.
- †Hoan: **ɕ^hyiā ~ ɕ^ha** (C, G).
- Ju-†Hoan: No lexicostatistical or etymological parallels.

59. NEW (+)

- PJ: ***ze** (Ju. *zé ~ zàṅ*, Kx. *zé*, Kg. *zε-ma*, Gr. *ze:*, OK. *ze*, Ek. *žèhè*). ◇ Preserved in all daughter dialects. Correspondences are mostly regular and trivial, with the exception of the tonal pattern that ranges from simple rising in Ju. to ultra-low in Ek. It is not quite clear if the Ju. form *zàṅ*, with a diphthong and pharyngealization, is merely a dialectal variant or a different root.
- †Hoan: **zà** (C, G).
- Ju-†Hoan: An acceptable lexicostatistical match. The vocalic correspondence between PJ **e* and †Hoan *a* (rather than *e* or *i*) is almost unique, but it should be noted that monophthongic **e* is very rare in PJ, and examples of PJ-†Hoan matches where it is present are even more scarce; for possible confirmation of regularity, cf. PJ **h=e* — †Hoan *ha* ‘this’ q.v., PJ **ge* ‘to stay /in a place/’ — †Hoan *ga* ‘to rise /pl./’ (although the latter example is semantically questionable). Additionally, it is not excluded that PJ **ze* ← **za-I* with suffixation (cf. in that respect the odd variant *zàṅ* in Ju.).

60. NIGHT (-)

- PJ: ***!ú** (Ju. *!ú*, Kx. *!u ~ !u:* ~ *!ú*, Kg. *!ú ~ !u*, Gr. *!ú:*, OK. *!ú*, Ek. *!ú*). ◇ Preserved in all daughter dialects.
- †Hoan: **ɕ^hǎo** (C, G).
- Ju-†Hoan: No lexicostatistical matches. Cf., perhaps, †Hoan *!u* (C) ‘yesterday’ as a possible etymological match for the PJ form (although the semantic link is problematic).

61. NOSE (-)

- PJ: ***ckxúŋ** (Ju. *ɕú*, Kx. *čũ*, Kg. *s:ù ~ cũ ~ càŋ ~ cáũ*, Gr. *ɕũ*, OK. *cuŋ ~ cáŋ ~ čn*, Ek. *čkxán ~ čŋ*). ◇ Preserved in all daughter dialects; however, correspondences here are rare and complex. In the initial position, Ek. and some other dialects point to an original affricate cluster **ckx-* that must have been phonologically opposed to the simple glottalized affricate **čʔ-* in PJ. The coda contains a velar nasal, presumably with a preceding labial vowel (reflecting the correspondence «Ju. *-ũ* : Ek. *-(a)ŋ*»), although this particular part of the reconstruction is provisional.
- †Hoan: **!q^ʔǒ** (C, G).
- Ju-†Hoan: Despite some phonetic similarity between the two forms (vocalism, glottalic articulation, etc.), there is no evidence to support click loss in PJ (or secondary click formation in †Hoan).

62. NOT (+)

- PJ: ***!ōā** (Ju. *!óá*, Kx. *!wa ~ !ua:* ~ *!á*, Kg. *!úú ~ !úi*, OK. *!wa ~ !we ~ !wí ~ kwé ~ kwí*, Ek. *!ōā*). ◇ Preserved in all daughter dialects. Note irregular click loss in some OK. subdialects, possibly caused by frequent usage of this auxiliary morpheme.

- ǀHoan: $|\text{r}^{\text{h}}\text{õ} \sim |\text{h}\text{õ}\text{?}\text{õ}$ (C, G). \diamond The Sasi form is recorded as $|\text{ũ}$.
- Ju-ǀHoan: Although correspondences between click effluxes are clearly irregular, unexplained fluctuation is already observable on the synchronic level within ǀHoan itself; taking into consideration the auxiliary (grammatical) function of this negative particle, making it more prone to various irregular developments (e.g. of an assimilative nature, or resulting from undetected contractions with other auxiliary morphemes, etc.), we tentatively count this pair, reduced to the basic shape $*|\text{U-}$, as an etymological and lexicostatistical match.

63. ONE (+)

- PJ: $*|\text{è}\text{?}\text{e}$ (Ju. $|\text{è}\text{?}\text{é}$, Kx. $|\text{e} \sim |\text{é}\text{é}$, Kg. $|\text{é}\text{é}$, Gr. $|\text{h}\text{é}$, OK. $|\text{é}$, Ek. $|\text{è}\text{?}\text{è} \sim |\text{è}$). \diamond Preserved in all daughter dialects.
- ǀHoan: $\text{õ}\text{ú}$ (C, G).
- Ju-ǀHoan: Despite apparent phonetic dissimilarity, correspondences between these two items are actually quite regular: ǀHoan õ is a perfect match for PJ $*|\text{/}$, and differences in vocalism are explained by the general labialization of vowels in ǀHoan after a labial click (actually, this is the same correspondence as in ‘head’ q.v.).

64. PERSON (\approx)

- PJ: $*\text{ǀ}\text{ù}$ (Ju. $\text{ǀ}\text{ù}$, Kx. $\text{ǀ}\text{u}$, Kg. $\text{ǀ}\text{ú} \sim \text{ǀ}\text{ú}$, OK. $\text{ǀ}\text{u} \sim \text{ǀ}\text{u}$). \diamond This is one of the few Ju words that may have relied on tonal alternations to form the plural, something that is still preserved in modern Ju. (pl. $\text{ǀ}\text{ú}$); cf., however, such plural forms as Kx. $\text{ǀ}\text{u}:\text{-si}$, Kg. $\text{ǀ}\text{u}:\text{-si}\text{h}$, indicating productive analogical reformation in various dialects. The word occasionally gets lost or semantically shifted: cf. Gr. atà ‘person’ (an unclear replacement, seemingly of non-native origin due to its violation of standard Ju phonotactics; in the plural number, however, the old word is still retained as $\text{ǀ}\text{ù}:\text{-} \sim \text{ǀ}\text{ú}:\text{-}$ ‘people’); Ek. $|\text{x}\text{ũ}\text{ú}$ ‘person’, with the old word $\text{ǀ}\text{ù}$ apparently shifting to the pronominal meaning ‘we /excl./’. Special mention must be made of the compound form $*\text{ǀ}\text{ù-}|\text{h}\text{õ}\text{ã}$, lit. ‘true person’, denoting North Khoisan-speaking people; given its presence in both Ju. and Ek., it is reconstructible for PJ as an archaic ethnic self-designation.
- ǀHoan: $\text{ǀ}\text{àm-kò:e}$ (G). \diamond Clearly a compound; second part may be a general morpheme for denoting people (cf. $|\text{á}\text{?}-\text{kò:e}$ ‘Bushman’) and is possibly of Central Khoisan origin (cf. Proto-Khoe $*\text{k}^{\text{h}}\text{oe}$ ‘person’). The first part, however, cannot be explained away as a borrowing. Suppletive plural: $\text{ǀ}\text{ò}\text{ò-}|\text{á}\text{?}\text{è}$ ‘people’ (C, G).
- Ju-ǀHoan: No lexicostatistical matches. In Ju., the word $\text{ǀ}\text{àm}$ means ‘south’; this agrees with the etymology of the exoethnonym ‘ǀHoan’ (= $|\text{Xó}\text{õ} \text{ǀ}\text{q}^{\text{h}}\text{ũ}\text{ã}$ ‘south’) and formally permits to reconstruct Proto-Ju-ǀHoan $*\text{ǀ}\text{am}$ ‘South’. If so, the Ju equivalent for ‘person’ is probably more archaic, which is made even more likely by its non-trivial paradigmatic features (tonal alternation as a grammatical means). Unfortunately, regular correspondences to PJ $*\text{ǀ-}$ in ǀHoan remain unknown, so the word $*\text{ǀ}\text{ù}$ could be compared to either (a) ǀHoan $\text{ǀ}\text{ò}\text{ò-}$ in $\text{ǀ}\text{ò}\text{ò-}|\text{á}\text{?}\text{è}$ ‘people’ (where the second component is an additional plural marker) or (b) ǀHoan $\text{ǀ}\text{a}$ ‘husband’. Comparison (b) is more phonetically similar in respect to consonantism, but not vocalism; comparison (a) is cumulatively better both phonetically and semantically, but would still need to be confirmed by further examples. In the absence of contradictory cases, we may still count it as a tentative etymological match.

65. RAIN (-)

- PJ: ***!ǎ** (Ju. *ǎ*, Kx. *ǎ*, Kg. *ǎ* ~ *ǎa*, Gr. *ǎa*, OK. *ǎ* ~ *ǎa* ~ *ga*, Ek. *ǎ*). ◇ Preserved in all daughter dialects. For future purposes (such as tracing various morphophonological processes in the history of Khoisan linguistic lineages), it is perhaps worth noticing the similarity with ***!ú** ‘water’ (see below), although the two roots were clearly distinct even on the PJ level.
- †Hoan: *ǎǎ* (C, G).
- Ju-†Hoan: No lexicostatistical or etymological parallels.

66. RED (+)

- PJ: ***!ǎe** ~ ***!ǎ** (Ju. *ǎe*, Kx. *ǎe*, Kg. *ǎe* ~ *ǎa*, Gr. *ǎe*, OK. *ǎe*, Ek. *ǎe*). ◇ Preserved in all daughter dialects. However, while the Northern dialect cluster points to PJ ***!ǎe**, the rest of the dialects rather agree on PJ ***!ǎ**. This may be interpreted either as a rare, non-trivial combination of features (e.g. some special nasalized diphthong), or, more likely, as two morphological variants, indicating that the original root was simply ***!ǎ-** and that it became fused with two different suffixal extensions (***!ǎ-e** vs. ***!ǎ-N**).
- †Hoan: *ǎǎ* (C, G).
- Ju-†Hoan: A perfect lexicostatistical and etymological match, especially if we interpret internal Ju evidence as reflecting original ***!ǎ-**. For the correspondence between PJ voiced and †Hoan voiceless effluxes, see corr. #38a.

67. ROAD (?)

- PJ: ***!ǎ** (Ju. *ǎ* ‘path’, Kx. *ǎ*, Kg. *ǎa*, OK. *ǎ*, Ek. *ǎ*). ◇ This is almost certainly the original PJ root denoting the default means of getting from one place to another (Dorothea Bleek seems to have mistranscribed a dental click for both Kx. and OK. instead of palatal articulation — a rather common error in her records). For modern languages, dictionaries occasionally observe lexicalized oppositions between the older ‘path’ and the more recent ‘road’ (= ‘enlarged path between settlements’), cf. Ju. *ǎǎ* (no etymology); Kg. *kú*, Ek. *kùhù* (originally, perhaps, = ‘footprint’, as this meaning is also attested for Kg.). Such oppositions are likely to reflect quite recent developments.
- †Hoan: *ǎǎ* (C, G) ← ***!dao**. ◇ The more archaic form *ǎǎ* is still preserved in the Sasi dialect.
- Ju-†Hoan: †Hoan **dao* ‘path, road’ is phonetically identical with !Xóǎ *dào* and, further still, with Proto-Khoe **dao* ‘road’; this is an areal word, ultimately of Khoe origin, that has replaced the original †Hoan term for this Swadesh meaning and must be excluded from lexicostatistical comparison.

68. ROOT (-)

- PJ: (?) ***!ǎni** (Ju. *ǎni*, Kx. *ǎni*). ◇ This Swadesh meaning is unstable in Ju, and semantic reconstruction is hindered by inadequacy of existing semantic descriptions. An alternate candidate for PJ status is the root ***!ǎri**, cf. Ek. *ǎri* ‘root; handle’, OK. *ǎri* ~ *ǎri* ‘branch (?); root fibre’, Kx. *ǎri* ‘root fibre’, Kg. *ǎri* ‘root fibre’, Ju. *ǎri* ‘root’ (copied by Dickens from the earlier dictionary of J. Snyman); it is, however, seen here that most of the old sources have it in the specific meaning ‘root fibre’, so it is unclear if it should be properly eligible for straightforward semantic comparison. Cf. also Kg. *ǎbe* ‘root’, of unclear origin, perhaps = Ju. *ǎbe* ‘species of shrub’. We very tentatively go along with Dickens' data on Ju., setting up ***!ǎni** (or ***!ǎri**, since diagnostic parallels in Northern dialects are lacking) as the potential protoform.

- ǀHoan: **!qʔai** (C). ◊ Since the form is not attested in texts, it is not clear whether this is truly the generic term for ‘root’ in the language. Cf. other attested terms meaning ‘a kind of root’, e. g. *!one*, *ǁgama* etc.
- Ju-ǀHoan: No lexicostatistical or etymological parallels.

69. ROUND (?)

- This meaning is almost never attested in any of the available lexicographic sources on either Ju or ǀHoan languages; the very concept of ‘roundness’ of an object is seemingly not inherent to these lineages, aside from a few very recent borrowings.

70. SAND (?)

- PJ: ***kxà**. ◊ Same word as ‘earth’ q.v.
- ǀHoan: Not attested. Possibly also same word as ‘earth’.

71. SAY (±)

- PJ: ***kò** (Ju. *kò*, OK. *ka* ~ *ke*, Ek. *kòè* ~ *kwèé* ~ *kūyā*). ◊ Reconstructed based on the isogloss between OK. and Ek.; original root vocalism is not quite certain due to elements of suffixation in daughter dialects (cf. *ka* ~ *ke* in OK.). Another technically possible candidate is recorded in older sources: Kx. *o=kxwi*, Kg. *o=kxwi* ~ *o=kxwì*, Gr. *o=kxwí*. It is a transparent compound from PJ **o* ‘to do, make’ + PJ **kxúí*, and both for Ju. and Ek. it is translated as ‘speak, talk (about smth.)’ rather than ‘say (smth. specific)’. Furthermore, Dickens even assigns the morpheme *kxúí* a pronominal rather than verbal meaning: ‘be thus, be so (e. g. of the sound, sight or way of doing something)’, although a more detailed analysis of contexts is needed to clarify the situation.
- ǀHoan: **kǀĩ** (C, G).
- Ju-ǀHoan: PJ **kò* and ǀHoan *kǀĩ* constitute formal consonantal class matches, but root vowel correspondences are highly irregular. Despite this, we may count the pair as a «weak» etymological match, due to the semi-auxiliary nature of the word and, consequently, the possibility of undetected suffixes or enclitics to influence its vocalism in either of the two compared taxa.

72. SEE (+)

- PJ: ***séŋ** ~ ***héŋ** (Kx. *se*: ~ *se:*, Kg. *s:ŋ*, Gr. *sŋ*, OK. *siŋ* ~ *síŋ* ~ *süŋ* ~ *hŋ*, Ek. *híŋ* ~ *hí* ~ *šíŋ*). ◊ For Ju. proper, Dickens translates the cognate form *sé* as ‘to look (at), look after, investigate’, while reserving the meaning ‘to see’ for Ju. *hò* ← PJ **hò(ò)* ‘to find’ (cf. Ek. *hò-hò* id.). Phonological reconstruction is problematic. Ju. and Kx. drop the nasal part of the coda in this root just as they do for the verb ‘to drink’ q.v., for not quite clear reasons; nevertheless, most of the dialects, including additional data from Snyman's general survey, confirm original **-ŋ*. Vocalism is tentatively reconstructed as **-e-* based on Ju. data (in most dialects the two codas, *-iŋ* and *-eŋ*, seem to have merged). Initial **s-* is occasionally found lenited to *h-*; this either means a unique positional development before a syllabic nasal, or reflects an original aspirated **sʰ-* (not enough data to reach a definite conclusion).
- ǀHoan: **cí** (C, G).
- Ju-ǀHoan: A solid lexicostatistical and etymological match; correspondences are regular (for PJ **s* : ǀHoan **c*, see ‘hear’; deletion of velar nasal coda in ǀHoan is all-pervasive, see corr. #16).

73. SEED (?)

- PJ: (?) *!ó (OK. !ó, Ek. !ó). ◊ Outside the Northern branch, this word is elicited as Ju. !ó ‘pip’, confirming PJ reconstructibility with a highly natural semantic narrowing in Ju. On the other hand, Ju. ||àʔá ‘seed, kernel, marrow’, with the former meaning more precisely expressed by the compound ||xàrà-||àʔá ‘plant seed’, corresponds to Ek. ||àʔà ‘bone marrow’, suggesting a more general/abstract semantics of ‘pith, core substance’ for the protolevel. In older sources, the meaning is attested quite poorly.
- †Hoan: (?) !uru: (T). ◊ Attested only in Traill 1973. Unreliable.
- Ju-†Hoan: It is preferable to exclude this word from comparison, since PJ reconstruction is not particularly secure, and neither is the †Hoan entry. The Swadesh meaning ‘seed’ (as a general term) is quite unstable in Ju-Taa languages on the whole.

74. SIT (+)

- PJ: *ʔǃǃ (Ju. ǃǃ, Kx. ǃǃ, Kg. ǃǃ ~ ǃǃ ~ ǃǃ-a, Gr. ǃǃ, OK. |ǃ ~ |ǃ, Ek. ǃǃ). ◊ Preserved in all daughter dialects. Northern forms indicate an original preglottalized nasal click (cf. also the variants ǃǃ /Western/, ǃǃ /Eastern/ in Heikkinen 1986: 23). Coda is probably the same as in ‘blood’ q.v. Singular action form; the corresponding plural stem is PJ *ǃǃ (Ju. ǃǃ, Ek. ǃǃ, Kx. ǃǃ, OK. !o:, etc.).
- †Hoan: ǃǃ (C, G). ◊ Suppletive plural action form: kí=ǃǃ (C, G).
- Ju-†Hoan: A perfect lexicostatistical match; note the corresponding preglottalized nasal clicks. The plural action stems, however, do not correspond to each other and cannot be etymologized on a mutual basis. ◊ HH: 21, 28.

75. SKIN (-)

- PJ: *ǃǃ (Ju. ǃǃ, Kx. ǃǃ, Kg. ǃǃ ~ ǃǃ ~ ǃǃwa ~ ǃǃä, Gr. ǃǃ: ~ ǃǃwà, OK. ǃǃ ~ ǃǃ, Ek. ǃǃ). ◊ Preserved in all daughter dialects. Correspondences are regular. Some old sources seem to reflect an additional suffixal variant *ǃǃ-a (or ǃǃ-ba?) that is not confirmed in more recently transcribed material.
- †Hoan: ǃǃ (C, G).
- Ju-†Hoan: No lexicostatistical or etymological parallels. Of note, perhaps, is the complete segmental correlation between †Hoan ǃǃ ‘skin’ and PJ *ǃǃ ‘house’, but since traditional San houses are made of branches and reeds rather than animal skins, the connection is highly dubious on semantic grounds.

76. SLEEP (+)

- PJ: *cǃǃ (Ju. cǃǃ, Kx. cá ~ cǃǃ, Kg. cǃ ~ cá, Gr. cǃǃ, OK. cǃǃ ~ ca, Ek. ǃǃ). ◊ Correspondences are regular and trivial, including the development *cǃ- → ǃ- in Ekoka.
- †Hoan: cǃǃ (C, G). ◊ HH: 21, 23.
- Ju-†Hoan: A phonetically perfect lexicostatistical match. See corr. #18 for lack of nasalization in †Hoan.

77. SMALL (≈)

- PJ: *cǃe ~ *cǃe-mà (Ju. cǃe/-mà, Kx. ce:-ma ~ ce-ma, Kg. cé-ma ~ cé-e-ma, Gr. cǃe:-ma, OK. ce:-ma, Ek. ǃǃe-mà). ◊ In Ju., the simple form cǃe is used after nouns with diminutive suffixes (e.g. ǃǃu-ma cǃe ‘small house’); the compound form cǃe-mà (where -mà itself is a diminutive suffix) is used more frequently. Considering the data from the rest of the dialects, this situation is reconstructible for the PJ level, i.e. the PJ root *cǃe must have been most frequently used with the diminutive suffix *-mà. There is also no solid evidence for *mà

having ever functioned as an independent adjective ‘small’ on the same chronological level (some old sources list *ma* ‘small’ as a separate word, but textual examples always show it as a diminutive suffix, appended to nominal roots).

- ǀHoan: **|xʷǀi** (C, G). ◊ Glossed as ‘narrow’ in some sources, but cf. *těma-si |xʷǀi* ‘the dog is small’, etc. (hardly ‘narrow’). Of special note is the common diminutive suffix *-si* (C, G), as well as *zina* (Sasi *dàna*) ‘small /child/’ (used primarily in the submeaning ‘young’, but also seemingly attested in auxiliary functions, cf. Sasi *dana-si* ‘a little’).
- Ju-ǀHoan: No direct lexicostatistical matches. However, ǀHoan *|xʷǀi* may be tentatively compared with Ju. *|ùʔí*, Kg. *|wí* ‘thin (e.g. of paper)’, even though the click efflux correspondences are irregular; it is possible to suggest some rare type of dissimilation (e.g. **|xʷuʔi* → *|uʔi* in PJ) to get past this obstacle. Likewise, it is not excluded that PJ **cʷe* is the same morpheme as the diminutive suffix *-si* in ǀHoan, but this also requires explaining an irregular correspondence (lenition due to the shifted clitical/suffixal status of the morpheme?). At least one of these etymologizations has a good chance of being correct, so we accept the situation as reflecting a «partial» match.

78. SMOKE (≈)

- PJ: ***šórè ~ *šórà** (Ju. *šòrà*, Kx. *šore ~ šori*, Kg. *šórre ~ šure*, Gr. *šorǀi*, Ek. *šúlè*). ◊ Preserved in most daughter dialects; only for OK. Bleek lists the form *|onu*, of unclear origin (the original root is still preserved as *čule ~ čuli* ‘tobacco, snuff’). Correspondences are largely trivial, except for the unclear vocalic variation in the second syllable (cf. also Ju. *šòrò* ‘tobacco’, with yet another variant).
- ǀHoan: **zǃóeʰ** (H&H) ← ***dǃóeʰ**.
- Ju-ǀHoan: The ǀHoan form corresponds precisely to Ju. *dǃǃǃ* ‘to smoke out (bees), to make someone inhale smoke for medicinal purposes, etc.’. Since the overall semantics of the Ju. word may be generalized as ‘to make use of smoke’, zero-derivation of this verb from an original noun ‘smoke’ seems far more likely than the opposite scenario, in which case PJ **šórǂ* should be understood as an innovation (could ‘tobacco’ actually be the original meaning here?). ◊ HH: 14, 22.

79. STAND (≈)

- PJ: ***|ʷǀǀ** (Ju. *ǀǀ*, Kx. *ǀǀ*, Kg. *ǀǀ ~ ǀǀ ~ ǀǀǀ-ǀ*, OK. *ǀǀǀ ~ ǀǀ*, Ek. *ǀǀ ~ ǀǀ*). ◊ Preserved in all daughter dialects. The Ek. form is listed as *ǀǀ* (Western dialect) vs. *ǀǀ* (Eastern dialect) in Heikkinen 1986: 25, conflicting with König & Heine's transcription of a simple nasalized click and impeding a precise reconstruction. Note that this is the singular subject action verb; the corresponding suppletive plural stem is PJ **ǀǀǀ* (Ju. *ǀǀǀ*, Ek. *ǀǀǀ*, etc.).
- ǀHoan: **ǀǀǀ** (C, G). Suppletive plural action form: *ǀǀǀ* (ibid.).
- Ju-ǀHoan: This is a rare situation where a precise etymological match may be set up for the plural action stem (PJ **ǀǀǀ* = ǀHoan *ǀǀǀ*), but not for the singular one: despite a certain degree of phonetic similarity, discrepancies between click effluxes and codas remain unexplainable (the vowel at least could be explained away as extra suffixation, but the total lack of nasality in the ǀHoan form is a grave problem that prevents common etymologization of both items). ◊ HH: 21, 27 (plural action stem).

80. STAR (+)

- PJ: ***ǀǀ** (Ju. *ǀǀʰ*, Kx. *ǀǃǃ*, Kg. *ǀǃ ~ ǀǀ*, Gr. *ǀǀǀ*, OK. *ǀǀ ~ ǀǃ ~ ǀǀ*, Ek. *ǀǀ*). ◊ Preserved in all daughter dialects.
- ǀHoan: **ǀǃ** (C, G). ◊ Recorded as *ǀǀ* for the Sasi dialect.

- ǀHoan: **θxūī** (C, G).
- Ju-ǀHoan: PJ and ǀHoan forms perfectly match each other in everything (even tone!) except for the most important segment — the regular correspondence for ǀHoan *θ-* in Ju is *ʃ-* rather than *ʔ-*. Strictly speaking, this should invalidate the comparison (it is not found, for instance, in Heine & Honken's list of comparanda), but since the discrepancy concerns a rare type of click phonation that is absent in Ju languages altogether, it is not 100% certain that PJ **ʃ-* : ǀHoan *θ-* exhausts all possible types of correspondences before a complete list of parallels, based on a representative ǀHoan dictionary, is presented. For now, it cannot be ruled out that ǀHoan *θ-* is an innovation rather than an archaism (for instance, caused by labialization of the click efflux in certain contexts before labial vowels), which means that, in theory, *θ-* could correspond to more than one click type in Ju. Taking this into consideration, we may define this pairing as a potential match¹⁹.

85. THAT (+)

- PJ: ***toʔà** ~ ***ndoʔà** (Ju. *tòʔà*, Kg. *doá*, OK. *doa*, Ek. *ndùʔà* ~ *ndòʔà* ~ *tùʔà* ~ *tòʔà*). ◇ Since word-initial nasal clusters are generally prohibited in Ju, the variation *t-* ~ *d-* ~ *nd-* must probably result from morphemic contraction: it is reasonable to assume that **to-* represents the original root, while *ndo-* ~ *do-* are variants with an additional preposed deictic morpheme (← **NV-to-* ~ **VN-to?*). Final *-à* is a general relative morpheme.
- ǀHoan: **çǃǃa** (C, G). ◇ The more archaic variant *tǃǃa* is preserved in the Sasi dialect (Collins & Gruber 2014: 40).
- Ju-ǀHoan: A perfect lexicostatistical match. ǀHoan shows no signs of voicing or nasalization as seen in Ju dialects, indirectly confirming that these variants are secondary.

86. THIS (±)

- PJ: (A) ***e** (Ju. =è, Kg. *e* ~ *éya*, Ek. *ē*); (B) ***ŋ** (OK. *ŋ*, Ek. *ŋŋ*). ◇ Ekoka is the only Ju dialect in which both of these simple morphemes are attested: according to König & Heine, *ŋŋ* “refers to objects close to the speaker or deictic centre” as well as *ē*, but *ē* “has a contrastive function (‘this, rather than any other one’)” (König & Heine 2001: 64–65). Even if this opposition is not confirmed beyond the Northern cluster, both pronominal forms look sufficiently archaic to suggest that it may have been inherited from PJ, with South-Central dialects simplifying it in favor of **e* (at least in Ju.: the situation with the other dialects remains insufficiently well described).
- ǀHoan: **hā** (C, G).
- Ju-ǀHoan: Despite some phonetic similarity (mostly in terms of root structure), it is not easy to trace PJ **e* and ǀHoan *hā* back to a single protoform. However, the vocalic correspondence is not unique (corr. #6), and ǀHoan *h-* may theoretically be equated with the Ju class prefix *h=* that typically precedes the pronominal morpheme (*h=è* ‘this’ for classes 1–4, opposed to *k=è* ‘this’ for class 5).

¹⁹ In a recent presentation, Sands (2018) mentions this parallel together with an additional possible example (Ju *ǀǃǃ* ‘burp’ : ǀHoan *ǃǃu* ‘heartburn’) as possible evidence for a special series of labio-velar clicks in Proto-Ju-ǀHoan. This actually echoes an earlier idea suggested in Starostin 2008: 358 («...some old influencing factor, for instance, a particular type of labial articulation after the click (either the click itself or the following vowel could be strongly labialized)»). However, due to the relative scarceness of evidence and lack of local typological support for separate labiovelarized clicks the suggestion remains somewhat speculative for now.

87. THOU (-)

- PJ: ***a** (Ju. *à*, Kx. *a-hi*, Kg. *a ~ á*, Gr. *à*, OK. *a ~ a-hi*, Ek. *à*). ◇ Preserved in all dialects, being encountered either as a simple monophonemic variant or in the emphatic variant **a-hŋ* ~ **a-hi*. Curiously, in a few dialects an additional variant with an extra labial phoneme is attested: (a) for Kg., Bleek lists a special subject form *m?a*, distinguishing it from the more common object and possessive form *a*; (b) for Ek., König & Heine list a special subject form *bà*, especially in sentence-internal position. These phenomena most likely have a common origin, but the exact provenance of this labial prefix and its shape in PJ remain to be clarified.
- †Hoan: **ù** (C, G). ◇ In the Sasi dialect, there is also an additional “in-focus” form *bù* ~ *bù:* (Collins & Gruber 2014: 77).
- Ju-†Hoan: Unlike the 1st p. sg. pronoun, forms for the 2nd p. sg. pronoun in Ju and †Hoan cannot be reconciled with each other. The situation could make sense from a more comprehensive perspective that also includes !Ui-Taa (South Khoisan) languages as part of the same family: considering that the system there is reconstructible as **a* ‘thou’ (sg.) vs. **u* ‘you’ (pl.), it is likely that Ju has preserved the original singular form, whereas †Hoan may have replaced it with the original plural. However, this solution formally lies beyond the scope of this binary analysis. It is also curious to note the similarity of the *b-* “in-focus” Sasi prefix to the *m- ~ b-* subject prefix in certain Ju dialects, even if the prefixes in question are joined to different root morphemes.

88. TONGUE (±)

- PJ: ***Thari** (Ju. *dhàri*, Kx. *tari*, Kg. *térri*, Gr. *ṅtháti*, OK. *tali*, Ek. *dhàli*). ◇ Preserved in all daughter dialects. Correspondences between the initial consonant show irregular fluctuation of laryngeal features, including even a completely unexpected and very rare case of prenasalization in Gr., as recorded by Doke. This is consistent with the typologically aberrant (both for Africa and other world areas) phonetic behavior of the word ‘tongue’, reflecting an odd phonosemantic phenomenon that is difficult to explain in historical terms.
- †Hoan: **cèlā** (C, G), **cìrà**: (SH). ◇ Recorded as *càla* in the Sasi dialect.
- Ju-†Hoan: Judgement on whether PJ **Thari* and †Hoan *cèlā* are cognate or not has to be postponed. On one hand, the correspondences are notably irregular, since †Hoan *c-* (rather than *ç-* ← **t-*) is always found in roots where PJ has affricates or sibilants (see ‘hear’, ‘see’ on this list); vocalic patterns do not present a clear match, either. On the other hand, since the word-initial consonant or cluster in PJ remains altogether unclear, and since the word ‘tongue’ tends to behave irregularly in Khoisan languages on the whole, unique historical developments in this case seem highly likely; probability of cognacy is weak, but should not be ruled out.

89. TOOTH (+)

- PJ: ***c̣au** (Ju. *c̣àù*, Kx. *c̣rou ~ cou ~ cau*, Kg. *c̣:au ~ caù*, Gr. *c̣áú:*, OK. *cau*, Ek. *č̣āō*). ◇ Preserved in all daughter dialects.
- †Hoan: **c̣íú** (C, G), **c̣íù** (SH). ◇ Plural form: *c̣ěō* (G), *c̣àō-qà* (SH). Recorded as *c̣áú* in the Sasi dialect.
- Ju-†Hoan: A perfect lexicostatistical match with trivial correspondences. ◇ HH: 17, 23.

90. TREE (-)

- PJ: ***ʔḥãŋi** (Ju. *!àṅṅ*, Kx. *!āi*, Kg. *!áŋ ~ !áŋ-a ~ !āũ ~ !aŋ*, Gr. *!ãŋ*, OK. *!áú ~ !ā ~ gáú*, Ek. *!àhŋ ~ !ḥãŋ*). ◇ Preserved in all daughter dialects. However, phonetic correspondences here

are complex and, in some aspects, unique. The basic structure of the word is more or less the same as in ‘neck’ q.v., which is reflected in the reconstruction of the coda **-aŋi* for both items. Seemingly random fluctuations are, however, observed in click efflux articulation (ranging from simple velar release to prevoicing to aspiration), additional vowel properties (breathy articulation in Ju.) and in the tonal scheme. The provisional reconstruction with **^h-* and ultra-low tone on the first mora merely reflects the fact that some particularly complex bag of features must have been present on the proto-level in order to yield such a large variety of reflexes.

- ǀHoan: *ǀǀó* (C, G). ◊ Recorded as *ǀú* in the Sasi dialect.
- Ju-ǀHoan: It is worth noting that ǀHoan *ǀó* (Sasi *ǀú*) is a perfect phonetic match for Ju. *ǀú* ‘hunting bow’; semantically, such a link is possible, since traditional Bushman bows were «as a rule prepared from the wood of the *Grewia flava*» (Schapera 1930: 128), but requires setting up a chain of semantic shifts that is hard to accept without additional evidence. Ju **^hǀǀi* finds no cognates in ǀHoan.

91. TWO (-)

- PJ: **cā* ~ **cā̃* ~ **cā̃̃* (Ju. *cā̃* ~ *cā̃̃̃*, Kx. *ca* ~ *ča*, Kg. *sā̃* ~ *sā̃̃̃* ~ *s:a* ~ *ca* ~ *cā̃* ~ *ca:̃* ~ *zā̃̃̃*, Gr. *sā̃*, OK. *cá* ~ *ca* ~ *ča*, Ek. *čā̃*). ◊ Fluctuation between different types of vocalic features remains unexplained (different variants are sometimes attested within the same well-described dialect, e.g. Ju.).
- ǀHoan: *ǀǀā* (C, G).
- Ju-ǀHoan: Unless the PJ entry can be shown to represent a rare case of click affricativization (**ǀā̃̃̃* → **cā̃̃̃*), which is not altogether excluded but requires far more confirming evidence, PJ and ǀHoan forms have to be kept apart from each other.

92. WALK (GO) (±)

- PJ: **ú* (Ju. *ú*, Kx. *ú*, Kg. *ù* ~ *ù:*, Gr. *iú* ~ *ǀú:*, OK. *ú*, Ek. *ú*). ◊ Preserved in all daughter dialects.
- ǀHoan: *ǀǀǎ* (C, G). ◊ Attested as *tǎ* in the Sasi dialect.
- Ju-ǀHoan: The ǀHoan verb **tǎ* may be tentatively analyzed as a fused formation from an early root **ta* (which is still in use as an auxiliary pre-verb, indicating motion with the purpose of completing an action) and an unspecified second component — which, incidentally, could be fairly well associated with PJ **ú* ‘to go’ (vocalic correspondences would be perfect except for unexplained pharyngealization). However, there are no definitive grounds for such a segmentation, and given the short monovocalic nature of the compared root, we can only accept this match as highly tentative (in any case, even if the suggested fusion were correct, it would constitute a near-complete lexical replacement in ǀHoan).

93. WARM (HOT) (-)

- PJ: **khúí* (Ju. *khúí*, Kx. *kwí* ~ *khwí*, Kg. *kwí* ~ *kwi* ~ *kwiya*, OK. *khwí*, Ek. *khúí*). ◊ Preserved in all daughter dialects. In all well documented sources the meaning is explicitly noted as ‘hot’ (antonymous to **ǀǎ̃̃̃* ‘cold’), and in PJ it seems to have been well distinguished from **ǀú* ‘warm’ (Ju., Ek. *ǀú*, etc.).
- ǀHoan: *kǀuru* (C, G). ◊ Always glossed as ‘hot’; the word is probably distinct from ‘warm’, for which cf. *ǀo:* ~ *ǀuo* ‘warm’ in Traill 1973: 32, possibly = *ǀqo* ‘warmth’ (C).
- Ju-ǀHoan: The words for ‘hot’ in PJ and ǀHoan, despite some phonetic similarity, are probably not related (initial consonants do not perfectly correspond to each other, and

fossilized morphology has to be assumed for both groups to justify the connection). The words for ‘warm’, on the other hand, are most likely cognates, but better data are needed for †Hoan to ascertain the semantics.

94. WATER (+)

- PJ: ***!ú** (Ju. *ǀú*, Kx. *ǀú* ~ *ǀu* ~ *ǀú*, Kg. *ǀù* ~ *ǀú*, Gr. *ǀǀú*, OK. *ǀú* ~ *ǀó*, Ek. *ǀú*). ◊ Preserved in all daughter dialects. Retroflex click articulation is seen in the Grootfontein dialect and is unambiguously reconstructible for the proto-level.
- †Hoan: **ǀǀǀ** (C, G).
- Ju-†Hoan: The correspondence between a retroflex click in PJ and an affricate in †Hoan (in this case, both phonemes even share the same characteristics of +voiced) is essentially the same as in the word for ‘hand’ (see above). Although the phonological and phonetic implications of this correspondence remain unclear, observing it specifically in two highly stable elements of the Swadesh wordlist makes coincidence highly unlikely. We count this as a direct lexicostatistical match.

95. WE (±)

- PJ: [exclusive] ***è** (Ju. *è*, Kx. *e*, Kg. *e*, OK. *e* ~ *e-hŋ*, Ek. *è*); [inclusive] ***m̄** (Ju. *m̄*, Kg. *hm*, Ek. *m̄-hm̄*). ◊ The basic opposition between exclusive and inclusive forms of the 1st p. pl. pronoun is observed in the majority of Ju dialects. For Ekoka, it is noted that *è* is now perceived as an archaic form, with speakers generally preferring the innovation *ǀǀù* (= ‘people’, see ‘person’ above). Both pronouns also have expanded (emphatic?) variants, well attested in Ju. (*è-!á*, *m̄-!á*) and several other dialects. Special dual forms are more rare and transparently recent (Ju. *è-cá*, *m̄-cá* ‘the two of us’, compounded with ‘two’ q.v.).
- †Hoan: [exclusive] **n-!āǀē** (C, G); [inclusive] **qǀā** (C, G). ◊ The variant of the inclusive pronoun in the Sasi dialect is slightly shorter: *qǀ* (Collins & Gruber 2014: 77).
- Ju-†Hoan: There are no clear-cut isomorphisms between the 1st p. pl. sub-systems in PJ and †Hoan. Surprisingly, the closest morphemes are the suffixal extensions — PJ **-!a* and †Hoan *-!āǀē*, which cannot be easily traced back to any recent grammaticalization patterns (but are probably further related to the productive diminutive plural ending *-!āǀā* in †Hoan). It is likewise reasonable to suggest a link between †Hoan *qǀā* and the regular plural ending *-qǀ* in the same language, but the exact nature of it is a matter of guesswork (one possible scenario, for instance, is that the original pronominal root was contracted/deleted before the ending, i.e. **n-qa* → *qa*).

It is also worth noting that a special dual variant of the pronoun is attested in †Hoan, where the nasal monophonemic root is represented by a labial allophone: *m-θōā* ‘us two’ (Collins & Gruber 2014: 71). This can be easily ascribed to assimilative influence of the following labial click (← **n-θōā*); on the other hand, the opposite scenario cannot be excluded, either, i.e. delabialization in front of a tightly adjacent nonlabial click: **m-!āǀē* → **n-!āǀē*. Due to the uniqueness of this phonotactic environment, no examples are available to confirm or disprove such a development, meaning that it is possible to set up a tentative weak match between the exclusive pronoun in †Hoan and the inclusive pronoun in Ju (the disagreement in clusivity should not be a problem, since †Hoan must have remodeled the old opposition anyway).

96. WHAT (+)

- PJ: ***hǀ-čí** (Ju. *hǀ-čé*) / (?) ***m-** (Ek. *m̄-*). ◊ Interrogatives in Ju are usually complex, consisting of a general interrogative marker, a nominal root, and (optionally) a final particle:

thus, Ju. *hã-čé* = *ha* (question marker) + *či* ‘thing’ + *-e* (final particle). There is significant variation between dialects concerning the selection of the components: thus, Ek. and some other Northern dialects show *m* instead of *ha*, cf. OK. *m-pai* (D. Bleek’s data, second component is unique and unclear); Ek. *m-čá* ~ *m-čē* ~ *m-či*. Data from J. Snyman’s comparative survey of Ju subdialects, although restricted to ‘who?’ rather than ‘what?’, show that variants of interrogative pronouns with initial *ha-* have a much wider distribution than variants with *m-*, the latter being largely restricted to the area between the Cuito and Cuando rivers. It may be suspected that *m*-initial forms are secondary in the Northern dialects, having penetrated them under the influence of Khoe, where **mã* ~ **ma* ‘who?, which?’ is one of the main interrogative stems (Vossen 1997: 379). However, this evidence is inconclusive; it is likewise possible that **m-* is archaic, surviving on the Northern and Northwestern periphery of the Ju cluster.

- ǀHoan: **ǀǎʳi-yà** (C, G). ◊ Apparently, ǀHoã has chosen the rare strategy of neutralizing the lexical opposition between ‘who?’ and ‘what?’ in favor of ‘who?’ (see notes on ‘who?’ below). The morpheme *yà* is a general question particle. Cf. also *çini* ‘what?’ in Traill 1973: 32, not confirmed in any of the later sources. The Sasi equivalent for ‘what?’ is *ndā*, “a question word which does not exist in ǀHoã” (Collins & Gruber 2014: 192).
- Ju-ǀHoan: Any comparisons between Ju and ǀHoan interrogatives may be made only on the level of the general interrogative particle. Etymological identity of PJ **hà* and ǀHoan *yà* seems quite likely in light of the completely identical correspondence between PJ **ha* ‘3rd p. sg.’ and ǀHoan *ya* id. (although the deictic/personal and interrogative morphemes themselves are probably just homonyms). The nominal extensions of the pronouns are different, due to the lexical renewals of the words for ‘thing’ and ‘person’ in one or both subgroups after the split. Nevertheless, since the main interrogative meaning is carried by the *ha/ya* morpheme, we count both ‘what?’ and ‘who?’ as lexicostatistical matches between PJ and ǀHoan.

97. WHITE (≈)

- PJ: ***!àʔū** (Ju. *!àʔú*, Kx. *!áú*, Kg. *!áo* ~ *!káo* ~ *!kxáo*, Gr. *!əʔàù*, OK. *!áú*, Ek. *!àʔö*). ◊ The structure of the stem, including a glottal stop between the two vowels, is very well confirmed by most modern sources (Ju., Ek.) as well as Doke’s Grootfontein data. It also explains the variation observed between the forms recorded for Kg., where the glottal stop may have been incorrectly interpreted as part of the click efflux. The item is well attested in the majority of dialects and safely reconstructible for the PJ level.
- ǀHoan: **ǀǎʔnna** (H&H). ◊ Listed as *ǀxaʳna* ~ *ǀaʳana* in Traill 1973: 32.
- **Ju-ǀHoan**: No lexicostatistical matches. PJ ***!àʔū** ‘white’ is most likely somehow linked to Proto-Khoe ***!ʔú** ‘white’ (Vossen 1997: 506), well represented in Kalahari Khoe and also preserved in Nama with an additional suffix (*!u-ri*). Borrowing from Proto-Khoe into PJ does not seem likely, since it is unclear why an original **!ʔu* should have yielded a more complicated vocalic structure in PJ; more probable is the reverse situation (PJ ***!àʔū** → Proto-Khoe ***!ʔú** with simplification of an unusual structure), or even the scenario according to which both forms are retained from a common ancestor of PJ and Proto-Khoe (in which case, of course, the PJ equivalent should automatically be projected onto the Proto-Ju-ǀHoan stage as well).

As for ǀHoan **ǀǎʔnna**, this is a rare case of a **CVna* structure for a ǀHoan adjective; provided that *-na* is historically of suffixal origin, a possible parallel may be seen in Ju. **ǀǎʔàbè** ‘shiny’, further relatable to *!Xóõ* (Taa) **ǀǎʔba** id. Our understanding of the productivity aspects of early Peripheral Khoisan derivational morphology is insufficient to

assess the plausibility of two different derivational suffixes for the same root, but since there is ample evidence to back up the very presence of such derivational patterns at those stages, the etymological match between **ʔaʔ-na* and **ʔaʔ-be* is acceptable. (Note that it does not necessarily invalidate the comparison of the Ju word with †Hoan ‘moon’, discussed above, since both forms can ultimately go back to the same root with different suffixes).

98. WHO (+)

- PJ: **hà-ʒù* (Ju. *hà-ʒòè*, Kg. *a-ʒu*) / (?) **m-* (Ek. *m-ʒē ~ m-ʒōē*). ◊ In all Ju dialects, the animate interrogative pronoun is formed from the general interrogative morpheme + **ʒù* ‘person’ q.v. (sometimes also further extended with the deictic stem **-e*: **hà-ʒù-e* → Ju. *ha-ʒoe*). For discussion of the interrogative morpheme, see ‘what’ above.
- †Hoan: *ʔá-rī-yà* (C, G). ◊ In the Sasi dialect, the phonetic shape is *ʔá-lī-yà*. A compound form, consisting of *ʔá-rī* ‘man’ q.v. and the general interrogative particle *yà*. See ‘what?’ for further notes.
- Ju-†Hoan: A lexicostatistical match on the level of the general interrogative particle. See ‘what?’ for a more detailed commentary.

99. WOMAN (≈)

- PJ: **ʒ^hau* (Ju. *ʒ^hàú*, Kx. *ʒau ~ ʒou ~ ʒóú*, Kg. *ʒau ~ ʒáú ~ ʒáú*, Gr. *zsháú*., OK. *cau ~ cáú*, Ek. *ʒ^hāō*). ◊ Preserved in all daughter dialects. Correspondences are generally regular and trivial; PJ **ʒ^h-* automatically becomes preglottalized in Ju., so there is no need to carry it over onto the proto-level as a phonological feature. In most modern dialects the word unambiguously denotes a female human being, and should be distinguished as such from PJ **de ~ di* ‘female (in general, incl. animals etc.)’.
- †Hoan: *ʔá-rī=||ái^f* (C, G). ◊ A compound form; the first part is *ʔá-rī* ‘man’ q.v., while the second part, when in independent usage, means ‘female’ (listed as *||a^fai ~ ||ai^f* in Traill 1973: 32). Suppletive plural form: *||q^hà* ‘women’ (Collins & Gruber 2014: 21). Distinct from *ʒú* ‘wife’ (Collins & Gruber 2014: 92).
- Ju-†Hoan: PJ **ʒ^hau* ‘woman’ is a perfect etymological match for †Hoan *ʒú* ‘wife’ (HH: 17), and it is safe to assume that this term may have been polysemous in the ancestral language (‘woman /in general/’ = ‘married woman’). Another transparent cognate from the same semantic field is PJ **de ~ di* ‘female’ = †Hoan *ʒe* ‘mother’ (HH: 16). Nevertheless, †Hoan seems to have undergone lexical replacement, substituting the original term for ‘woman’ for a compound expression in which the ‘feminine’ part of the meaning is now denoted by the morpheme *||ái^f*, very likely diffused in †Hoan under Taa influence, cf. !Xóõ *||ái^f* ‘female’, also N|u|len *tu ||ai* ‘woman’, lit. ‘person-female’, i. e. the same model of compounding as in †Hoan. Therefore, we cannot qualify this situation as a lexicostatistical match.

100. YELLOW (-)

- PJ: **||āŋu* (Kx. *||āũ*, Kg. *||áŋ ~ ||āŋ*, Gr. *||a:ŋ ~ ||āŋ*, OK. *||āŋ*). ◊ Same word as ‘green’ q.v.; most of the old sources on Ju dialects indicate no lexical distinction between the basic ‘green’, ‘blue’, and ‘yellow’. In more modern and more detailed sources, we occasionally encounter separate entries for ‘yellow’, e.g. Ju *||ð^fnì-||ú*, lit. ‘/the color of/ the jewel beetle’s (*||ð^fnì*) belly (*||ú*)’, and Ek. *||úũ* = ‘egg’ q.v. Naturally, these have to be interpreted as recent (completely transparent) semantic innovations.

- ǀHoan: **zǀʔa** (T). ◇ Same word as ‘green’; attested only in Traill 1973, thus not highly reliable from the phonetic or semantic aspect.
- Ju-ǀHoan: No lexicostatistical or etymological matches.

101. FAR (-)

- PJ: ***ǀxǎ** (Ju. ǀxǎ, Kx. ǀxǎ: ~ ǀa ~ ǀa: ~ !ǎ: ~ !^hǎ: ~ |xǎ: ~ |ǎ, Kg. ǀxǎ ~ ǀxǎ ~ ǀa, OK. ǀxa ~ |xa, Ek. !ǀxǎ). ◇ The reconstruction is based on precise correspondences between Ju. and Ek. Old sources show a lot of fluctuation between the palatal and the dental (more rarely, the alveolar) click; this is not well understood (palatal clicks are frequently transcribed erroneously by L. Lloyd, D. Bleek and others, but the fluctuations look rather extreme in this particular case).
- ǀHoan: **ǀǎ** (C, G).
- Ju-ǀHoan: No lexicostatistical or etymological matches.

102. HEAVY (-)

- PJ: ***ǀĩ^h** (Ju. ǀĩ^h, Kx. ti, Kg. tí ~ tǐ, Ek. ǀĩhi). ◇ Preserved in all daughter dialects (where attested). Reconstructible for PJ with breathy vowel articulation and ultra-low tone.
- ǀHoan: **ǀqô** (C, G).
- Ju-ǀHoan: No lexicostatistical or etymological matches.

103. NEAR (+)

- PJ: ***toǀm** (Ju. tǀm, Kx. tǀm ~ toma ~ tum ~ dom, Kg. tǀumm ~ tumma ~ tamma, OK. tum, Ek. tǀm). ◇ Preserved in all daughter dialects. Some of the attested forms represent the complex «junctive» variant **toǀm-a* (supposedly followed by a complement).
- ǀHoan: **ǀǎ^fm** (C, G). ◇ Should go back to an earlier **ta^fm*.
- Ju-ǀHoan: Consonantal correspondences between PJ and ǀHoan are perfect; vocalism remains more complicated, but cf. HH: 18, where several additional examples of the same pattern (PJ **o* : ǀHoan *a*) are adduced. The authors provisionally interpret them as reflecting Ju-ǀHoan **ao*, without mentioning that all such instances occur exclusively before the labial nasal coda (e. g. ǀHoan ǀna^fm ‘springhare’ = PJ **ǀo^fm* id., ǀHoan ǀám ‘ripe, cooked’ = PJ **ǀom* id., etc.). The likeliest solution is that the coda simply influenced the original vocalism in PJ. Less clear is the correlation between glottalic articulation of the vowel in PJ vs. pharyngealization in ǀHoan, but this, too, is not unprecedented (cf. ‘cold’ above, or PJ **ǀǎ^fma* ‘to enter’ — ǀHoan *ǀa^fm* id.). Overall, while the phonological and phonetic details still deserve closer scrutiny, the etymology as a whole can be evaluated as highly reliable. ◇ HH: 22.

104. SALT (?)

- PJ: ***ǀgúí** (Ju. ǀgúí, Kg. ǀgwi, OK. ǀgwi, Ek. ǀgúí). ◇ This is the most widespread and the least etymologically suspicious equivalent for ‘salt’ in Ju. Another root, attested as Kx. *dabe*, Ju. *díbi* and also recorded by Snyman for several other Ju subdialects, is most likely of Khoe origin (cf. Proto-Khoe **dobe* ‘salt’ in Vossen 1997: 481).
- ǀHoan: **ǀqǎ^fnǎ** (C, G). ◇ Clearly the same word as !Xóǀ *qǎ^fna* ‘salt’.
- Ju-ǀHoan: No lexicostatistical or etymological matches. Judging by the situation in Ju, the meaning ‘salt’ is rather easily diffused across different Khoisan lineages, so it is highly likely that ǀHoan *ǀqǎ^fnǎ* is a borrowing from Taa (rather than both being inherited from Proto-Peripheral Khoisan).

105. SHORT (-)

- PJ: *!ò ~ *!ò-mà (Ju. !ò-mà, Kx. !ò:, Kg. !o-ma ~ !o:-ma, OK. !o-!o, Ek. !ò). ◇ Preserved in most dialects. The root can be used by itself or in conjunction with the diminutive suffix *-ma.
- †Hoan: †éū (C, G).
- Ju-†Hoan: †Hoan †éū is etymologically comparable with Ju. †àð^h ‘to lack, be short of’ (front vocalism in †Hoan is apparently caused by palatal influence of the click), but the Ju. form itself is not safely reconstructible for PJ. Additionally, there are still problems with phonetics (prosody) and semantics, so the etymology is not fully convincing.

106. SNAKE (≈)

- PJ: *†àgà ~ *†àwà (Ju. †àgà-mà ~ †à:-mà, Gr. †àwà, Ek. !àwà). ◇ The generic term for ‘snake’, reconstructible for PJ, has an atypical bisyllabic structure, since *-ga ~ *-wa is not one of the few common syllables allowed in coda position. Most likely, the form is originally a compound, although the phonetic and semantic properties of its source morphemes are unclear. There are several terms denoting specific types of snakes in PJ that are even more widespread and simpler in structure, e.g. *†ɰ ‘python’, *||kxãũ ‘blind snake’, *!e ‘puff-adder’; the common word for the entire suborder may be some descriptive term (of a euphemistic nature?). It is useful to note that some old sources occasionally quote words for specific types of snakes in the general meaning ‘snake’: e.g. Kx. !e: ~ !i ‘snake’ (= ‘puff-adder’), OK. !áú ~ !wẽ ‘snake’ (= ‘blind snake’). Naturally, it is impossible to correctly assess the semantic scope of these forms from existing data.
- †Hoan: !ái (C, G).
- Ju-†Hoan: The †Hoan term for ‘snake’ is a near-perfect correspondence for PJ *!ai ‘puff-adder’ (see ‘claw /nail’ on the possible correspondence between PJ *! and †Hoan !), but there are no parallels in †Hoan for PJ *†àgà — another indirect hint at the non-archaic nature of this compound.

107. THIN (-)

- PJ: *žã^fm (Ju. žã^fm, Kx. žã^fm, Kg. žã^fm, Gr. žãm). ◇ A common Ju morpheme. The Ek. equivalent is ||kxài, perfectly corresponding to Ju. ||kxài ‘wrinkled’ and probably semantically innovative.
- †Hoan: |xolo (C).
- Ju-†Hoan: No lexicostatistical or etymological parallels.

108. WIND (-)

- PJ: (?) *†a (Kx. †ã ~ †a:, Kg. †a ~ †ã). ◇ Although this root is quite widely distributed throughout the dialects (judging by J. Snyman's comparative data), it is oddly missing in both of the best described Ju varieties. In Ju., it has been replaced by mã^f, originally a verbal stem with the meaning ‘to blow /of wind/’ (cf. Ek. mã^f ‘to blow’, etc.). In some of the Northern dialects we see a different replacement: OK. ||uli, Ek. ||ðhlì ~ ||ðhlì-gō, likely cognate with Ju. ||òrò ‘whirlwind’, i.e. originally ‘strong wind’. On the other hand, the similarity between this root and Proto-Khoe *†ã ‘wind’ (Vossen 1997: 507) suggests an alternate scenario — namely, areal borrowing from Khoe sources along the same lines as ‘fish’ q.v. If so, PJ *ma^f could have very well been both a verbal (‘blow’) and nominal (‘wind’) root, with narrow specialization to verbal usage after the nominal functions were taken over by the Khoe borrowing. A more insightful evaluation of the probabilities will only be possible in the context of a general study on the scope and nature of Ju-Khoe areal contacts.

- ǀHoan: ǀqʷui (C). ◊ Similarity with !Xóǀ ǀqʰuē ‘wind’ is hardly accidental, but in this case, borrowing is not an immediately obvious explanation, since there are visible phonetic discrepancies (glottalized click efflux in ǀHoan vs. aspirated in !Xóǀ) that should not be characteristic of recent contact.
- Ju-ǀHoan: No lexicostatistical or etymological parallels (regardless of whether the comparison is made with PJ *ǀa or *maʰ).

109. WORM (?)

- PJ: Not properly reconstructible due to lack of attestation. Only the Ju. word is known: ǀũǀú.
- ǀHoan: ǀǃǃǃ: ~ ǀũǃǃ: ~ ǀǃǃu: (T). ◊ Not attested in any reliable sources.
- Ju-ǀHoan: Although there is too little information for an etymological or lexicostatistical decision, it is curious that the Ju. and ǀHoan forms are extremely similar to each other. However, the correspondence between a palatal click in Ju. (or PJ) and an alveolar click in ǀHoan would be highly irregular, unless Traill's phonological transcription is in error — but no other examples of such errors could be detected upon careful analysis of the data in Traill 1973.

110. YEAR (-)

- PJ: (?) *kuri (Ju. kúri, Kx. kuri, OK. kuri). ◊ Although the form is quite widely spread across Ju dialects, its projection onto the PJ level is highly dubious — like ‘fish’ and possibly ‘wind’ (see above), this is most likely a borrowing from Khoe *kúri ‘year’ (Vossen 1997: 454). There is, however, very limited data on alternate candidates. In Ek., the meaning ‘year’ is expressed by the same word as ‘rain’ (ǀǃ) — possibly an archaism, but explicitly limited to just one dialect. For the Grootfontein dialect, Doke records ǀǃǃ ‘year’, an isolated form with no parallels whatsoever.
- ǀHoan: kʰǃǃ (C, G).
- Ju-ǀHoan: No lexicostatistical or etymological parallels.

Data analysis

The table below summarizes all our findings, once again classifying all matches into «solid» (confirmed by recurrent correspondence patterns), «dubious» (containing no more than one strong violation of observed patterns), «etymological» (potential cognates are only attested with a semantic shift), and non-existent. The 10 additional items (101–110) are marked separately (e. g. «8+1» means that there are 8 matches in the main wordlist and 1 more among the ten additional items).

Match type	List half	Cases	Wordlist items
Solid	1st	21	‘blood’, ‘die’, ‘ear’, ‘eat’, ‘eye’, ‘hand’, ‘head’, ‘hear’, ‘horn’, ‘I’, ‘kill’, ‘louse’, ‘name’, ‘new’, ‘not’, ‘one’, ‘star’, ‘tooth’, ‘water’, ‘what’, ‘who’
Solid	2nd	8 + 1	‘all’, ‘belly’, ‘earth’, ‘red’, ‘see’, ‘sit’, ‘sleep’, ‘that’, ‘near’
Dubious	1st	6	‘claw /nail/’, ‘drink’, ‘mouth’, ‘tail’, ‘tongue’, ‘we’
Dubious	2nd	5	‘bite’, ‘cold’, ‘say’, ‘this’, ‘walk /go/’
Etymological	1st	5	‘foot’, ‘meat’, ‘moon’, ‘smoke’, ‘stone’
Etymological	2nd	9 + 1	‘big’, ‘come’, ‘knee’, ‘many’, ‘person’, ‘small’, ‘stand’, ‘white’, ‘woman’, ‘snake’
No matches	1st	18	‘ashes’, ‘bird’, ‘black’, ‘bone’, ‘dog’, ‘dry’, ‘egg’, ‘fire’, ‘hair’, ‘heart’, ‘leaf’, ‘night’, ‘nose’, ‘rain’, ‘sun’, ‘thou’, ‘tree’, ‘two’

Match type	List half	Cases	Wordlist items
No matches	2nd	18 + 6	'breast', 'fat', 'feather', 'fly', 'give', 'good', 'green', 'know', 'lie', 'liver', 'long', 'man', 'mountain', 'neck', 'root', 'skin', 'warm /hot/', 'yellow', 'far', 'heavy', 'short', 'thin', 'wind', 'year'
Excluded (lack of data)		8 + 1	'bark', 'burn', 'cloud', 'full', 'round', 'sand', 'seed', 'swim', 'worm'
Excluded (borrowings)		2 + 1	'fish', 'road', 'salt'

The following conclusions may be drawn from these statistics.

1. Percentage of lexicostatistical matches between Proto-Ju and #Hoan on the 100-item wordlist may vary from 32% (29/90, only counting the «solid» matches) to 44% (40/90, counting «solid» and «dubious» matches together).

Since the disintegration of Proto-Ju itself, based on lexicostatistical calculations between modern dialects, is tentatively dated to about 200 AD (Starostin 2013: 321), with the average Ju dialect replacing about 10–12% by the present day, this, according to Sergei Starostin's glottochronological method, yields a highly approximate figure of about 5000–5500 years of separation between modern Ju varieties and #Hoan in the worst case (all «dubious» matches discarded), or of about 4000–4500 years in the best case (all «dubious» matches included). The latter is an age roughly comparable with the most common glottochronological datings for such Eurasian families as, for instance, Fenno-Ugric (without Samoyed) or Kartvelian (together with the highly divergent Svan).

2. The number of direct solid lexicostatistical matches within the first («more stable») half of the Swadesh list vastly exceeds the number of such matches within the second half (21 against 8). This is significant evidence in favor of a genetic rather than areal connection between Ju and #Hoan, with the imminent underlying assumption of a common linguistic ancestor.

3. Conversely, the number of «etymological» matches is higher for the «less stable» part of the wordlist (9 against 5). This is an interesting observation that seems to agree with basic logic, since «less stable» lexical items should be expected to also be more prone to semantic change, in addition to outright elimination; however, it remains to be seen whether it may be generalized, since statistical data on this type of correlation has yet to be collected for representative samples.

4. There is currently no evidence that a majority, or even a significant portion, of lexical replacements that took place between Proto-Ju-#Hoan and Proto-Ju or modern #Hoan are due to massive borrowing from other sources. We have been able to reliably identify no more than three borrowed items (of Taa or Khoe origin), and suspicions have been raised about a few more (e.g. 'sun'), but on the whole, it seems as if the general process of disintegration was largely driven by internal factors.

Finally, in light of the «Ju-Taa», or «Peripheral Khoisan», hypothesis that interprets the similarities between North Khoisan (Ju) and South Khoisan (!Ui-Taa) in terms of genetic relationship, the following observations must be made:

- on one hand, binary comparisons between Ju and !Ui-Taa that do not find any parallels in #Hoan should not be regarded as significantly less reliable, since #Hoan is an isolated language, and its percentage of irretrievably lost Proto-Ju-Taa items should predictably be higher;
- on the other hand, caution must be exercised when dealing with exclusive #Hoan-Taa isoglosses (such as 'salt', etc.) that do not find parallels in either !Ui or Ju languages, particularly when these isoglosses are exact or near-exact phonetic matches; most likely, such cases reflect recent contact that should not distort our general perspective of distant genetic relationship between these taxa.

Appendix: List of observed phonetic correspondences between Ju and #Hoan

The table below lists all cases of phonetic correspondences that have been observed between Proto-Ju and Eastern #Hoan on the data of basic (Swadesh) lexical items *and* additional lexical items discussed in the main body of the paper (non-Swadesh meanings are listed in italics).

It must be noted that this list does not aim at systematic completeness; thus, there are quite a few segments reconstructible for Proto-Ju (mostly in the non-click consonant domain) that find no #Hoan correlates in this table, and vice versa. Likewise, the table does not contain a special column for Proto-Ju-#Hoan reconstructed phonemes, and while in quite a few cases one-to-many correspondence types are commented upon as to the issue of possible complementary distribution of reflexes, this is not always the case — for instance, there is currently no clear understanding of the principles that govern the reflexes of such vowel qualities as nasalization, glottalization, and pharyngealization, or of the seemingly chaotic distribution of voiced and voiceless reflexes of click phonemes. Such principles may or may not be uncovered at the next stages of etymological research on Ju-#Hoan; in the meantime, what matters most is the *recurrent* nature of such correspondences, proving or at least increasing the probability of their non-accidental nature.

The following types of correlations are included in the table:

(a) phonetically identical segments between #Hoan and Proto-Ju (for such cases, especially if the involved phonemes are rare, recurrence is not necessarily required);

(b) phonetically similar segments between #Hoan and Proto-Ju, differing by no more than one distinctive and commonly unstable feature²⁰ such as +/- voice (for consonants) or +/- raised (for vowels). If the correlation is one-to-one, with no alternate correspondences for either member of the pair, recurrence is not required. If there are conflicting one-to-many correspondences, it is recommended to establish complementary distribution (cf. #1 vs. #1a vs. #1b), or to provide at least as many examples as there are for group (c) cases;

(c) «non-trivial» correspondences, such as #35b, in which the segments differ significantly from each other. To judge such cases as recurrent correspondences, we need to have no fewer than three examples of each (with precise matching semantics or meanings connected by the most trivial of semantic shifts).

For additional examples of possible correspondences and additional comments on those listed in the table below, see Starostin 2008 and Heine & Honken 2010.

Ju	#Hoan	Items	#
a	a	'cold', 'come/fetch', 'earth', 'hear', 'red', 'sleep', 'stand', 'interr. morpheme', '3rd p. sg.', 'enter'	1
	oa	'eye', 'sky' ²¹	1a
	i	'moon/shiny', 'dove', 'refuse/dissuade' ²²	1b
ae	ae	'meat/cut meat'	2
	i	'die'	2a
ai	ai	'puff-adder/snake'	3
	i	'be able'	3a

²⁰ Unstability of features is well demonstrable through the analysis of closely related dialectal forms attested in the Khoisan-speaking area, where fluctuations between voiced/voiceless or high/mid articulation are well known, but the laws that govern such fluctuations have not been described to general satisfaction.

²¹ Regular development after labial clicks.

²² As a result of assimilation in *CaCi ~ *CaCe type structures.

Ju	‡Hoan	Items	#
ao	eu	'short/lack'	4
au	u	'foot/track', 'duiker'	5
	iu	'hand', 'tooth', 'woman/wife', 'dig' ²³	5a
e	e	'female/mother'	6
	a	'new', 'this'	6a
	u	'head', 'one' ²⁴	6b
i	i	'big/many', 'mouth'	7
o	o	'not'	8
	au	'lie/sit'	8a
oa	oa	'stone', 'that'	9
oe	oe	'smoke'	10
	ue	'all'	10a
	ui	'tail', 'take off / drop off'	10b
om	am	'near', 'springhare', 'ripe/cooked'	11
	em	'knee/kneel' ²⁵	11a
u	o	'belly', 'nail', 'horn', 'kill', 'name', 'star', 'water', 'steenbok' ²⁶	12
ui	ui	'small/thin'	13
	oe	'ear'	13a
-ṃ	-am	'eat'	14
-m	-m	'enter'	15
-/V/ŋ	-i	'blood', 'louse', 'see'	16
	-a	'sit'	16a
V>	V	'eye', 'moon/shiny', 'one', 'small/thin', 'that' ²⁷	17
	V ⁱ	'cold', 'nail', 'meat/cut meat', 'enter', 'sky'	17a
V	Ṽ	'die', 'ear', 'head', 'hear', 'mouth', 'not', 'sleep', 'stand'	18
	V>	'foot/track', 'red', 'refuse/dissuade'	18a
Ṽ	Ṽ	'kill', 'star', 'steenbok'	19
	V	'stone'	19a
V ⁱ	V ⁱ	'smoke', 'springhare'	20
V ⁱ >	V ⁱ	'stone'	21
m	m	'I'	22

²³ Regular development after coronal affricates and fricatives.

²⁴ Possibly a regular development after labial clicks.

²⁵ Only as a result of assimilation before an additional front vowel suffix.

²⁶ Mid vowel *o* is a much more frequent ‡Hoan correspondence for PJ **u* than ‡Hoan *u*, which is why HH's interpretation of this correspondence as reflecting a typologically unusual diphthong **ou* in Proto-Ju-‡Hoan (HH: 17) is barely credible. There are a few reliable cases of ‡Hoan *u* : PJ **u* attested as well (HH: 16), but if it turns out to be impossible to prove complementary distribution, it is more likely that additional vowel qualities will have to be set up for the ancestral state, e.g. +/-ATR differentiations (these are known to be phonologically relevant at least for Khoe languages, unlike labial diphthongs such as *ou* or *uo*, virtually unknown in Khoisan languages).

²⁷ Correlations between different types of vowel phonation in PJ and ‡Hoan are clearly very complex. The complexity may be caused by different combinations of features in the protolanguage; the base timbre of the vowel they are associated with; and various types of assimilative / dissimilative interactions with click accompaniments. At present, we lack the data to conduct a more thorough investigation, and list all the possible patterns without evaluating them on behalf of the degree of their regularity.

Ju	ǀHoan	Items	#
t	*t → ɸ	'that', 'near'	23
d	*d → ʒ	'smoke', 'female/mother'	24
cʷ	cʷ	'louse', 'sleep', 'tooth'	25
č	č	'come/fetch'	26
s	c	'hear', 'see'	27
ʒ	z	'new'	28
ʒ ^h	ž	'woman/wife'	29
kx	kx	'earth'	30
h	y	'interr. morpheme', '3rd p. sg.'	31
		'blood', 'ear', 'not', 'sit', 'small/thin', 'refuse/dissuade', 'steenbok'	32
	θ	'eye', 'head', 'one', 'sky', 'duiker'	32a
!	!	'belly', 'bone/spine', 'horn', 'kill', 'lie/sit', 'name', 'red', 'ripe/cooked'	33
†	†	'big/many', 'cold', 'knee/kneel', 'moon/shiny', 'star', 'short/lack', 'be able', 'dove'	34
!!	!	'nail', 'puff-adder/snake', 'foot/track', 'enter'	35
	⊙	'tail' ²⁸	35a
	š/ž	'die', 'hand', 'water', 'dig'	35b
l	l	'meat / cut meat', 'stand', 'stone'	36
C	C	'nail', 'cold', 'die'	37
	ᶑ	'moon/shiny', 'foot/track'	37a
	C ^h	'not', 'dove' ²⁹	37b
	Cxʷ	'small/thin'	37c
ᶑ	ᶑ	'puff-adder/snake', 'stand', 'water'	38
	C	'belly', 'eye', 'hand', 'red', 'enter', 'dig'	38a
č̣	č̣	'ripe/cooked', 'sky'	39
	C ^h	'stone'	39a
Cʷ	Cʷ	'bone/spine', 'duiker'	40
	Cqʷ	'blood' ³⁰	40a
C ^h	C ^h	'big/many', 'horn', 'kill'	41
ᶑ ^h	Cq ^h	'lie/sit' ³¹	42
Cʷ ^h	Cʷ ^h	'knee/kneel'	43
	Cq ^h	'ear', 'steenbok'	43a
Cx	Cx	'tail'	44
ɾč̣	ɾč̣	'head', 'sit', 'be able', 'springhare'	45

²⁸ Very dubious, based on one example only; however, all the other segments in 'tail' match each other so precisely that it is tempting to suggest some sort of rare positional development (perhaps labialization of an original retroflex click before a labial vowel?).

²⁹ Cf. also #39 below. This conflicts with #41, where aspiration is supposed to be preserved in both branches of the family. However, the examples are too semantically precise to be dismissed.

³⁰ PJ does not differentiate between glottal stop and post-velar accompaniments; presumably, ǀHoan is more archaic here, whereas in PJ they generally merged without a trace (see also corr. #43a).

³¹ Dubious, not because of the distinction in voice, but rather because uvular accompaniments in ǀHoan would rather be expected to yield glottalization in PJ (see #43a). Nevertheless, Heine & Honken list several additional examples of similar cases (HH: 29), so this does look like a realistic correlation whose conditions are yet to be properly investigated.

References

- Bell, Arthur, Chris Collins. 2001. †Hoan and the typology of click accompaniments in Khoisan. In: Arthur Bell, Paul Washburn (eds.). *Khoisan: Syntax, Phonetics, Phonology and Contact*. (Cornell working papers in linguistics, 18): 126–153. Ithaca, NY: CLC Publications.
- Bleek, Dorothea F. 1956. *A Bushman Dictionary*. New Haven: American Oriental Society.
- Collins, Chris. 2001. Aspects of plurality in †Hoan. *Language* 77(3): 456–476.
- Collins, Chris. 2002. Multiple verb movement in †Hoan. *Linguistic Inquiry* 33(1): 1–29.
- Collins, Chris, Jeffrey Gruber. 2014. *A Grammar of †Hòã*. Köln: Rüdiger Köppe Verlag.
- Dickens, Patrick. 1994. *English-Ju'hoan / Ju'hoan-English Dictionary*. Köln: Rüdiger Köppe Verlag.
- Doke, Clement M. 1925. An outline of the phonetics of the language of the Chû: Bushmen of north-west Kalahari. *Bantu studies and general South African anthropology* 2(3): 129–165.
- Greenberg, Joseph H. 1966. *The Languages of Africa (2nd edition)*. The Hague: Mouton & Co.
- Gruber, Jeffrey. 1975. Plural predicates in †Hòã. In: Anthony Traill (ed.). *Bushman and Hottentot linguistic studies*. (Communications from the African Studies Institute, No. 2): 1–50. Johannesburg: University of the Witwatersrand.
- 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 Publishing Company.
- Heikkinen, Terttu. 1986. Phonology of the !Xū, dialect spoken in Ovamboland and western Kavango. *South African Journal of African Languages (Pretoria)* 6 (1): 18–28.
- Heine, Bernd, Henry Honken. 2010. The Kx'a Family: A New Khoisan Genealogy. *Journal of Asian and African Studies* 79: 5–36.
- Honken, Henry. 1977. Submerged features and proto-Khoisan. In: A. Traill (ed.). *Khoisan linguistic studies 3. Communications from the African Studies Institute* 6: 145–169. Johannesburg: University of the Witwatersrand.
- Honken, Henry. 1988. Phonetic correspondences among Khoisan affricates. In: R. Vossen (ed.). *New perspectives on the study of Khoisan*. Quellen zur Khoisan-Forschung/Research in Khoisan studies, Bd 7: 47–65. Köln: Rüdiger Köppe Verlag.
- Honken, Henry. 2013. Phonetics and phonology: Eastern †Hoan. In: Rainer Vossen (ed.). 2013. *The Khoesan Languages*: 84–87. London/New York: Routledge.
- Kassian, Alexei, George Starostin, Anna Dybo, Vasily Chernov. 2010. The Swadesh wordlist: an attempt at semantic specification. *Journal of Language Relationship* 4: 46–89.
- Kassian, Alexei, George Starostin, Mikhail Zhivlov. 2015. Proto-Indo-European-Uralic Comparison from the Probabilistic Point of View. *Journal of Indo-European Studies* 43 (3/4): 301–392.
- König, Christa, Bernd Heine. 2008. *A Concise Dictionary of Northwestern !Xun*. Köln: Rüdiger Köppe Verlag.
- Lionnet, Florian. 2009. *Lucy Lloyd's !Xun and the Ju dialects*. Linguistics colloquium, Institute for Asian and African Studies, Humboldt University, Berlin, 10 November 2009. Available online at: <http://www.iaaw.hu-berlin.de/afrika/linguistik-und-sprachen/veranstaltungen/afrikalinguistischeskolloquium/papers-wintersemester-2009-10/lucy-lloyds-xun-and-the-ju-dialects>.
- Miller, Amanda L. 2013. Phonetics and phonology: Northern Khoesan. In: Rainer Vossen (ed.). 2013. *The Khoesan Languages*: 45–51. London/New York: Routledge.
- Pratchett, Lee James. 2017. *A documentation of Groot Laagte †Kx'aoll'ae and dialectal diversity in Southeastern Ju (Kx'a)*. Ms.
- Sands, Bonny. 2010. Juu subgroups based on phonological patterns. In: M. Brenzinger, Ch. König (eds.). *Proceedings of the 1st International Symposium, January 4-8, 2003, Riezlern / Kleinwalsertal*: 40–73. Köln: Rüdiger Köppe Verlag.
- Sands, Bonny. 2018. *From Calgary to the Kalahari: An exploration of language diversity and endangerment*. Presentation given at the University of Calgary, March 2, 2018.
- Sands, Bonny, Henry Honken. 2014. †Amkoe body part terminology in comparative perspective. In: Tom Güldemann, Anne-Maria Fehn (eds.). *Beyond 'Khoisan'. Historical relations in the Kalahari Basin*: 233–254. Amsterdam: John Benjamins Publishing Company.
- Snyman, Jan W. 1980. The relationship between Angolan !Xu and Žul'õasi. In: Jan W. Snyman (ed.). *Bushman and Hottentot linguistic studies* 1979: 1–58. Pretoria: University of South Africa.
- Snyman 1997: Snyman, Jan W. A preliminary classification of the !Xū and Zhu|hõasi dialects. In: W. Haacke, E. Elderkin (eds.). *Namibian languages: reports and papers*: 21–106. Köln: Rüdiger Köppe Verlag.

- Starostin, George. 2003. A lexicostatistical approach towards reconstructing Proto-Khoisan. *Mother Tongue VIII*: 81–128.
- Starostin, George. 2008. From modern Khoisan languages to Proto-Khoisan: The value of intermediate reconstructions. In: Smirnov, Ilya (ed.). *Aspekty komparativistiki III. Orientalia et Classica: Trudy Instituta vostochnyx kul'tur i antichnosti XIX*: 337–470. Moscow: RSUH Publishers.
- Starostin, George. 2016. From wordlists to proto-wordlists: reconstruction as «optimal selection». *Faits de langues* 47/1: 177–200.
- Starostin, Georgij. 2013. *Jazyki Afriki: opyt postrojenija leksikostatisticheskoy klassifikacii. Tom I: Metodologija. Kojanskije jazyki*. Moscow: Jazyki slav'anskoj kul'tury.
- Traill, Anthony. 1973. “N4 or S7”: another Bushman language. *African studies* 32 (1): 25–32.
- Traill, Anthony. 1994. *A !Xóǀ dictionary*. Köln: Rüdiger Köppe Verlag.
- Traill, Anthony, Hirosi Nakagawa. 2000. A Historical !Xóǀ-ǀGui Contact Zone: Linguistic and other Relations. In: Batibo, Herman M., J. Tsonope (eds.). *The State of Khoesan Languages in Botswana*: 1–17. Mogoditshane: University of Botswana/University of Tromsø, Tasalls Publishing and Books.
- Treis, Yvonne. 1998. Names of Khoisan Languages and their Variants. In: Mathias Schladt (ed.). *Language, Identity, and Conceptualization among the Khoisan*: 463–495. Köln: Rüdiger Köppe Verlag.
- Vossen, Rainer. 1997. *Die Khoe-Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Köln: Rüdiger Köppe Verlag.
- Vossen, Rainer (ed.). 2013. *The Khoesan Languages*. London/New York: Routledge.
- Westphal, Ernst O. J. 1974. Notes on A. Traill: “N4 or S7?”. *African studies* 33 (4): 243–247.
- Westphal, Ernst O. J. 1966. Linguistic research in SWA and Angola. In: *Die ethnischen Gruppen Südwestafrikas. Wissenschaftliche Forschung in Südwestafrika, Bd. 3*: 125–144. Windhoek: Südwestafrikanische Wissenschaftliche Gesellschaft.

Г. С. Старостин. Лексикостатистические исследования по койсанским языкам I: родство между языками жу и ǀхоан

Статья представляет собой первое из серии планируемых исследований по сравнительной лексикостатистике ряда языковых семей, традиционно (со времен Дж. Гринберга) причисляемых к гипотетической койсанской макросемье. В настоящей публикации проводится детальный лексикостатистический анализ данных по двум таксонам: языкам жу, или северной койсанским (пучок относительно близкородственных диалектов), и языку восточный ǀхоан, который до недавнего времени рассматривался как изолят, но сегодня все же скорее считается ближайшим родственником языков жу. На основании как поверхностного (фонетические сходства), так и этимологического (фонетические соответствия) анализа возможных когнатов между языками жу и ǀхоан число лексикостатистических сходжений между ними определяется в диапазоне от 32% до 44%, что примерно соответствует такой же глубине родства, как между финно-угорскими или картвельскими языками. Помимо этого, анализируется также дистрибуция когнатов между различными слоями базисной лексики (более/менее устойчивыми), что дает основание утверждать именно о генетическом родстве, а не об ареальных связях между обоими таксонами.

Ключевые слова: койсанские языки, жу языки, язык ǀхоан, лексикостатистика, глоттохронология, сравнительно-исторический метод.

Отражение южнославянского *r в заимствованиях из старорумынского в цыганский¹

Сплошной просмотр северновлашских цыганских (кэлдэрарских и кишинёвских) заимствований из румынских диалектов подтверждает, что: (1) румынский язык до позднего времени безупречно сохранял латинское *-rr-* внутри слова, (2) в балканороманских диалектах, в т. ч. в румынском языке, в начале слова лат. *r-* > **rr-*. Кроме того, можно сделать новое наблюдение, что (3) в южнославянском диалекте, из которого заимствовали много слов носители «раннерумынского» языка, было два отчётливых аллофона /r/: мягкий — перед **ě* (и, возможно, перед **b*) и твёрдый — во всех остальных случаях (в т. ч. перед **e* и, вероятно, перед **i*), которые давали рум. *-r-* и *-rr-* соответственно.

Ключевые слова: цыганский язык, румынский язык, южнославянские языки, языковые контакты, историческая фонетика

Здесь пойдёт речь о слабо изученном румынском архаизме — сохранении «долгого» */r̄/. В современном румынском литературном эта фонема полностью слилась с обычным /r/, а её следы в ряде случаев видны лишь по следующим задним гласным: *i* > *î* и *e* > *ă*. Чаше всего это проявляется в начале слова: лат. *rīsu-* > **rrisu* > **rrīsu* > рум. *rîs* ‘смех’; в середине слова примеров мало: ср. лат. *interrītāre* > рум. *întârîtă* ‘раздражать’, но лат. *ēricius* > **arici* > рум. *arici* ‘ёж’ (не ***arîci*).

То же развитие гласных, что и после */r̄/, имело место после отвердевших согласных (в диалектах): **sita* (из слав.) > диал. *śitā*, литер. *sitā* ‘сито’; **zidu* (из слав.) > диал. *zîd*, литер. *zid* ‘стена’; лат. *vessica* > диал. *băşîcă*, литер. *băşică* ‘пузырь’; **žiru* (из слав.) > диал. *ĵîr*, литер. *jir* ‘жёлудь’ (Rosetti 1986: 625).

Старая долгая фонема */r̄/ («*rr*») сохраняется маргинально в части румынских диалектов, чьи описания ненадёжны и обрывочны. В большинстве случаев следы старого состояния почти затёрты. Рукописные свидетельства дают довольно мало, т. к. особый знак для /r̄/ писался непоследовательно (Rosetti 1986: 480; Densusianu 1938: 121)².

Однако от внимания исследователей, насколько нам известно, ускользает, что это */r̄/ сохраняется в цыганских (северновлашских) заимствованиях из румынского. Цыганские влашские диалекты формировались начиная примерно с XIII–XV вв. в контакте с румынским языком. В них эта фонема (обозначаемая как /r̄/ и реализуемая раскатисто или картаво, т. е. заднеязычным [ʁ]³) фигурирует не только в заимствованиях из румын-

¹ Работа выполнена при поддержке фонда РФНФ, грант №16-34-01044: «Изучение и описание влашских диалектов цыганского языка в России». Автор благодарит С. Г. Болотова, Марка Гринберга, Мате Каповича, А. С. Касьяна, К. А. Кожанова, С. А. Лашина, Орсата Лигорио, В. А. Панова, Войцеха Смочиньского и анонимного рецензента. Условные обозначения: * — реконструированная, не засвидетельствованная форма; ** — не существующая (в некоторых случаях ожидаемая) форма; > — даёт фонетически; < — происходит фонетически из; ← — заимствовано из; ⇐ — получено морфологически из; ~ — находится в свободном варьировании с; : — находится в отношении чередования с; + — сложение с.

² Впрочем, Росетти пишет, что в XVI в. в части памятников обозначение *ř* постоянно, но приводит примеры, из которых, если мы правильно их понимаем, следует скорее обратное.

³ Как, например, факультативно в португальском, ср. порт. *riso* ‘смех’ [ʁrizu] или [ʁizu].

ского, но и в исконной (индийской) лексике, где восходит к индийским ретрофлексным зубным *ṭ* и *ḍ* в ряде позиций (Ослон 2018б). Это значит, что ко времени контакта с румынами, например, др.-инд. *ḍomba-* ‘член касты домов’ (в новоиндийских языках: *ḍom*) > цыг. *ṛom* ‘цыган; муж’ скорее всего звучало как [ṛom] (с долгим, раскатистым сонантом, как сейчас во многих диалектах и говорах) и так же, вероятно, звучало румынское */ṛ/, которое заимствовалось в цыганский именно в этом виде. Больше всего примеров на такие заимствования показывает кэлдэрарский диалект.

В связи с румынским пластом заимствований в кэлдэрарском (как и вообще в северновлашских диалектах) имеется несколько загадок. Так, кэлдэрары и близкие им группы по сей день сохраняют старорумынские формы ряда слов, ср. кэлд. *vr'ămia* ‘время’ (ст.-рум. *vreāte*, совр. *vréme*) и др., притом что в румынских диалектах, кажется, нет никаких следов таких форм уже несколько сот лет. Сохранение /ṛ/ у северновлашских цыган не менее загадочно. Складывается впечатление, что они активно имели дело с румынами в средние века, но потом в течение долгого времени или вообще с ними не общались (что совершенно невероятно, ведь они были крепостными у румынской знати, а также бродячими ремесленниками в исключительно румынском языковом окружении), или владели только каким-то особым архаичным вариантом румынского языка, проявлявшим поразительную стойкость против влияния современных ему румынских диалектов.

Приведём более или менее исчерпывающие списки заимствований с фонемой /ṛ/ в кэлдэрарском (в основном наши данные (Ослон 2018а), в противном случае из словаря ДД) и кишинёвском⁴ (Кожанов рук.) диалектах, механическую реконструкцию источника заимствования («раннерумынская»⁵ форма)⁶, современную литературную румынскую форму, а также краткую этимологическую справку. Поочерёдно рассмотрим начальную, срединную и конечную позиции в слове.

Начало слова

В начале слова северновлашские /ṛ/ и /r/, как и во всех остальных позициях, противопоставлены фонологически. В дорумынских словах (т. е. исконных индийских и грецизмах) здесь может быть как ṛ-, так и r-, ср. *ṛaj* ‘палка’ : *raj* ‘господин’; *ṛyl* ‘flatulentia’ : *riś* ‘медведь,’ *riṇ* ‘напильник’ (грецизм); *ṛoj* ‘ложка’ : *rovlì* ‘трость’. Однако в заимствованиях из румынского здесь бывает только /ṛ/. Начнём с исконнорумынских, т. е. романских слов:

кэлд. *ṛyso* ‘очень’, ‘смех’ (ДД) ← раннерум. **rrîsu* (рум. *rîs*; в памятниках: *rride*, ‘смеяться’, Densusianu 1938: 121) < лат. *rîsus*;

кэлд. *ṛyра* ‘яма’ ← раннерум. **rrîpa* (рум. *rîpă*) < лат. *ripa*;

кэлд. *ṛădăcîna* ‘корень’ ← раннерум. **rrădăcîna* (рум. *rădăcînă*; в памятниках: *rradăcinra*, Densusianu 1938: 121) < лат. *rădîcîna*;

кэлд. *ṛăvda* ‘терпение’ ← раннерум. **rrăbda*, неясно, но ср. рум. *răbdă* ‘терпеть’ (в памятниках: *rrebda*, Densusianu 1938: 121);

кэлд. *ṛăca* ‘утка’ ← раннерум. **rrăța* (рум. *rața*) < (?) **ratia* (Skok 1971: 92);

⁴ Кишинёвский диалект — ответвление лаэшкового, самого распространённого диалекта нынешнего государства Молдавия, массово представленного также на территории Румынии.

⁵ К «раннерумынскому» языку мы здесь условно относим праформы румынских слов (с фонемой /ṛ/), в разное время (?) заимствованных северновлашскими цыганами. «Старорумынскими» мы называем слова, фиксируемые румынскими памятниками до XVIII в. включительно.

⁶ Судя по всему, кэлдэрары заимствовали румынские существительные в определённой форме (т. е. с постпозитивным артиклем): с гласным исходом на —*a*, с согласным — на —*i*.

- кэлд. *řàta* ‘колесо’ ← раннерум. **rrōáta* (рум. *roată*) < лат. *rota*;
 кэлд. *řàkàrà* ‘прохлада’ ← раннерум. **rrākoárea* (рум. *răcoáre*) ⇐ **rrāce* (рум. *rece*, диал. *rāce* ‘холодный’ < лат. *recens* ‘свежий’ ← ‘недавний’, DER: 685, 694);
 кэлд. *řŭncedo* ‘прогорклый’ ← раннерум. **rrŭncedu* (рум. *rŭnced*) < лат. *rancidus*;
 кэлд. *řuzŭna* ‘ржавчина’ ← раннерум. **rrugŭna* (рум. *rugŭnă*) < лат. **aerŭgina* (DER 708);
 кэлд. *řŭta* ‘червяк’ ← раннерум. **rrŭta* (рум. *rŭtă*) ⇐ *rŭtă* ‘нюхатъ’ < лат. *rimāre*;
 кэлд. *řagŭla* ‘семейный круг’ (?) ← раннерум. **rrāgŭla* (рум. *răgŭla* ‘род, племя’) < лат. *rēgula*;
 кэлд. *řŭja* ‘злой’ (прил. ⇐ сущ.) ← раннерум. **rrŭja* (рум. *rŭie* ‘парша’) < **rrānja* < лат. *arānea* ‘паутина’;
 кэлд. *řavŕil* ‘моросит’ ← раннерум. **rrŭurá* (рум. *rŭurá*) ⇐ **rrŭci* (рум. *rŭi* ‘река’; в памятниках: *rriu*, Densusianu 1938: 121) < лат. *rŭvus*;
 кэлд. *řudil* ‘молит’ ← раннерум. **rrugá* (рум. *rugá* ‘просить’; в памятниках: *rruga*, Densusianu 1938: 121) < лат. *rogāre*;
 кэлд. *řagušŭl* ‘хрипнет’ ← раннерум. **rrāgušŭ* ⇐ лат. приставка *re-* + рум. *gúšă* ‘зоб’;
 киш. *řàva* ‘роса’ ← раннерум. **rrōáva* (рум. *róuă*) ⇐ лат. *rō[s]-* (как рум. диал. *zíuă* ⇐ *zi* ‘день’) (DER: 707).

Таким образом, в начале слова наблюдаем ситуацию, аналогичную таковой в иберийских языках, ср. исп. */ř/áiz* ‘корень’. Среди балканороманских языков имеем протезу в арумынском, ср. (вариант) *arāmān* ‘арумын’ < **rromāni* < лат. *rōmānus* ‘римлянин’. Всё это говорит о том, что балкано-романские народнолатинские говоры фонетически развили раскатистый приступ: лат. *r-* > **rr-*, как в иберийской латыни. Начальное раскатистое *rr-* в румынском в какой-то мере отражено ранними памятниками (Densusianu 1938: 37).

Точно так же, как в цыганских заимствованиях из румынского исконно романских слов, имеем только начальное */ř/* во всех славизмах, попавших в цыганский через румынский:

- кэлд. *řŭndo* ‘ряд’ ← раннерум. **rrŭndu* (рум. *rŭnd*; в памятниках: *rrŭnd*, Densusianu 1938: 121) ← юсл. **rĕdъ*;
 кэлд. *řŭda* ‘дышло’ ← раннерум. **rrŭda* (рум. *rŭdă*) ← юсл. **rŭdà*;
 кэлд. *řatija* ‘водка’ ← раннерум. **rrākŭja* (рум. диал. *răchie*) ← юсл. *rakŭja* ← тур. *rakŭ*;
 кэлд. *řòfa* ‘платье’ ← раннерум. **rrókŭja* (рум. *róchie, róche*)⁷ ← юсл. (?), ср. схрв. *ròklja* ← венг. *rokol(y)a* ← нем. *Röcklein* (Skok 1971: 156);
 кэлд. *řŭbla* ‘рубль’ ← раннерум. **rrŭbla* (рум. *rŭblă*) ← юсл. (?) ← рус. *рубль*;
 кэлд. *Řusŭja* ‘Россия’ ← раннерум. диал. **Rrusŭja* (рум. стар. *Rusŭa*, Гинкуловъ: 5, теперь *Rusia*) ← юсл. *Rusŭja* (новошток. *Rŭsija*);
 киш. *Řusàle* ‘Троица’ ← раннерум. **Rrusalije* (рум. *Rusălie, Rusăle*) ← юсл. или через другое посредство (с фонетич. трудностями) из лат. *Rosălia* (DER: 710);
 кэлд. *řatèzo* ‘щеколда’ ← раннерум. **rrātèzu* (рум. *rătez, retez*) ← юсл. ?, ср. рус. *пéтязь*, пол. *rzeciądz* (венг. *retesz*);
 кэлд. *řygyjil* ‘рыгает’ ← раннерум. **rrŭgŭjŭ* (рум. *rŭgŭjŭ*, наст. 3. ед. *rŭgŭie*) ← юсл. **rŭgati* (ударение как в рум.!);
 кэлд. *řăspil* ‘рассеивает’ ← раннерум. **rrăšŭpŭ* (рум. *risipŭ*, вост. диал. *rŭšŭpŭ*, стар. *răsipŭ*) ← юсл. **rasip-* < **raz-syp-*;
 кэлд. *řazgajil* ‘нежит’ ← раннерум. **rrăzgŭjŭ* (рум. *răzgŭjŭ*) ← юсл. ? (ср. болг. *разгалям*).

⁷ В старорумынском *kl' > k'* не позднее XV в., причём, возможно, в разных диалектах по-разному (Rosetti 1986: 480). Примерно в это время начался контакт влашских диалектов цыганского языка с румынским.

Середина слова (между гласными)

В срединной позиции видно, что сохраняется латинское *-rr-*:

кэлд. *càřa* ‘народ’ (*‘страна’) ← раннерум. **řarra* (рум. *řarã* ‘страна’) < лат. *terra*;
кэлд. *kəřàřa* ‘пробор’ ← раннерум. **kărràřea* (рум. *căřare* ‘тропинка; пробор’) < лат. *carraria*;

киш. *kyřica* ‘повозка’ ← раннерум. **kărrúřa* (уменьш.; рум. *căruřã*) ⇐ **carru* < лат. *carrus* (DER: 140);

киш. *kyřyl* ‘носит’ ← раннерум. **kărrá* (рум. *cărá*) < **carrāre* ⇐ *carrus* (DER: 140);

киш. *skuřyl / skořyl* ‘выжимает; уменьшается’ ← раннерум. **skurrá* ⇐ лат. *excurrere* ‘убегать’ (в т. ч. о жидкости), очевидно, в румынском смешение с корнем *cur-* ‘чистить’ (DER: 266), однако, как видим, цыганский сохраняет долгий согласный;

кэлд. *bəřandəvo* ‘потроха (блюдо)’ ← раннерум. **borrindăcu* (рум. *borindău* ‘блюдо с кровью и мясом’) неясного происхождения, но с первой частью, возможно, под влиянием глагола *borí* ‘блевать’ < **abhorrire* (DER: 98);

киш. *čořoj* ‘ворон’ ← раннерум. **čorrói* (рум. *cioroi*) при *cioárã* ‘ворона’ < **čorra* неизвестного происхождения, но ср. алб. *sorrë* (соответствие *rr*);

кэлд. (*v*)*uřyto* ‘противный’ ← раннерум. причастие **urrítu* (рум. *urít*) от глагола **urrí* (рум. *urí* ‘ненавидеть’) < **horrire* (DER: 877).

Сюда же кэлд. *cųřa* ‘чуть-чуть’ ← раннерум. **řirra* (рум. *řirã*) без этимологии.

Там, где в латыни обычное *-r-*, имеем обычное *-r-* и в цыганском:

кэлд. *vəro* ‘двоюродный брат’ ← раннерум. **văru* (рум. *văr*) < лат. *verus*;

кэлд. *múra* ‘ягода’ ← раннерум. **múra* (рум. *múřã*) < лат. мн. *mora* и мн. др.

Ср. также кэлд. *-ri-* ← раннерум. *-ri-* < лат. *-lī-*:

кэлд. *burika* ‘пупок’ (смена рода) раннерум. **buriku* < лат. **umbilicus* (рум. *buríc*)

Но есть пример (один?) на «лишнее» долгое /ř/:

кэлд. *pàřirřa* ‘рогоз’ ← раннерум. **pàřirra* (рум. *pápurã* ‘рогоз’) при предполагаемом лат. **papurra* ⇐ лат. *papyrus* ← гр. *πάπυρος*.

При этом в срединной позиции между гласными в славизмах, полученных цыганами через румынское посредство, /ř/ встречается в большинстве примеров:

кэлд. *văřo* ‘известь’ ← раннерум. **vărru* (рум. *var*) ← юсл. **vârŕb*;

кэлд. *dăřo* ‘взнос от гостя на свадьбе’ ← раннерум. **dărru* (рум. *dar*; в памятниках: *darrŭ*, Rosetti 1986: 480) ← юсл. **dârŕb*;

кэлд. *komàřa* ‘клад’ ← раннерум. **komqarra* (рум. *comoárã*) ← юсл. **komðra* < псл. **kombra* (← лат. *càtara* ← гр. *καμάρα* ‘свод’);

кэлд. *pàřa* ‘пламя’ ← раннерум. **pàrra* (рум. *pářã*; в памятниках: *parra*, Densusianu 1938: 121) ← юсл. *pãra*;

кэлд. *sųřo* (ДД) ‘серый’ ← раннерум. **sųrru* (рум. *sur*) ← юсл. **sŭrŕb*;

кэлд. *zgųřa* ‘зола в печи’ ← раннерум. **zgúrra* (рум. *zgúřa*, *sgúřa* ‘окалина’) ← юсл. **zgir...* (?), ср. болг. *сгурѝя*, *згурѝя* ‘шлак’, с озвончением *z-* < **ск-* (по аналогии с болг. *горѝя* ‘гореть’ (?), БЕР VI: 570) ← нгр. *σκουρ(ι)ά* < *σκωρία* ‘ржавчина, накипь, шлак’ (DER: 755).

Во всех приведённых примерах раннерум. /r̄/, как представляется, является **тривиальным отражением** славянского *r перед непередними гласными — здесь перед *a или *o. Этого не понимал О. Денсушиану: «... le même texte [Psaltirea Hurmuzaki] offre en outre rr à l'intérieur de quelques mots... là, rr semble avoir été introduite arbitrairement par le copiste» (Densusianu 1938: 121); при этом он приводит примеры (пропускаем однокоренные): *amarră, hotarră, izvorrele, oborri, omorri, orriiașul, păharri, parră, părrasi, vorroavă, zborrulu*. Замечательно, что во всех примерах (кроме *amarră*) здесь ожидается именно *rr* — всё это славизмы (кроме германизма *păharri* и мадьяризма *orriiașul*, а также, возможно, *hotarră*)⁸.

Есть, однако, также цыганские заимствования, где можно подозревать какие-то передние гласные. Это несколько глаголов, где также имеем цыг. ř (и рум. r перед î):

кэлд. *pyřyl* ‘доносит, предаёт’ ← раннерум. *pārrî (рум. *pîrî*) ← юсл. *p̄vr-, ср. болг. диал. *прѣ се* ‘ссорюсь’;

кэлд. *vyřyl* ‘суёт’ ← раннерум. *vārrî (рум. *vîrî*) ← юсл. *v̄vr-, ср. болг. *врѣ* ‘сую’;

кэлд. *tyřyl* ‘тащит’ ← раннерум. *tārrî (рум. *tîrî*, но и стар. *tirî*, Oczko 2014: 284) ← юсл. *t̄vr-, ср. схрв. *trēm, tārēm* ‘тру’ (болг. *трѣя* ‘тру’).

Однако по крайней мере в двух глаголах — *r*:

кэлд. *năzărîl* ‘виднеется’ ← раннерум. *nāzārî (рум. *năzărî* ‘мерещиться; завидеть’) ← юсл. *nazvr̄ēti;

кэлд. *părîl* ‘шпарит’, киш. ‘пáрит’ ← раннерум. *pārî (только с приставкой рум. *opārî* ‘ошпарить’ и др., см. ниже) ← согласно DER: 582, слав. *opariti (но см. ниже).

В чём же тут дело? Непонятно, из каких именно форм заимствованы румынские глаголы первой группы: *pîrî, vîrî, tîrî*. Например, *tîr-î* может быть из чего-то типа *tr̄-ti, *tr-ŏ (ср. схрв. *t̄rti, trēm/tārēm*), при возможных вариантах в инфинитиве типа сербск.-церк.-слав. *тр̄пти* с ятем (< *terti), ср. рус. *терѣть*. Как бы то ни было, здесь тот же рефлекс, что и перед *a или *o (в именах).

Что же касается второй группы: *năzărî, opārî*, — то эти глаголы по крайней мере в инфинитиве **содержат ять, который влияет на качество -r-**. В случае *năzărî* ← юсл. *nazvr̄ēti имеем ещё рум. *zărî* ‘видеть смутно, завидеть’ ← юсл. *zvr̄ēti. Других вариантов инфинитива нет. А глагол *opārî*, который на первый взгляд заимствован из *o-pār̄iti, на самом деле представляет собой смешение двух глаголов: *pār̄iti и *pvr̄ēti в пользу последнего, что явствует из его значений, среди которых присутствует ‘преть’: (1) ‘обливать кипящей жидкостью’, возвр. ‘ошпариться’ (← *pariti), (2) возвр. (обычно о детях) ‘вызывать у себя раны или раздражения, обычно на суставах, потом или мочой (← *pvr̄ēti), (3) возвр. (диал. о кукурузе, муке, сухих растениях) ‘преть, портясь’ (← *pvr̄ēti), (4) возвр. (о растениях) ‘вянуть от жары, плесневеть от солнца после дождя (← *pvr̄ēti)⁹. Это значит,

⁸ Из написаний *vorroavă* ‘слово’, *vorrovi* ‘говорить’ в румынских памятниках следует высокая вероятность славянского происхождения этого гнезда. Само слово *vorová* важно ввиду его вероятной связи с рум. *vorbă* ‘тж.’, о чьей этимологии давно спорят (вторая гипотеза связывает его с лат. *verbum* ‘тж.’), см. (Pamfil 2003: 21). Возможно, и *vorová*, и *vorbă* — отглагольные имена от двух вариантов глагола (если это варианты): *vorovi* и *vorbi*. Тогда источником заимствования может быть что-то вроде укр. диал. *вогоріти* ‘говорить’ (с другой метатезой).

⁹ **OPĂRÍ**, *opăresc*, vb. IV. 1. A turna peste ceva un lichid clocotit (apă, leșie, lapte etc.) pentru a spăla, a curăța de coajă, a găti etc. ♦ Tranz. și refl. A provoca sau a căpăta arsuri, turnând sau vărsând apă clocotită sau alt lichid foarte fierbinte; a (se) arde. 2. Refl. (Mai ales despre copiii mici) A face răni și iritații în unele părți ale corpului, mai ales la încheieturi, din cauza transpirației, a urinei etc. 3. Refl. (Reg.; despre mălai, făină, plante uscate) A se

что, хотя в цыганском имеется только значение ‘шпарить’, форма этого глагола в румынском источнике не имеет отношения к славянской форме *pariti.

В позиции после согласного или перед согласным выступает только обычный /r/ (видимо, в раннерумынском не было **Crr и **rrC, хотя в исконноцыганских словах Cř, řC вполне возможны), в т. ч. в славизмах, ср. в глаголах:

кэлд. *primil* ‘принимает’ ← раннерум. **primí* (рум. *primí* ‘принимать’, стар. *priimí*, DER: 666 ← юсл. **pri-jьm-*, ср. болг. *приема, приемам, приймам* ‘принимаю’);
кэлд. *opril* ‘запрещает’ (ДД) ← раннерум. **opri* (рум. *opri* ‘остановить’) ← юсл. **o-p(ь)r-* (с падением ера; ср. *pîri* ← **pьr-* без приставки), ср. болг. *опра* ‘обопру’ и мн. др.

Из всего этого делаем вывод, что между гласными в языке-источнике славизмов встречалось два аллофона фонемы /r/. Один из них — мягкий (перед ятем в славянском источнике) передавался румынской фонемой /r/, а другой (во всех остальных случаях, в т. ч., видимо, всегда в начале) передавался румынским /r̄/ (раскатистым). Эта разница точно воспроизводилась цыганами с помощью двух соответствующих фонем.

В цыганском есть также простое /r/ в существительном, где можно предполагать (с натяжкой) славянский исход на *-rb (т. е. перед ерем), чему, однако, противоречит написание в старорумынском:

кэлд. *iforo* ‘метель’ ← раннерум. *?*iforu* (рум. *vîhor* ‘вихрь’, диал. *vîvor*; в памятниках: *vîhorrulu*, *Densusianu* 1938: 121) ← юсл. **vixorb*, ср. близкую форму (с *-b-, а не *-o- в основе) болг. *вихър*; конечное *-b имеем в рус. *вихрь* — впрочем, тогда непонятно, откуда окончание раннерум. *-и.

Кроме того, в цыганском простое /r/ выступает в суффиксе -ar- (частично славянского происхождения):

кэлд. *bufâri* ‘книга, кошелёк’ ← раннерум. **bufâri* (рум. стар. *bucvâriu*, *bucvâri*, *bucvâr*, *bufâr* ‘церковная книга’) ← юсл. **bukъvâr(j)ь*.

В новогреческих заимствованиях через румынский — только /r/:

кэлд. *zâharo* (ДД), *zâro* ‘сахар’ ← раннерум. **zâharu* (рум. *zâhar*) ← нгр. ζάχαρ(ο);
кэлд. *vorèzo* ‘рис’ ← раннерум. **orèzu* (рум. *oréz*) ← (?) нгр. ορούζι, ορούζα, с неясным окончанием и корневым гласным в румынском (DER: 584);
кэлд. *kârâmida* ‘кирпич’ ← раннерум. **kârâmida* (рум. *cârâmîdă*) ← ср.-гр. κεραμίδα (DER: 142).

Это значит, что румынский язык по-разному воспринимал славянское и греческое /r/.

Что касается других языков-источников, то по цыганскому видно, что немецкое и турецкое /r/ по крайней мере в части случаев румынский передавал через */r̄/:

кэлд. *pahârô* ‘стакан’ ← раннерум. **pahârri* (рум. *pahâr*; в памятниках: *pâharri*, *Densusianu* 1938: 121) ← свн. *behhari* (неясно окончание);
кэлд. *tatârô* ‘мусульманин’ ← раннерум. **tatârri* (рум. *tâtâr* с редукцией) ← тур. *tatar*;
кэлд. *katîrô* ‘мул’ (ДД) ← раннерум. **katîrru* (рум. *catîr*) ← тур. *katir*.

încinge, alterându-se. 4. Refl. (Despre plante) A se ofili, a se veşteji (din cauza căldurii); a se măna din cauza soarelui prea fierbinte apărut după ploaie (DEX).

Загадочное слово:

кэлд. *kàpara* ‘залог в виде подарка невесте’ ← раннерум. ? (рум. *capără*) ← ит. *caparra* ‘залог’ при схрв. *kàpara* и нгр. *καπάρο* (среднего рода) (DER: 136); непосредственный источник заимствования в румынский неясен, отсутствие /r̄/ загадочно; может быть, это пример новорумынской формы в кэлдэрарском?

Конец слова (?)

В кэлдэрарском есть всего одно слово (слово-высказывание), для которого в «раннерумынском» можно предполагать отсутствие гласного окончания. Это слово (в роли союза) румыны должны были заимствовать, не добавляя никакого гласного в конце:

кэлд. *bàtăř* ‘ей-богу’ ← раннерум. **bátărr* (рум. диал. *bátîr*, *bátâr* ‘хотя; по крайней мере’) ← венг. *bátor* ‘quoique, quand même’ (Cihac 1879: 481).

Возможно, такое конечное венгерское *-r* всегда передавалось как /r̄/.

Возвращаясь к рефлексам южнославянского **r*, в целом понятно, что они не различаются перед слав. **i* и **y*. В начале слова всегда имеем **rrî-*, однако это не значит, что на стадии заимствования из славянского в румынском не было **ri-*¹⁰:

рум. *Rîm* ‘Рим’ (при, видимо, более новом *Rim*) < **Rrîmu* ← юсл., ср. болг. *Рим*, схрв. *Rîm*, *Rîma* < **Rîmъ*; также *rîmleán*;

рум. *rîbiță* ‘нек. виды мелких рыб’ < **rrîbița* ← юсл., ср. болг. *рѹбица* < **rŷbica*; сюда же рум. *rîmnic* ‘пруд’;

и мн. др. (в т. ч. *rîza* ‘риза’, *rîs* ‘рысь’).

На начальное слав. **rě-* надёжных примеров нет, но есть один на начальное слав. **rb-*, которое тоже даёт твёрдость (?):

рум. *rîvni* ‘жаждать’ < **rrăvni* ← юсл., ср. болг. диал. *рѣвна се* ‘нравлюсь’ < **rŷvni-*; самые ранние написания: *răvni*, затем *rîvni*, *rîmni*, *revni* (уподобление гласных?) (Oszko 2014: 253).

В середине слова имеется всего несколько примеров на слав. **-ri-* с обоими рефлексами в румынском, ср. *-rî-* (по крайней мере, один пример)¹¹:

рум. *tărîță*, обычно мн. *tărîțe* ‘отруби’ < **tărrîța* ← юсл., ср. болг. мн. *трѹци*, схрв. мн. *trîce* ‘тж.’ < **tŷrîca*, очевидно, ← **tŷr-* ‘тереть’; таким образом, здесь не павший ер, так же, как в однокоренном *tîrî* ‘тащить’ (см. выше), что может говорить о раннем времени заимствования;

¹⁰ Исконные румынские слова на *ri-* типа *risipi* ‘рассыпать’, *ridică* ‘поднимать’, *ridiche* ‘репа’, по-видимому, во всех случаях — результат ассимиляции *_ă_i > _i_i* (ср. варианты *răsipi*, *rădică*, *rădiche*). То же касается слов на *re-*, где *_ă_e > _e_e*, ср. *retéz* ‘щеколда’, *reveni* ‘вымачивать’ и др. Можно думать, что эта ассимиляция стала возможна, т. к. к этому моменту уже стало возможным простое начальное *r-*. Это в какой-то степени так уже с XVII в. (хотя формы типа *risipi* ещё редки) (Rosetti 1986: 450). Простое начальное *r-* выступает в поздних заимствованиях, ср. рум. *rip* ‘грязный’ ← нгр. *ρύπος* (DER: 701), *rif* ‘мера длины’ ← тур. *rif* (DER: 699).

¹¹ Сюда явно не относится рум. *țărînă* ‘пахотная земля’, которое иногда возводят к южнославянскому, ср. схрв. *čarina* ‘пошлина’. Впрочем, вероятно, имела место контаминация, судя по вариантам *țărînă*, *țarină* (DER: 826); латинское происхождение от **terra* подтверждается написанием в румынских памятниках: *țerrină*, наряду с *țarrănă* (Rosetti 1986: 480).

Но и *-ri-*:

рум. *coriță* ‘корыто’ ← юсл. ?, ср. болг. *korúto*, схрв. *kòrito* < **korýto*;

рум. *verigă* ‘звено; кольцо’ ← юсл. ?, ср. болг. *verúga*, схрв. *vèriga* < **verìga* (рум. *verigă* → кэлд. *viriga* ‘кольцо’);

рум. *pérină*, вар. *pernă* ‘перина’ (первый вариант, видимо, более старый, однако встречается только с конца XVII в., Oczko 2014: 215) ← юсл. ?, ср. болг. *périna*, схрв. *pèrina* < **perina* (рум. *pérină* → кэлд. *pèrina* ‘перина’).

Славизмы с суффиксом *-ica* дают *-ri-* (все, кроме *țărîță*, см. выше): *vèveriță* ‘белка’ ← **věverica*; *iriță* ~ *ierîță* ‘яровая пшеница’ ← **jarîca* и нек. др. с менее ясными источниками (Pascu 1916: 254); ср. также с колебанием *toporîște*, *toporîște* ← *toporiște* (из какого языка?) (Pascu 1916: 253).

Как бы то ни было, видно, что слав. *у прямо не отражается в заимствованиях как рум. *î*. Возможно, на начальных стадиях любое слав. *rV, где V = *i, *y, давало раннерум. *rrî (тогда *-ri-* в румынском объясняется поздним заимствованием из славянского). Без этого предположения распределения между рум. *rrî и *ri в славизмах-именах обнаружить не удаётся. Поэтому вряд ли уместно рассматривать такие заимствования отдельно¹².

Исходя из этого, можно предположить, что в южнославянском диалекте, из которого заимствовали слова носители «раннерумынского» языка, *у и *i уже слились. Из наличия нескольких примеров на *rrî может следовать, что это слияние произошло скорее в пользу *у ввиду предшествующего */r/. Возможно, другие согласные смягчались сильнее, ср. рум. *cobilă* ‘кобыла’ ← юсл. **kobila* < **kobýla*; рум. *vidră* ‘выдра’ ← юсл. **vidra* < **výdra* и т. д. (Densusianu 1901: 273). В любом случае, в источнике не было двух фонем /i/ и /y/¹³.

Как было показано выше, немногочисленные славянские глаголы на *-Vr-*, попавшие через румынский в цыганский, распределены в нём между *-řy-* и *-ri-* по наличию *ě в исходе основы славянского глагола (*pyřyl*, *tyřyl*, *vyřyl*, но *nazərıl*, *pərıl*). При этом в самом румынском это противопоставление, видимо, в какой-то момент начало затуманиваться, так что глаголы на *-rî* сейчас нередко имеют варианты на *-rî*, и наоборот. Эти варианты в какой-то степени уже присутствуют в румынских памятниках (выше упомянут вариант *tirî* XVI в.), но цыганским не отражаются (правда, примеров мало).

В книге Oczko 2014 приведены многочисленные примеры славизмов из румынских памятников в основном XVI–XVII вв. (автор расписала данные нескольких словарей). В словарной части книги даётся заглавная форма слова, а затем указания на год и источник данных, а также варианты. Выбор заглавной формы обусловлен современным литературным вариантом, но именно он — среди этих глаголов, — судя по материалу, чаще всего является единственным в памятниках (только у одного глагола есть вариант). Приведём интересующие нас славянские глаголы, данные в этой работе, не попавшие в цыганский (попавшие мы дали выше) с их вариантами (и источниками заимствований).

¹² Общепринятое мнение на этот счёт нам неясно. Обычно слова со слав. *i и *y разделяются, ср. Oczko 2014: 64, 65, где говорится, что *у даёт *î* только после *r*. Это создаёт иллюзию непосредственного соответствия, но точно так же отражается и *i (кроме того, приводимые Очко *pretcă* при схрв. *pritka* и *vedeală* при болг. «*vidělo*» не могут быть доводом в пользу отдельной судьбы *-i-*, т. к. в первом случае первична форма с ятем *preatcă*, мн. *pretci*, *-e*, а во втором — дело явно в румынском корне *ved-*, DER: 889). Неясна и позиция Росетти (Rosetti 1986: 307).

¹³ Этому, на первый взгляд, противоречит рум. *prîsni* ‘брызгать’, приводимое как заимствование из слав. *prysnoti* (Oczko 2014: 65). На самом деле, вероятно, *prîsni* — не что иное как фонетический вариант *pîrsni* (по Scriban 1939: 990 *pîrsni* — на востоке) ← юсл. **přsnuti* (схрв. *přsnuti*, болг. *прѣсна*).

На *-rî*:

- рум. *doborî* и *oborî* ‘валить’ (в памятниках: *oborri*, Densusianu 1938: 121) ← юсл. **do-/o-borîti* (ср. схрв. *obòriti* ‘свалить’);
 рум. *omorî* ‘убивать’, *zamorî* ‘уморить’ (в памятниках: *omorri*, Densusianu 1938: 121) ← юсл. **u-?/za-morîti* (ср. схрв. *mòriti* ‘морить’);
 рум. *coborî* ‘спускаться’, м. б., ⇐ рум. *pogór* ‘спуск’¹⁴ ← юсл. ?, ср. др.-рус. *погорь* ‘вниз’ (поэтому, м. б., это отымённый глагол, а не заимствованный);
 рум. *oṭârî* ‘злиться’ ← юсл. **o-cěriti* (ср. схрв. *cěriti se* ‘скалиться’);
 рум. *zădărî* ‘дразнить, задирать’ ← юсл. **za-dbr-* (ср. схрв. *zàdrēm* ‘задерж’);
 рум. *ocărî* (вар. *ocări*) ‘оговаривать’ ← юсл. **u-?korîti* (ср. схрв. *kòriti* ‘корить’).

В части случаев показательно написание *-rri* в памятниках при совр. рум. *-rî*:

- рум. *pojari* ‘зажечь’ ← юсл. **požariti*; в памятниках: *pojerrri-se-vor* (Densusianu 1938: 121) (в Oczko 2014: 225 приводится с *-ri*, очевидно, по недосмотру).

Отсутствие такого написания в памятниках непоказательно (часть памятников совсем не отражает *-rr-*, в части отражение непоследовательно); так, только с одинарным *-r-*: *dvori* ‘служить’, *blagodari* ~ *blagodări* ‘благодарить’.

Только на *-rî*:

- рум. *zărî* ‘видеть смутно, завидеть’ (и с приставками: *năzărî*, см. выше) ← юсл. **zvrěti* (ср. схрв. *zrěti* ‘зреть, видеть’);
 рум. *dogorî* ‘разгораться’, *prigorî* (возвр.) ‘жариться’ ← юсл. **do-/pri-gorěti* (ср. схрв. *gòrjeti* ‘гореть’).

Возможно, сюда же можно отнести:

- рум. *preapări* ‘убедить’ ← юсл. **prěpъrěti*; (в памятниках: *preapări*, но деепр. *preapărînd*, Densusianu 1938: 524).

Все три глагола на *-rî* — ятевые (сюда же относится *opări*, о котором см. выше), а глаголы на *-rî* — нет. Тем самым как будто подтверждается (хоть и на весьма скудном материале) замеченное нами распределение.

В современном румынском это распределение почти совсем затёрто. Так, имеются варианты: *oṭârî* ~ *oṭări* ‘злиться’; *ogorî* ~ *ogorî* ‘выжечь [поле]’; *dogorî* ~ *dogorî* ‘разгораться’; только *izgorî* (возвр.) ‘греться, преть’ (о зерне); *zădărî* ~ *zădări* ‘дразнить’, сюда же *ijderî* (*ijdări*, *izdări*, *jidări*) ~ *jidări* ‘изобретать, тормозить’; *zătărî* ~ *zatări* (также без еря: *zătrî*) ‘губить’; *dîrî* ~ *dîri* ‘топтать, следить’ (DER: 292) и мн. др. глаголы-славизмы с более или менее ясной этимологией. Имеется также некоторое число глаголов без варианта *-rî*, ср. *găvări* ‘говорить’, вероятно, более новых. Итак, современное распределение совершенно непоказательно. Однако, полагаясь на работу Oczko 2014, мы предположили, что эта ситуация вторична. Если так, то можно считать, что северновлашские цыганские диалекты удивительным образом сохраняют важный румынский архаизм.

В любом случае, разбор упомянутых древних румынских славизмов (в их современном цыганском облике), судя по всему, проливает свет на довольно древнее состояние фонологии каких-то южнославянских диалектов (раннеболгарских?), а именно — позволяет с высокой точностью вообразить реализацию в них двух аллофонов /r/, отразившихся как две фонемы в румынском. Эти наблюдения могут помочь в т. ч. при уточне-

¹⁴ Видимо, сюда же *pogorî*, значение = *coborî* (DER: 2201).

нии датировки миграции цыган и, возможно, также раннего румынско-славянского культурного взаимодействия.

Кроме того, мы показали, что в румынских диалектах, окружавших цыган, совершенно явственно сохранялись две фонемы: */r̄/ и */r/, которые непоследовательно отражаются румынскими памятниками. Из того, что у цыган (живших на румыноязычной территории до середины или конца XIX в.) это различие осталось нетронутым, следует, по-видимому, что в румынском языке оно исчезло совсем недавно.

Итак, в раннерумынском языке могло быть следующее распределение (для исконных слов и славизмов; о других заимствованиях и о конечной позиции см. выше):

	в начале	в середине (V_V)
унаследованные латинские слова	r- > rr-	-r- > -r- -rr- > -rr-
заимствования из южнославянского	*r- → rr-	*-r- (перед *ǣ, *b) → -r- *-r- (проч. случаи) → -rr- (возможно, с колебаниями перед *i)

Сделанные нами выводы не приводят к каким-либо суждениям о фонематическом статусе «твёрдого» и «мягкого» -r- в южнославянской эпохи его активных контактов с дакорумынским. Однако важно, что акустически они были достаточно различны, чтобы передаваться разными романскими фонемами.

Сама потеря противопоставления псл. *y и *i широко демонстрируется всеми славянскими языками региона. Так, эти фонемы слились во всех славянских языках, кроме части восточнославянских и западнославянских диалектов. Разные рефлексy дают следующие языки: русский, белорусский, маргинальные украинские диалекты, польский, лужицкие, в чешском и словацком это различие в части случаев видно по предшествующим согласным. В остальных это противопоставление пропало довольно рано. Судя про рассмотренному в этой статье материалу, сначала слияние этих двух фонем фонетически прошло в пользу *y (такова современная ситуация в большинстве украинских говоров, в т. ч. в литературном). Вероятно, таков был и ранний южнославянский диалект, с которым контактировали румыны, и именно этот факт отразился в северновлашских цыганских диалектах.

Литература

- БЕР = Во. Георгиев, Ив. Гълъбов, Й. Заимов, Ст. Илчев, М. Рачева, Т. Ат. Тодоров и др. (ред.). 1971–2002. *Български етимологичен речник*. София: Академично издателство «Проф. М. Дринов».
- Гинкуловъ, Я. Д. 1840. *Начертание правилъ валахо-молдавской грамматики, составленное Я. Гинкуловымъ*. Санктпетербургъ. Въ типографіи Императорской Академіи Наукъ.
- ДД = Деметер, Р. С., П. С. Деметер. 1990. *Цыганско-русский и русско-цыганский словарь (кэлдэрарский диалект)*. Под редакцией Л. Н. Черенкова. Москва: Апостроф.
- Кожанов, К. А. Рукопись. *Грамматика кишинёвского диалекта; Русско-цыганский словарь: кишинёвский диалект*. Неопубл. рукопись.
- Ослон, М. В. 2018а. *Язык котляров-молдова. Грамматика кэлдэрарского диалекта цыганского языка в русскоязычном окружении*. Москва: Языки славянской культуры (в печати). Электронная версия доступна по адресу: [www.rromanes.org/pub/Oslon/Ослон М.В. Язык котляров-молдова.pdf](http://www.rromanes.org/pub/Oslon/Ослон%20М.В.%20Язык%20котляров-молдова.pdf).
- Ослон, М. В. 2018b. *Историческая фонетика цыганских диалектов*. Неопубл. рукопись. Электронная версия доступна по адресу: [www.rromanes.org/pub/Oslon/Ослон М.В. Историческая фонетика цыганских диалектов.pdf](http://www.rromanes.org/pub/Oslon/Ослон%20М.В.%20Историческая%20фонетика%20цыганских%20диалектов.pdf).

References

- BER = Vo. Georgiev, Iv. Gălăbov, J. Zaimov, St. Ilčev, M. Račeva, T. At. Todorov et al. (eds.). 1971–2002. *Bălgarski etimologičen rečnik*. Sofia: Akademično izdatelstvo «Prof. M. Drinov».
- Cihac, Alexandru de. 1879. *Dictionnaire d'étymologie daco-romane*. Francfort s/M: Ludolphe St-Goar.
- DD = Demeter, R. S., P. S. Demeter. 1990. *Cygansko-russkij i russko-cyganskij slovar'* (kêldêrarskij dialekt). Pod redakcij L. N. Čerenkova. Moskva: Apostrof.
- Densusianu, Ovide. 1901. *Histoire de la langue roumaine. Tome 1. Les origines*. Paris: Leroux.
- Densusianu, Ovide. 1938. *Histoire de la langue roumaine. Tome 2. Le seizième siècle*. Paris: Leroux.
- DER = Cioranescu, Alejandro. 1966. *Diccionario etimológico rumano*. Universidad de la Laguna.
- DEX = Academia Română. *Dicționarul explicativ al limbii române (ediția a II-a revăzută și adăugită)*. 2009. București: Institutul de Lingvistică „Iorgu Iordan – Alexandru Rosetti”, Editura Univers Enciclopedic.
- Ginkulov, Ja. D. 1840. *Načertanie pravil valaxo-moldavskoj grammatiki, sostavlennoe Ja. Ginkulovym*. Sankt-peterburg. V tipografii Imperatorskoj Akademii Nauk.
- Kožanov, K. A. Ms. *Grammatika kišinjovskogo dialekta; Russko-cyganskij slovar': kišinjovskij dialekt*. Unpublished manuscript.
- Scriban, August. 1939. *Dicționarul Limbii Românești*. Iași.
- Oczko, Anna. 2014. *Rumuńska słowiańszczyzna. Zapożyczenia południowosłowiańskie w języku rumuńskim w XVI i XVII wieku*. Kraków: Collegium Columbinum.
- Oslon, M. V. 2018a. *Jazyk kotljarov-moldovaja. Grammatika kêldêrarskogo dialekta cyganskogo jazyka v russko-jazyčnom okruženii*. Moskva: Jazyki slavjanskoj kul'tury (in print). Online version available at: [www.rromanes.org/pub/Oslon/Oslon M.V. Язык котляров-молдовая.pdf](http://www.rromanes.org/pub/Oslon/Oslon%20M.V.%20Jazyk%20kotljarov-moldovaja.pdf).
- Oslon, M. V. 2018b. *Istoričeskaja fonetika cyganskix dialektov*. Unpublished manuscript. Online version available at: [www.rromanes.org/pub/Oslon/Oslon M.V. Историческая фонетика цыганских диалектов.pdf](http://www.rromanes.org/pub/Oslon/Oslon%20M.V.%20Istoričeskaja%20fonetika%20cyganskix%20dialektov.pdf).
- Pamfil, Carmen-Gabriela. 2003. *O etimologie controversată: vorbă < (v. rom. dvorbă <) sl. dvoriba*. *Anuar de Lingvistică și Istorie Literară XXXIX-XLI*: 7–30.
- Pascu, Gîorge. 1916. *Sufixele românești*. București: Socec & Co.
- Rosetti, Alexandru. 1986. *Istoria limbii române. I. De la origini pînă la începutul secolului al XVII-lea. Ediție definitivă*. București: Editura Științifică și Enciclopedică.
- Skok, Petar. 1971. *Etimologijski rječnik hrvatskoga ili srpskoga jezika*. Zagreb: Jugoslavenska Akademija znanosti i umjetnosti.

Mikhail Oslon. The reflexes of South Slavic *r in borrowings from Old Romanian into Romani

A thorough examination of North Romani (Kelderar and Kišyniovar) borrowings from Romanian dialects confirms that (1) Romanian has, until quite recently, rigorously preserved Latin -rr- word-internally, and that (2) in Balkan Romance dialects, including Romanian, Lat. r- > rr- word-initially. It also leads us to a new observation that (3) the South Slavic dialect which was the source of numerous loans into “Early Romanian” possessed two distinct allophones of /r/: a soft (palatalized) one — before *ě (possibly, also before *b) — and a hard (non-palatalized) one — in all other contexts (in particular, before *e and probably before *i) — which yielded Romanian -r- and -rr-, respectively.

Keywords: Romani, Romanian, South Slavic, language contact, historical phonology

А. И. Коган / Anton Kogan

(Институт востоковедения РАН, Москва)

Ответ на статью А. С. Крыловой

«Лексикостатистика новоиндоарийских языков: взгляд полевого лингвиста»

В своей недавней статье (Крылова 2017) А. С. Крылова предложила ряд поправок и дополнений к индоарийской лексикостатистической базе данных, приведенной в нашей работе (Коган 2016). Кроме того, ей были высказаны некоторые замечания, касающиеся самих принципов отбора лексики, включаемой в стословные списки. Все эти поправки, дополнения и замечания, безусловно, интересны, но в значительной своей части небесспорны, а потому требуют детального обсуждения. Прежде, однако, хотелось бы поблагодарить А. С. Крылову за пополнение базы материалами, собранными ей в ходе полевой работы. Речь идет, в первую очередь, о материалах по языкам куллуи и ория. Учет этих новых данных, вне всякого сомнения, позволит уточнить родословное древо и сделает его лучше отражающим реальную картину генетических отношений новоиндийских языков.

Разбор замечаний А. С. Крыловой представляется целесообразным производить, следуя порядку и рубрикации, предложенным в ее статье.

Замечания к семантике

1. Индоарийские местоимения, выступающие в стословном значении ‘all’ и продолжающие др.-инд. *sarva*- ‘все’ (хинди-урду *sab*, пандж. *sabh*, неп. *sabai* и т. д.), объявляются этимологически не связанными с местоимениями, восходящими к прототипу *sāra*- ‘весь, целый’ (потх., хиндко *sārē*, лахнда *sārā* и т. д.¹). Утверждается, что два данных этимона не связаны на праиндоевропейском уровне (Крылова 2017: 282). Между тем, связь между ними представляется вероятной, на что в свое время указывал еще Жюль Блок (Bloch 1919: 420). Оба они, скорее всего, восходят к производным и.-е. **solo-*, **sol(e)u-* ‘целый’². Таким об-

¹ Полный перечень отражений этих двух основ в привлеченных нами для рассмотрения новоиндийских языках см. в статье (Коган 2016: 239).

² Отражения данной основы в других индоевропейских языках см. в: Pokorny 1959: 979-980.

разом, присвоение отражениям др.-инд. *sarva*- и *sāra*-³ одного номера в базе данных представляется нам вполне оправданным. Включение же в стословные списки обоих слов в качестве синонимов, как это предлагает А. С. Крылова, учитывая сказанное выше, нецелесообразно.

5. Утверждается, что в стословном списке хинди-урду в значении ‘neck’ дано название горла, а не шеи (Крылова 2017: 283). Однако у хинди-урду *galā*, включенного нами в базу данных, имеется как значение ‘горло’, так и значение ‘шея’ (Бархударов и др. 1972: 452)⁴. Поскольку данное слово является исконным, предпочтение при составлении списка было отдано именно ему, а не персидскому заимствованию *gardan*.

Замечания к этимологиям

1. А. С. Крылова отрицает возможность исконного происхождения названия красного цвета в большинстве новоиндийских языков (хинди-урду, пандж., лахнда, гудж. *lāl*, синдхи *lālu*, бенг., асс. *lal* и т. д.⁵), предлагая выводить его из перс. *la'l* ‘рубин’ (Крылова 2017: 283–284). Следует сказать, что связь индоарийских слов с персидским постулируется и нами (Коган 2005). Однако нам представляется более вероятным не предполагаемое А. С. Крыловой заимствование с последующим появлением у существительного ‘рубин’ дополнительной семантики ‘красный’, а контактирование данного существительного с исконным индоарийским прилагательным. Это последнее мы считаем отражением вероятного прототипа **lōhila-*, а влиянием со стороны семантически близкого персидского слова объясняем появле-

³ Не исключено, что данная основа, вопреки предположению Р. Тернера (Turner 1966: 770), не является этимологически тождественной др.-инд. *sāra*- ‘сердцевина; суть, сущность’.

⁴ Ср., например, такие контексты употребления данного слова, как *galē mē hāth dālnā* ‘обнять за шею’, *galē milnā* ‘обниматься’.

⁵ Полный список новоиндийских форм см. в Коган 2016: 250.

ние в новоиндийских формах долгого \bar{a} на месте ожидаемого \bar{o} ⁶. А. С. Крылова ставит под сомнение саму возможность реконструкции данного прототипа, а между тем, предполагавшаяся Р. Тернером (Turner 1966: 650) праформа **lōhila-*, вероятнее всего, представляет собой региональное отражение др.-инд. *rudhira-* ‘красный’. Утверждение А. С. Крыловой о том, что в средневековых новоиндийских диалектах существительное ‘рубин’ затруднительно отличить от прилагательного ‘рубиновый, красный’ ввиду несформированности системы послелогов, вызывает немало сомнений и при этом едва ли может считаться весомым аргументом. Примеры из языка Кабира, приведенные в (Крылова 2017: 283), демонстрируют отсутствие формальных средств маркировки генитивных определений, однако мы не знаем несомненных примеров функционирования последних в качестве цветообозначений в индоарийских языках. Напротив, известные нам названия цветов, этимологически связанные с существительными, характеризуются морфологическим оформлением, отличным от такового у генитивных определений⁷. Кроме того, если принять точку зрения А. С. Крыловой, остается неясным, почему прилагательное ‘красный’ и существительное ‘рубин’ различаются в современном индоарийском, где на синхронном уровне они, несомненно, представляют собой два разных слова, что в некоторых языках фиксируется в орфографии (ср., например, разное написание этих слов в урду: لال ‘красный’ и لؤلؤ ‘рубин’). Весьма сомнительным представляется и тезис о том, что «основные исконные цветообозначения хинди, как и большинство исконных базовых прилагательных, являются изменяемыми и заканчиваются в исходной форме

⁶ В нашей монографии «Дардские языки: генетическая характеристика» данная гипотеза формулируется следующим образом: «Наличие нерегулярного \bar{a} в современных индоарийских формах, возможно, объясняется аналогией с бытующим во многих языках севера Индостана заимствованным из персидского существительным *lāl* ‘рубин’» (Коган 2005: 157). Неясно, почему описанный сценарий историко-фонетического развития совершенно не упоминается А. С. Крыловой, указавшей в числе аргументов против нашей этимологии на отсутствие долгого \bar{a} в отражениях праформы **lōhila-*, приведенных в словаре Р. Тернера.

⁷ Ср., например, некоторые цветообозначения современного хинди-урду: *baīganī* ‘фиолетовый’ от *baīgan* ‘баклажан’ при генитивном образовании от последнего слова *baīgan kā; sunahrā, sunahlā* ‘золотистый’, несомненно связанное с *ṣōṇā* ‘золото’, но вовсе не идентичное послеложному сочетанию *ṣōṇē kā*, функционирующему как генитивное определение.

на \bar{a} » (Крылова 2017: 284). В лучшем случае это гипотеза, требующая проверки. Осуществление такой проверки, по всей видимости, предполагает расписку больших словарей, а простое перечисление прилагательных из стословного списка, которым ограничивается А. С. Крылова, не может служить надежным обоснованием. Прилагательные хинди-урду *gōl* ‘круглый’ и *bhāri* ‘тяжелый’ представляют собой показательные контрпримеры, причем, вопреки утверждению А. С. Крыловой, их фонетический облик отнюдь не дает оснований «трактовать их как санскритские заимствования»⁸. Отсутствие форм типа **lāl* в новоиндийских языках, где произошел переход $l > !$ (Крылова 2017: 284), может иметь синхронно-фонетическое объяснение: подобные формы попросту трудны для артикуляции.

5. Новоиндийские названия семени, этимологически связанные с др.-инд. *bija-* и не утратившие интервокальную аффрикату *j* (хинди-урду *bīj*, синдхи *b'iju*, каччи *bijj*, куллуи *bejja*, бенг. *bij*, асс. *biz* и др.⁹), должны, по мнению А. С. Крыловой, быть признаны санскритизмами (Крылова 2017: 285). Данное утверждение, однако, по всей видимости неверно. Подобные формы могут рассматриваться как рефлексы засвидетельствованной в древнеиндийском основы *bijya-*, на что указывал еще Р. Тернер (Turner 1966: 523). Конечная гемината в каччи, а также краткий гласный в каччи и синдхи однозначно указывают на более ранний (среднеиндийский) прототип **bijja-*, вполне закономерно продолжающий др.-инд. *bījya-*.
6. Утверждается, что продолжения др.-инд. *sūrya-/sūriya-* ‘солнце’ «утратили в первом случае *r*, во втором *y*» (Крылова 2017: 285). Если первая часть данного тезиса несомненна, то вторая представ-

⁸ Относительно этих двух слов и их трактовки в статье А. С. Крыловой необходимо высказать несколько дополнительных замечаний. Существительное хинди-урду *gōl* ‘круг, шар’ является фонетически регулярным отражением др.-инд. *gōla-* ‘шар’. Адъективное значение ‘круглый’ у этого слова зафиксировано уже в раннесреднеиндийскую эпоху (ср. пали *gōla-* ‘круглый; шар’), поэтому в данном случае едва ли можно вслед за автором статьи предполагать процесс, параллельный образованию *lāl* ‘красный’ от *lāl* ‘рубин’. Прилагательное же *bhāri* ‘тяжелый’ в принципе не может являться санскритизмом: в древнеиндийском в соответствующем значении используется основа *guru-*, а слово, фонетически и семантически тождественное хинди-урду *bhāri* попросту не засвидетельствовано. Традиционно прототипом для этого прилагательного считается др.-инд. *bhārika-* ‘forming a load, heavy’ (Turner 1966: 539).

⁹ Полный список форм см. в Kogan 2016: 252.

ляется спорной. В упомянутой А. С. Крыловой работе К. Масики отмечена возможность перехода интервокального *-y-* > *-jj-* в среднеиндийских диалектах (Masica 1991: 169). В новоиндийскую эпоху гемината регулярным образом упрощалась и закономерное отражение др.-инд. *sūriya-* должно было принять вид *sūrVj*, что и обнаруживается в реально засвидетельствованных новых индоарийских формах (ср. дакхини *sūrīj*, хинди-урду, пандж., гудж., кум. *sūraj* и т. д.¹⁰). Объявлять подобные формы «частично адаптированными санскритизмами», как это делает А. С. Крылова, мы не видим веских оснований.

Поправки к списку хинди

4. Вопреки утверждению А. С. Крыловой (Крылова 2017: 286), хинди *mās* ‘мясо’ едва ли может рассматриваться как основной синоним, который следует включать в стословный список. Слово это относится к книжному стилю, характерно только для литературного хинди и неупотребительно в урду. Его незаимствованное происхождение представляется маловероятным. По всей видимости, исконным отражением др.-инд. *māṃsa-* ‘мясо’ в хинди-урду является устаревшее и практически вышедшее ныне из употребления *mās*¹¹. Включенное же в нашу лексикостатистическую базу персидское заимствование *gōšt* стилистически нейтрально, и необходимости заменять его каким-либо другим словом мы по этой причине не видим.
5. Существительное *mēh* ‘дождь’ имеется во всех доступных нам словарях хинди и урду. Едва ли есть основания считать это слово малоизвестным, как утверждает А. С. Крылова (Крылова 2017: 286). Нет веских причин и объявлять его диалектным заимствованием. Во всяком случае, в пользу этого не свидетельствуют никакие факты исторической фонетики. Поэтому мы считаем вполне оправданным включение в базу именно этого слова, а не его заимствованного синонима *bāriṣ* (< кл.-перс. *bāriṣ* ‘осадки’).

¹⁰ Полный список форм см. в Kogan 2016: 253. Данное историко-фонетическое развитие обнаруживается и в других примерах. Ср., например, хинди-урду *dhīraj* ‘самообладание, стойкость, терпение’ < др.-инд. *dhīrya-/ dhīriya-*.

¹¹ Слово засвидетельствовано, например, в словаре Platts 1884. Ср. родственные формы, восходящие к древнеиндийской незаализованной основе *mās*, в наиболее генетически близких к хинди-урду индоарийских языках: дакхини, пандж. *mās-*, авадхи, кум. *māsu*.

Поправки к списку бенгали

3. Предполагаемое А. С. Крыловой (Крылова 2017: 287) исконное происхождение бенг. *agun* ‘огонь’, хотя и допускается Р. Тернером в качестве одной из возможных альтернатив (Turner 1966: 821), представляется маловероятным, поскольку закономерным отражением др.-инд. *agni-* в бенгали является устаревшее бенг. *agi* ‘огонь’. Скорее в данном случае можно говорить об имевшей место в прошлом контаминации старого адаптированного санскритизма¹² с исконным словом, приведшей к появлению в современной бенгальской форме начального *a* (< *ā*) вместо ожидаемого *ɔ*.

Отдельно хотелось бы остановиться на предлагаемом в работе А. С. Крыловой методе отбора лексики для включения в лексикостатистическую базу. В начале статьи автор заявляет: «Сразу отмечу, что мои представления о методике научного исследования не предполагают ни возможности получить «правильный» стословный список, пользуясь только словарём, ни возможности получить его от какого-то одного, пусть идеального, информанта» (Крылова 2017: 279). В чем состоят эти самые «представления о методике научного исследования», остается во многом неясным, поскольку нигде в дальнейшем тексте А. С. Крылова их полностью и эксплицитно не излагает. Лишь однажды, говоря о трудностях, возникающих при составлении стословного списка хинди, она кратко объясняет причину своего скептического отношения к словарям и нашему методу работы с ними: «... как в двуязычные словари ..., так и в толковые или синонимические ... входит лексика самых разных диалектов, зачастую без поясняющих помет. Почти каждое слово стословника имеет 5-10 синонимов... В этой ситуации тактика А. И. Когана избегать включения в список заимствований при наличии любого синонима, не противоречащего исторической фонетике, может привести к включению в стословный список слов, принадлежащих другим идиомам, расположенным в зоне хинди, и употребляющимся только на ограниченной территории» (Крылова 2017: 285).

Прежде всего, следует сказать, что утверждение об отсутствии поясняющих помет далеко не всегда соответствует действительности. Так, в использо-

¹² Такую трактовку допускает и Р. Тернер (Turner 1966: 821).

вавшемся нами и цитируемом А. С. Крыловой словаре Бархударов и др. 1972 диалектизмы снабжены специальной пометой *диал.* (диалектное). Разумеется, нельзя исключить, что какие-то диалектные элементы остались неотмеченными, но каждый раз, когда исследователь предполагает подобное, ему необходимо обосновывать свое этимологическое решение, т. е. показывать высокую вероятность заимствования того или иного слова из определенного диалекта. Сделать это без обращения к фактам исторической фонетики, как правило, попросту невозможно. В любом случае, едва ли оправдано судить обо всех словарях, исходя из «презюмции виновности», а *ргіогі* приписывая их авторам неспособность или нежелание решать проблему выявления диалектизмов¹³. В связи с этой проблемой А. С. Крылова утверждает также следующее: «Самым простым способом отсечь такие диалектизмы будет приблизительный подсчёт сравнительной употребительности синонимов в корпусе текстов» (Крылова 2017: 285). К сожалению, не было пояснено, в чем секрет эффективности такого способа. Совершенно не понятно, каким образом результаты подсчёта употребительности слова могут указывать на его этимологию.

Подсчётам по корпусам А. С. Крылова вообще отводит, как представляется, незаслуженно важную роль. Предлагаемый ей алгоритм отбора слов для стословника, по-видимому, можно кратко сформулировать так: отбирать слова, обнаруживающие максимальное число вхождений в веб-корпусах. Адекватность данного алгоритма далеко не бесспорна. Необходимым условием включения в лексикостатистическую базу той или иной лексической единицы является, как известно, стилистическая немаркированность последней. Можно ли считать наибольшую встречаемость в текстовом корпусе свидетельством принадлежности к нейтральному стилю? Ясно, что ответ на этот вопрос неоднозначен и зависит как от характера текстов, подобранных для создания корпуса, так и от специфики конкретного языкового материала. Так, в языке хинди стилистические различия в лексике

проявляются прежде всего в использовании в разных стилях заимствований из разных источников. При наличии двух синонимов, один из которых усвоен из санскрита, а другой — из персидского или (через посредство последнего) арабского языка, санскритизм практически всегда принадлежит к книжному или официальному стилю, персизм и арабизм же стилистически нейтральны. Если создатели корпуса отдавали предпочтение текстам, написанным на нормированном литературном языке, встречаемость санскритизмов может превысить встречаемость персизмов и арабизмов. По-видимому, с подобной ситуацией и столкнулась А. С. Крылова, получившая большее число вхождений для таких санскритских заимствований, как *mās* ‘мясо’ и *mahilā* ‘женщина’ в сравнении с их синонимами персидского происхождения *gōšt* и *aurat* соответственно.

Описанный случай представляет собой яркую (и притом, вероятнее всего, не единственную) иллюстрацию того факта, что результаты подсчётов по корпусам иногда могут скорее ввести в заблуждение, нежели прояснить реальную картину. Следует признать, что полученные А. С. Крыловой цифры едва ли могут рассматриваться в качестве надежных указаний на стилистическую характеристику слова. Это в свою очередь заставляет нас задаться вопросом о том, в какой мере эти цифры вообще показательны для лексикостатистических штудий.

Весьма странными представляются нам отдельные решения А. С. Крыловой, касающиеся замены конкретных слов в конкретных списках. Речь идет, прежде всего, о предложениях заменить одни заимствования на другие (поправка №3 к списку хинди, поправка №2 к списку бенгали) или включить в базу два синонима с отрицательными номерами (поправка №7 к списку хинди, поправки №6¹⁴ и 10 к списку бенгали). Влияние таких поправок на результаты лексикостатистических подсчетов, очевидным образом, равно нулю, и оправданность их поэтому представляется по меньшей мере сомнительной.

Учитывая все вышесказанное, мы не считаем необходимым вносить в стословный список хинди

¹³ В данной связи нужно также отметить, что хотя развитая синонимия, действительно, характерна для словарного состава хинди и урду, главным фактором, породившим такую ситуацию, является вовсе не диалектная неоднородность, а влияние более престижных в разные эпохи литературных языков. Это видно, в частности, из того факта, что основная часть синонимов приходится на санскритские, персидские, арабские и английские заимствования. Выделение подобной лексики обычно не сопряжено с какими-либо трудностями.

¹⁴ Поправка №6 к списку бенгали вызывает особенно сильное недоумение. Предлагается внести в стословник англицизм *libhar* ‘печень’ (<англ. liver), употребляемый в бытовом языке «преимущественно в кулинарии» (Крылова 2017: 287). Лексику с подобной сферой употребления, несомненно, никоим образом нельзя считать «стословной». Ср. диагностические контексты для слова «печень», предложенные в статье (Kassian et al. 2010: 67): «Он вырезал печень из туши», «Человечья печень больше собачьей».

поправки, предложенные в статье (Крылова 2017). Поправки к спискам бенгали (исключая рассмотренные и отвергнутые выше), ория и куллуи принимаются нами, причина чему — более глубокое в сравнении с нашим знакомство А. С. Крыловой с материалом этих языков. Результаты лексикостатистических подсчетов по исправленной базе данных позволяют построить родословное древо, изображенное на рис. 1.

В заключение хотелось бы отметить, что даже с учетом всех наших возражений работа А. С. Крыловой дала чрезвычайно интересные результаты. Многие высказанные в ней соображения в силу своей дискуссионности, несомненно, будут способствовать более активному обсуждению как вопросов генетической классификации индоарийских языков, так и некоторых методологических проблем лексикостатистики и глоттохронологии.

Сокращения

асс. — ассамский
 бенг. — бенгали
 гудж. — гуджарати
 др.-инд. — древнеиндийский
 и.-е. — индоевропейский
 кл.-перс. — классический персидский
 кум. — кумауни
 неп. — непали
 пандж. — панджаби
 перс. — персидский
 потх. — потхохари

Сокращения названий языков на родословном древе

ASS — ассамский; AWD — авадхи; BNG — бенгали; BNJ — банджари; BRJ — брадж; DGR — догри; DKH — дакхини; DUM — думаки; GJR — годжри; GRH — гархвали; GUJ — гуджарати; HIM — химачали; HND — хинди-урду; HNK — хиндко; KCH — каччи; KNK — конкани; KUL — куллуи;

KUM — кумауни; LHD — лахда (мултани); MAI — майтхили; MAL — мальдивский; MAR — маратхи; MEW — мевати; MND — мандеали; NEP — непали; ORY — ория; PNJ — панджаби; PTH — потхохари; PRY — парья; RAJ — раджастхани (марвари); ROM — цыганский; SND — синдхи; SNG — сингальский; WGD — вагди; WPH — котгархи

Литература

- Бархударов, А. С., В. М. Бескровный, Г. А. Зограф, В. П. Липеровский (под редакцией В. М. Бескровного). 1972. *Хинди-русский словарь в двух томах*. М.: Советская энциклопедия.
 Коган, А. И. 2005. *Дардские языки. Генетическая характеристика*. Москва: Восточная литература.
 Крылова, А. С. 2017. Лексикостатистика новоиндоарийских языков: взгляд полевого лингвиста. *Вопросы языкового родства* 15(3-4): 279–298.

References

- Barkhudarov, A. S., V. M. Beskrovnyj, G. A. Zograf, V. P. Liperovskij. 1972. *Hindi-russkij slovar' v dvoukh tomakh*. Moskva: Sovetskaja enciklopedija.
 Bloch, Jules. 1920. *La formation de la langue marathe*. Paris: É. Champion.
 Kassian, Alexei, George Starostin, Anna Dybo, Vasilij Chernov. 2010. The Swadesh wordlist. An attempt at semantic specification. *Journal of Language Relationship* 4: 46–89.
 Kogan, Anton. 2016. Genealogical classification of New Indo-Aryan languages and lexicostatistics. *Journal of Language Relationship* 14(3-4): 227–258.
 Kogan, A. I. 2005. *Dardskije jazyki. Geneticheskaia kharakteristika*. Moskva: Vostochnaja literatura.
 Krylova, A. S. 2017. Leksikostatistika novoindoarijskikh jazykov: vzgl'ad polevogo lingvista. *Journal of Language Relationship* 15 (3-4): 279–298.
 Masica, Colin P. 1991. *The Indo-Aryan languages*. Cambridge: Cambridge University Press.
 Platts, John Tompson. 1884. *A dictionary of Urdu, classical Hindi, and English*. London: W. H. Allen & Co.
 Pokorny, Julius. 1959. *Indogermanisches etymologisches Wörterbuch*. Bern / München: Francke Verlag.
 Turner, R. L. 1966. *A comparative dictionary of Indo-Aryan languages*. London: Oxford University Press.

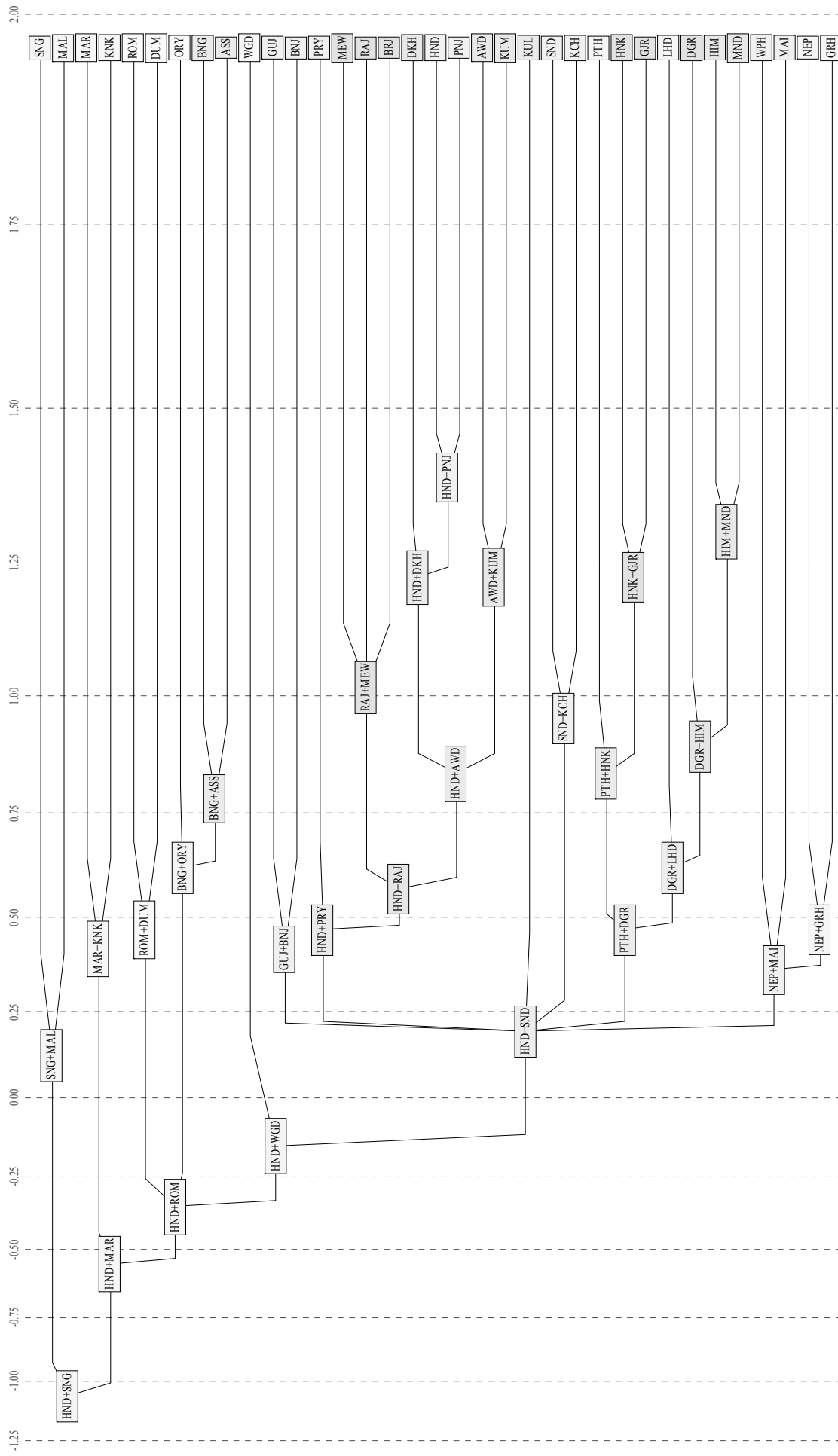


Рис. 1. Родословное древо индоарийских языков, построенное с учетом принятых нами замечаний А. С. Крыловой